The Effects of Multimedia Presentations on Information Processing

Eye-movement analyses of text and picture integration
in a multimedia-based learning scenario

Doctoral Thesis

submitted by

Ying-Hua Guan

Faculty of Linguistics and Literature

University of Bielefeld

Advisors and Reviewers:

Prof. Dr. Hans Strohner, University of Bielefeld

Prof. Dr. Gert Rickheit, University of Bielefeld

Dr. Lorenz Sichelschmidt, University of Bielefeld

Bielefeld, October, 2002

Acknowledgments

I am truly indebted to my parents, brother, and sister for their love and support throughout my graduate years and I am grateful to a number of my friends for encouraging me and taking care of me when I was ill or in trouble. This thesis would not have been possible without the support of my family and those friends. I cannot thank them enough for what they have done for me.

I am also indebted to the following people (listed in alphabetical order) for their inspiration, supervision, proof-reading, and technical and financial support:

Table of Contents

1. INTRODUCTION

Multimedia is widely used in human-human and human-machine communication. In the context of educational practice, multimedia is regarded as a powerful tool for presenting learning materials, as compared with the traditional text-oriented medium. The primary advantage of multimedia presentations (MPs) is that information can be presented in different codes (e.g., verbal and non-verbal) and in different modalities (e.g., in visual and auditory formats) at the same time. Aside from the diverse ways of displaying information, some multimedia learning environments also allow learners to interact with the media. In light of these advantages, it appears convincing that multimedia can deliver information more effectively as well as motivate learners to engage themselves more deeply in the learning process. This has led to a general conviction that multimedia is superior to traditional learning media in every respect. However, during the last decade, empirical studies on the effects of MPs on learning have yielded inconclusive results. In some cases, learning has been promoted by deploying MPs, whereas in some other cases learning did not improve or was even impaired by deploying MPs.

After reviewing the most important theories and empirical studies dealing with the role of MPs in learning, I feel motivated to investigate the following issues:

1. How do different MPs influence the way in which people process information?
2. Under which conditions do the various kinds of MPs facilitate learning?
3. How do different modalities for presenting information interact with regard to the cognitive load on the side of the recipients?

Many researchers in educational psychology focus on examining the relationship between MPs and learning outcomes while ignoring how people process multimedia information. However, in my view, we cannot completely understand the role of MPs in learning without more closely examining the cognitive side of information processing. Consequently, this thesis does not just restrict itself to investigating learning performance, but also aims to provide detailed analyses of the effects of MPs on information processing. The technique employed in order to study the way in which people process multimedia materials is eye-tracking. Eye movements give insight into how people allocate their attention among the different components of a multimedia display in order to integrate information from various

sources. Thus, eye-movement data enable us to reconstruct the cognitive processes by which people utilize multimedia materials.

Another motivation for this work was to empirically discern the characteristics of effective MPs. In order to establish a factual basis that would answer the question regarding under which conditions the various kinds of MPs would facilitate learning, I not only collected findings from a number of published studies, but also conducted two experiments in which diverse ways of presenting information were systematically varied.

Finally, I intend to clarify the influence of MPs on the efficiency of information processing in the recipient's working memory. To this end, I shall provide theoretical and empirical arguments against the claim that information processing becomes more efficient when more modalities are involved in MPs.

1.1 Some terminological preliminaries

There is no standard definition of "multimedia" to be found in the literature. In general, the term "multimedia" refers to an integrated use of text, pictures, video, and audio. Weidenmann (1997) has pointed out that media should be objects, technical devices, or configurations in which information is stored and transmitted. In his view, multimedia is characterized by an integrated use of technical devices such as a PC and video recorder. According to Mayer and Sims (1994), "multimedia" involves the use of more than one presentation medium. Based on their understanding, presentation media comprise, for example, animation and narration. In my opinion, however, a medium is a carrier of information. Thus it is not necessarily tied to a technical device as suggested by Weidenmann because a PC can also play videos without the need for a video recorder. It appears to me that what is generally termed "multimedia" pertains to the way in which information is presented, rather than to the devices that are involved.

Typically, multimedia-based information is presented in different codes (e.g., text and graphics) and in different modalities (e.g., auditory and visual). Multimodal presentations involve information that has to be perceived by different sensory channels. Verbal information, for instance, can be rendered as visual (written) or auditory (spoken) text. However, a problem arises in that moving pictures (e.g., animation or videos) and static pictures (e.g., illustrations or diagrams), though belonging to the same modality, convey information in different ways. I regard motion or animation as a subcategory of visually-based information. In order to avoid terminological confusion and to distinguish between the two ways of

presenting pictorial information, I shall use the term "modality" to refer to visual or auditory presentation, while using the term "presentation mode" to refer to the way in which information is presented. That is, static and moving pictures are taken as two different presentation modes of pictures that fall under the same modality. Finally, for the sake of convenience, I shall use the term "multimedia presentation" to refer to a presentation that is either multicodal or both multicodal and multimodal. The same rationale underlies my use of the term "multimedia information".

1.2 Structure of the thesis

The thesis in hand mainly consists of two parts. The first part (from Chapter 2 to Chapter 6) refers to the theoretical background to my research topic, and the second part (Chapter 7 and Chapter 8) contains new empirical data on multimedia learning. Altogether, the thesis is structured as follows:

- In Chapter 2, I shall provide an overview of theories concerned with the mental representation of verbal and pictorial information.
- In Chapter 3, I shall elucidate how human memory is structured and how the memory systems encode, store, and retrieve multicodal and multimodal information.
- In Chapter 4, I shall review theories and empirical studies regarding learning with text and pictures, and examine the conditions under which pictures facilitate learning.
- In Chapter 5, I shall introduce theories of multimedia-based learning that are widely referred to in instructional design.
- In Chapter 6, I shall explain the connections between eye movements and the underlying cognitive processes. In so doing, I shall discuss a few studies that have investigated eye movement behavior during learning using multimedia materials.
- In Chapters 7 and 8, I shall report on two experiments which I have conducted to investigate the determinants of multimedia learning.
- In Chapter 9, I shall give a few conclusions and possible applications of this research.

2. MENTAL REPRESENTATIONS OF VERBAL AND PICTORIAL INFORMATION

In the field of cognitive science, there are different theoretical approaches that try to explain how information is mentally represented. However, in light of the complex and flexible way in which human cognitive systems operate, the mystery of mental representations has not been completely solved. In this chapter, I shall introduce different theories concerning mental representations of verbal and pictorial information.

2.1 Mental representations of verbal information

2.1.1 The propositional model

When reading or listening to a text, people mentally represent the text they perceive. These mental representations consist of different levels (cf. Kintsch, 1994): 1) The surface level: A text is represented by its phonetic, phonological, graphemic, or lexical features. 2) The propositional level: This includes the syntactic and semantic characteristics of a text. "Propositions can be viewed as semantic molecules that represent the meaning of a text in a rudimentary way." (Rickheit and Sichelschmidt, 1999: 23). The propositional representation of a text describes how those semantic molecules are structured in the recipient's mind. 3) The situation model (or discourse model) level: The situation model is "the representation of the situation described by the text, rather than the text itself." (Kintsch, 1994: 731). That is, readers or listeners may elaborate with additional information that goes beyond the information explicitly given in the text. According to Kintsch (1994), a situation model is not necessarily propositional but might be procedural, abstract, or take the form of a mental image. With respect to text comprehension, the propositional and situational representations are the most crucial ones which I would like to address further.

The propositional model proposed by Kintsch (1974, 1978) deals with the cognitive processes of comprehension and memorization of text. The model assumes that the meaning (i.e. the semantic base) of a text is represented by an ordered list of propositions. A proposition contains a predicate (i.e. a relational concept) and one or more arguments. Predicates may be verbs, adjectives, adverbs, conjunctions,

prepositions, or quantifiers. Arguments may be concepts or other propositions, which perform different semantic functions such as agent, recipient, object, goal, etc. The sentence "Mary gave John a book but he lost it," for example, has three propositions (Kintsch, 1994: 726):

GIVE (agent: MARY, recipient: JOHN, object: BOOK)
LOSE (agent: JOHN, object: BOOK)
BUT (GIVE (MARY, JOHN, BOOK), LOSE (JOHN, BOOK))

The propositions that represent the meaning of a text are ordered hierarchically. The superordinate proposition is shared by several arguments, each of which in turn is shared by some other subordinate propositions within the hierarchy. Kintsch and van Dijk (1978) suggested that the text base be processed in cycles because of the limited capacity of working memory. They assumed that working memory can only process n_i propositions or chunks of propositions at a time, where n_i is contingent upon text and reader/listener characteristics. The working-memory buffer, which holds the most relevant parts of the text base in its current state of development, is of limited size s. In each processing cycle, n_i new propositions and s propositions in the memory buffer are involved, by which connections between the new propositions and those held in the buffer are searched. If any connection is found, the new propositions are added to the previous propositional structure. If none is found, recipients have to search for propositions stored in long-term memory (or eventually re-read the text), or else they must draw appropriate knowledge-based inferences. Propositions that are currently processed in a processing cycle may be stored in long-term memory and reproduced later. An example that demonstrates the processing cycles is taken from Kintsch and van Dijk (1978: 376). The text shown below is an excerpt from a research report entitled "Bumperstickers and the Cops":

> "A series of violent, bloody encounters between police and Black Panther Party members punctuated the early summer days of 1969. Soon after, a group of black students I teach at California State College, Los Angeles, who were members of the Panther Party, began to complain of continuous harassment by law enforcement officers. Among their many grievances, they complained about receiving so many traffic citations that some were in danger of losing their driving privileges. During one lengthy discussion, we realized that all of them drove automobiles with Panther Party signs glued to their bumpers. This is a report of a study that I undertook to assess the seriousness of their charges and determine whether we were hearing the voice of paranoia or reality. (Heussenstam, 1971, p. 32)"

Figure 1 shows the proposition list for the text. Figure 2 demonstrates the processes of cyclical construction for the coherence graph. Figure 3 depicts the complete coherence graph in which the number of boxes shows the number of extra cycles required in processing.

Proposition number	Proposition
1	(SERIES, ENCOUNTER)
2	(VIOLENT, ENCOUNTER)
3	(BLOODY, ENCOUNTER)
4	(BETWEEN, ENCOUNTER, POLICE, BLACK PANTHER)
5	(TIME: IN, ENCOUNTER, SUMMER)
6	(EARLY, SUMMER)
7	(TIME: IN, SUMMER, 1969)
8	(SOON, 9)
9	(AFTER, 4, 16)
10	(GROUP, STUDENT)
11	(BLACK, STUDENT)
12	(TEACH, SPEAKER, STUDENT)
13	(LOCATION: AT, 12, CAL STATE COLLEGE)
14	(LOCATION: AT, CAL STATE COLLEGE, LOS ANGELES)
15	(IS A, STUDENT, BLACK PANTHER)
16	(BEGIN, 17)
17	(COMPLAIN, STUDENT, 19)
18	(CONTINUOUS, 19)
19	(HARASS, POLICE, STUDENT)
20	(AMONG, COMPLAINT)
21	(MANY, COMPLAINT)
22	(COMPLAIN, STUDENT, 23)
23	(RECEIVE, STUDENT, TICKET)
24	(MANY, TICKET)
25	(CAUSE, 23, 27)
26	(SOME, STUDENT)
27	(IN DANGER OF, 26, 28)
28	(LOSE, 26, LICENSE)
29	(DURING, DISCUSSION, 32)
30	(LENGTHY, DISCUSSION)
31	(AND, STUDENT, SPEAKER)
32	(REALIZE, 31, 34)
33	(ALL, STUDENT)
34	(DRIVE, 33, AUTO)
35	(HAVE, AUTO, SIGN)
36	(BLACK PANTHER, SIGN)
37	(GLUED, SIGN, BUMPER)
38	(REPORT, SPEAKER, STUDY)
39	(DO, SPEAKER, STUDY)
40	(PURPOSE, STUDY, 41)
41	(ASSESS, STUDY, 42, 43)
42	(TRUE, 17)
43	(HEAR, 31, 44)
44	(OR, 45, 46)
45	(OF REALITY, VOICE)
46	(OF PARANOIA, VOICE)

Figure 1.
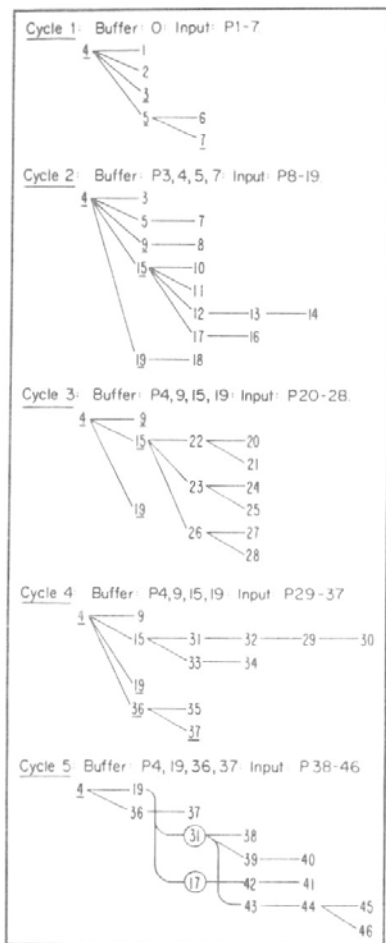
Figure 2.
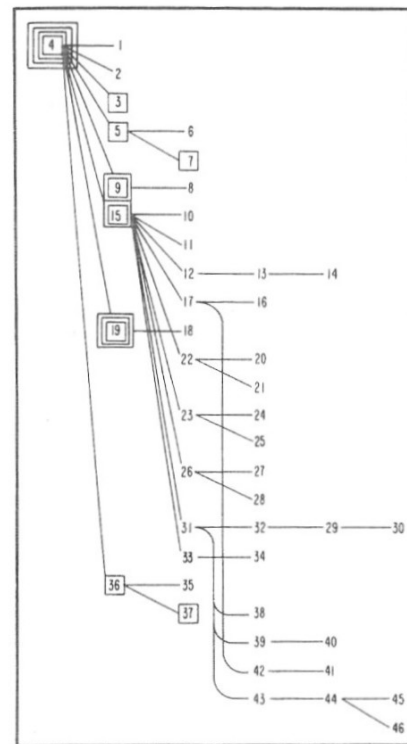
Figure 3.

During each cycle, a subset of *s* relevant propositions is selected and held over in the buffer for the next processing cycle. In effect, the most relevant propositions (usually those that are high in the hierarchy) will participate in processing cycles more frequently, i.e. be often activated in the working-memory. This can explain why those propositions are remembered better than the less relevant ones (those that are low in the hierarchy).

There are, however, some serious limitations of Kintsch and van Dijk's (1978) propositional model. First, the referential identity of argument concepts was taken as the basis for the coherence relationships within a text. Nevertheless, the referential identity does not guarantee coherence. For example, "His favorite animal is the dog. Dogs are a kind of mammal. Cats and dogs are enemies." This is a string of sentences that share a common referent (dog), but that cannot be regarded as a coherent text. Secondly, the model does not clearly explain how and to what extent inferences are drawn. Besides, drawing inferences should not be viewed as 'the last resort' for establishing coherence because recipients do often use their knowledge

during text processing to infer information that goes beyond the text. Thirdly, the propositional model is able to describe the microstructure (i.e. local structure) of a text. However, for a longer text, it fails to describe the macrostructure (i.e. global structure) appropriately. Kintsch (1994) has pointed out that "Understanding a text is not just a matter of understanding each phrase and sentence and linking them together in a coherent network. It also has a global processing component." (Kintsch, 1994: 733). The global coherence of a text involves discourse understanding, that is, building a model of the situation described by the text. To construct a situation model, the pragmatic and social context must be considered as well. Nonetheless, the discourse understanding in the propositional model is only restricted to the semantic level.

To account for those problems, Kintsch (1988) proposed *the construction-integration model*. According to this model, text comprehension begins by constructing a network of the representational units (i.e. concepts) and their interrelationships as stipulated by the text. The processes of construction are not necessarily precise. For instance, words or sentences with ambiguous meanings (e.g., homonyms: 'bank') are initially represented by their possible meanings in the network at the same time. The construction process is followed by the integration process, which is postulated as a process of spreading activation within the network. Through this mechanism, strongly interconnected parts in the network (i.e. contextually relevant concepts) are strengthened, whereas isolated parts (i.e. contextually irrelevant concepts) are deactivated (Kintsch, 1994). Consequently, any contextually inappropriate meanings of the ambiguous words or sentences are filtered out during the integration processes, which serve as the context effect on text comprehension. While the schema theories assume that the schemata (scripts or frames) existing in knowledge control the context-sensitive operation in constructing the situation model in the first place, the construction-integration model assumes that context sensitivity of knowledge activation is "an uncontrolled, bottom-up process, determined only by the strength of the associations between items in long-term memory and the text." (Kintsch, 1994: 733).

As to the processing cycles, the model assumes that when a new sentence is processed, the most strongly activated proposition(s) from the previous sentence is (are) always held in the focus of attention (or short-term memory buffer) to maintain the coherence of the network. This is based on the assumption that information kept in the focus of attention is linked to its related information in long-term memory. In this case, the connected information in long-term memory becomes readily available for further processing. According to Ericsson and Kintsch (1995), the effective

capacity of working memory could therefore be increased, which they have termed "long-term working memory".

The construction of macrostructure is assumed to be carried out strategically. The mental representation of the macrostructure of the text in readers' or listeners' minds might or might not correspond to the author's intention, depending on their individual comprehension strategies or goals. All in all, the construction-integration model has taken the context effect and inference during text comprehension into account. The model allows the construction of a macrostructure as well as a situation model of the discourse.

Of course, this model is only one of the several theoretical approaches to the issue of mental representations of verbal information. Other theoretical approaches such as schema theories are concerned with how knowledge in long-term memory affects the cognitive processes during the construction and reproduction of information. In the following section, I shall briefly outline the schema theories.

2.1.2 Schema theories

In contrast to the propositional model, schema theories argue in favor of top-down processing components. Schema theorists maintain that semantic memory is structured in schemata, which are much larger knowledge structures than propositions. "Schemata may be viewed as hierarchically ordered, abstract, stereotypical representations of objects or states of the world which provide empty slots to be specified in individual contexts." (Rickheit and Sichelschmidt, 1999: 26). These organized knowledge structures are essential for information processing because they influence the way in which people comprehend, interpret, and remember information.

Empirical evidence supporting schema theory was first brought out by Bartlett (1932). He showed his subjects a North American Indian folk tale and found that when subjects recalled the story, they either omitted the part that did not fit their prior expectation or schemata, or distorted some unfamiliar situations described in the story (reported in Baddeley, 1997). Following Bartlett (1932), a schema is an organized structure of our knowledge and expectations of a certain aspect of the world. When learning or remembering new information, people elaborate the meaning of the to-be-learned information with the help of schemata. If the new information is not compatible with the schema, distortions occur.

Some other theoretical terms such as 'story grammar' (Rumelhart, 1975), scripts (Schank and Abelson, 1977), or frames (Minsky, 1975) are similar to the concept of schema. A story grammar refers to the rules describing the structure

underlying a story, which influences how people remember a story. It was proved in a study by Thorndyke (1977) that 1) the violation of story grammar leads to poor recall of the story; 2) propositions that are essential to comprehension tend to be better recalled; 3) there is an interaction between the text structure and the importance of propositions. In the less structured text, important propositions are not better recalled (reported in Baddeley, 1997). Frames and scripts, on the other hand, refer to knowledge structures that represent conventions of social events, for example being in a restaurant. Such knowledge structures are important to discourse understanding, and help people to behave appropriately in social situations. Similar to schemata, scripts are organized in a hierarchical order. In a study by Bower et al. (1979), it was found that people tended to recall information that was not explicitly uttered in the story, but which was consistent with the script. This result is analogous to Bartlett's finding.

Taken together, the schema theories attempt to explain how knowledge is structured in long-term memory and how those structures affect people as they process and remember new information. The notion of schema does not only apply to describing knowledge structures in memory, but is also used in some learning theories (see Chapter 3, Section 3.4.1). The functions of schema in information processing include: 1) *selection*: Only the information that is relevant to the existing schemata will be incorporated; 2) *abstraction*: Only the meaning (not the form) of information is processed; 3) *interpretation*: The information to be processed is interpreted in regard to the already existing schemata; 4) *integration*: The processed information will be either connected to or absorbed by the existing schemata (Thorndyke and Yekovich, 1980; Alba and Hasher, 1983; cited in Rickheit and Strohner, 1993).

2.2 Mental representations of pictorial information

The format of mental representations of pictorial information (i.e. mental images) is still controversial. There has been debate regarding the nature of mental images. Two opposing theoretical approaches are *the analogical* versus *the propositional* approaches. The debate is concerned with whether mental images are like pictures (or percepts) that are analogous to what we see, or whether they are represented as propositions. An example of these two different formats of mental representation is shown in Figure 4.

"A BALL IS ON A BOX"
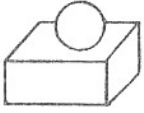
Propositional Representation (Description)	Quasi-pictorial Representation (Depiction)
ON (BALL, BOX)	
1. Relation	1. No distinct relation
2. Argument(s)	2. No distinct arguments
3. Syntax	3. No clear syntax
4. Truth value	4. Truth value only when under a particular description
5. Abstract	5. Concrete
6. Not occur in spatial medium	6. Occurs in spatial medium
7. No abstract spatial isomorphism	7. Abstract spatial isomorphism
a) No necessary part/whole relations	a) Necessary part/whole relations
b) Size and orientation optional	b) Size and orientation necessary
c) Arbitrary marks	c) Non-arbitrary marks
8. No abstract surface-property isomorphism	8. Abstract surface-property isomorphism
a) No necessary part/whole relations	a) Necessary part/whole relations
b) Shape not necessary	b) Shape necessary
c) Arbitrary marks	c) Non-arbitrary marks

Figure 4: Properties of propositional and quasi-pictorial formats
(Taken from Kosslyn, 1980: 31)

Empirical evidence in favor of the analogical approach was provided by Kosslyn (1973, 1980) and Shepard (Shepard and Metzler, 1971). In a series of studies, Kosslyn (1980) asked his subjects to mentally scan the pictures they had just studied and to report whether they saw the target items (given by the experimenter) on their mental images. He found that subjects typically required more time to give an answer when their current view point on the mental image and the target item were a great distance apart (the distance effect), or when the target item was embedded in a very small image (the size effect). Based on those findings, Kosslyn suggested that subjects in those experiments indeed used their "mind's eye" to scan their mental images, which must be analogical in nature. If mental images were represented propositionally, the reaction time should not have been affected by the distance or the size. Similarly, Shepard and Metzler (1971) showed their subjects a pair of three-dimensional objects on each occasion (see Figure 5) and asked them to judge whether they were the same objects. Since the objects were portrayed in different perspectives, subjects had to mentally rotate one of the objects to make a decision. The results indicated that the reaction time was a linearly increasing

function of the angular difference in the orientations of the two objects portrayed in the stimuli, which also favors the analogy view.

(A)   (B)   (C)
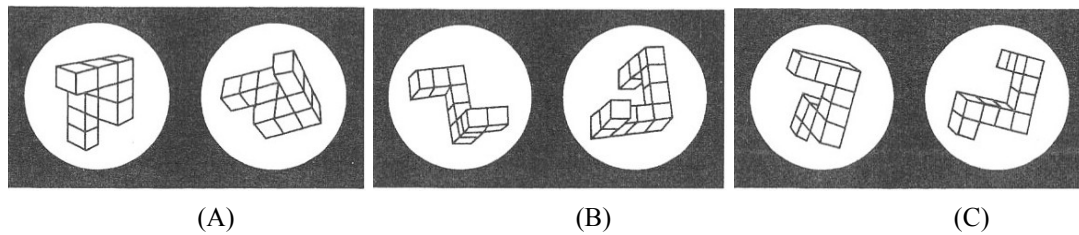
Figure 5: Examples of the stimuli used for mental rotation tasks
(Taken from Shepard and Metzler, 1971: 702)

In contrast to the propositional approach, Pylyshyn (1973, 1979, 1981) argued that mental images are propositional in character because the human mind only employs propositional representations which are abstract and are not tied to any particular sensory modality (cf. Douglas and Brian, 1992). The "depictive" representations are taken as "general reasoning adherents" (Pylyshyn, 2001:4) that accompany a set of underlying processes that operate upon spatial information in terms of discrete propositions. That is, the "picture-like" mental imagery that people experience is not constructed automatically when processing pictorial information. The proponents of the propositional approach criticize that the observations made by the proponents of the analogical approach are dubious because their research methods confound the results that they obtained. Pylyshyn (2001, 2002) claimed that subjects in Kosslyn's experiments experienced "seeing an image with the mind's eye," which was only based on an illusion. "…the experiments were revealing what subjects believed about what would happen if they were looking at a certain scene and not the inherent nature of an imagery medium or mechanism." (Pylyshyn, 2001: 2). In other words, the fact that subjects reported using depictive mental images to perform the task was an illusion elicited by the experimenter's instruction.

Besides, the effects commonly attributed to imagery can be explained on a propositional basis as well. For instance, the distance effect observed by Kosslyn (1980), did not necessarily result from mentally scanning a quasi-pictorial imagery. The same effect can also be explained by propositional representations if one considers the following example (see Figure 6). According to the distance effect, it is assumed that if one were mentally focusing on the left end of the speedboat, it would take less time to see the motor than to see the porthole, and less time to see the porthole than to see the anchor. Nonetheless, subjects could possibly construct propositional representations of the speedboat, like the one shown in Figure 7. The

greater the distance between the subject's focus and the target object, the more links that must be traversed to reach the target in the graph.
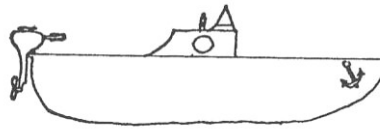
Figure 6: An example of the line drawings used as stimuli by Kosslyn (1973). (Taken from Kosslyn, 1980: 36).
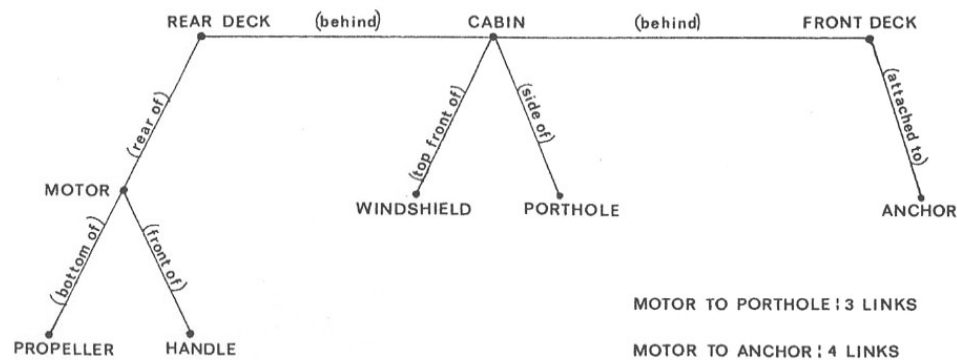
Figure 7: A propositional representation of the drawing of the speedboat. (Taken from Kosslyn, 1980: 39)

Furthermore, the mental images that subjects formed are not supposed to be analogous to the visual stimuli, but are rather constructed based on their knowledge of the world. "…both "mental scanning" and "mental rotation" transformations can be critically influenced by varying the instructions given to subjects and the precise form the task used and that the form of the influence is explainable in terms of the semantic content of subjects' beliefs and goals—that is, that these operations are cognitively penetrable by subjects' beliefs and goals." (Pylyshyn, 1981: 16).

The imagery debate has continued for about two decades. Kosslyn (1994) tried to use neuropsychological evidence to support the analogy claim. It was found that the primary visual cortex (Area 17) is activated when generating visual images, and that during imagery, a retinotopic, quasi-pictorial display is generated on the surface of the visual cortex. That is, mental images should be in a depictive form, just like in a two-dimensional picture. Pylyshyn (2001, 2002) argued that this kind of two-dimensional retinotopic imagery is only literal. It does not represent the form of the functional mental representations involved in vision because the visual inputs we perceive from the world are substantially more complicated than a two-dimensional

picture. For example, how people mentally rotate a three-dimensional object cannot be accounted for by a two-dimensional projection on the primary visual cortex. In other words, the neuropsychological evidence does not further enhance our understanding of the nature of mental images.

The imagery debate has not come to an end. In general, the propositional approach does not deny that people would deploy depictive mental images to perform certain tasks such as solving problems involving geometric displays. As Pylyshyn stated, "This is not a case of believing that images do not exist or are "epiphenomenal". It is a question of whether theories of mental imagery that posit 2D displays or "depictive representations" are empirically correct, or perhaps even coherent." (Pylyshyn, 2001: 3).

In my opinion, the ultimate solution has yet to be found, probably because we cannot precisely measure or trace how the human brain processes information through any currently available methods or technical instruments. In my opinion, mental images could comprise both analogical as well as propositional representations, and whether or not imagery is more analogical or more propositional might be contingent upon the nature of the tasks people are dealing with. Moreover, even though mental images are penetrable by tacit knowledge, so that what people "see" does not correspond to what they perceive in reality, it is not necessary to reject the usefulness of depictive representations. For example, when performing a mental-rotation task, I personally believe that we need both analogical and propositional representations because we must use our knowledge (the propositional representations) to rotate an object mentally, but we cannot simply rotate something without "picturing" its shape.

2.3 Dual coding theory

Dual coding theory (DCT) was proposed by Allan Paivio (1967, 1969, 1971, 1986). The theory was developed from a large number of studies on the role of imagery in associative learning. The imagery debate just mentioned was originally elicited by DCT. Based on the findings of those studies, DCT assumes that:

1) Human memory consists of modality-specific components for information processing. Information is represented in memory in a multimodal fashion, which is in contrast to the view that information is represented by abstract, amodal propositions.

2) DCT proposes two separate subsystems for human cognition. One specializes in the representation and processing of nonverbal information, whereas the other

specializes in the representation and processing of language (see Figure 8). "…the language system is peculiar in that it deals directly with linguistic input and output (in the form of speech or writing) while at the same time serving a symbolic function with respect to nonverbal objects, events, and behaviors. Any representational theory must accommodate this functional duality." (Paivio, 1986: 53). As to the nonverbal system, it deals with information in different modalities—visual, auditory, haptic, gustatory, olfactory, and affective— DCT research was, however, more focused on the visual one.

3) The representational units of the verbal and visual systems are supposed to be "modality-specific perceptual-motor analogues." (Paivio, 1991: 258). The units are hierarchically organized structures. The concept of 'unit' is similar to that of 'chunk', which is flexible in size. The representational units of the visual system are called imagens, whereas those of the verbal system are called logogens. Imagens are mental images that are analogous to the events they denote. The concept "logogen" was first used by Morton (1969). It was taken as a word template or feature pattern that accounts for word-recognition performance. Morton (1979) further postulated modality-specific logogens (visual vs. auditory logogens) and distinguished input- from output-logogens. The concept of logogens used in DCT, however, is broader and more flexible. The terms "imagens and logogens serve mainly to distinguish the underlying (hypothetical) cognitive representations from their expressions as consciously experienced images and inner speech, or overt behaviors such as drawing and speech." (Paivio, 1986: 59). Both imagens and logogens can function as "integrated informational structures or response generators for some purposes." (Paivio, 1986: 59). Information in the verbal system is processed sequentially, whereas information in the visual system is processed in parallel.

4) Verbal and nonverbal systems are independent. During representational processing, logogens are directly activated by linguistic inputs, and imagens are directly activated by nonverbal inputs through representational connections. However, the systems are interconnected by referential links, so that the activation from one representational unit to the other(s) between systems is possible. In light of referential processing, "verbal and nonverbal codes corresponding to the same object can have additive effects on recall." (Paivio, 1991: 259).

5) Pictures are recalled better than words (picture superiority), and concrete words or sentences are recalled better than abstract words or sentences (concreteness effect) because pictures and concrete words or sentences are coded both in verbal and imaginal formats in memory (the conceptual-peg hypothesis). Dual coding facilitates recall because one representational unit in a system may trigger the activity of the corresponding one in another system through the referential connection. It should be

noted that the interunit processing is optional. That is, Paivio did not claim that picture naming or imagining concrete words or sentences is automatic even though it is highly likely to occur under some circumstances (cf. Paivio, 1986: 62).
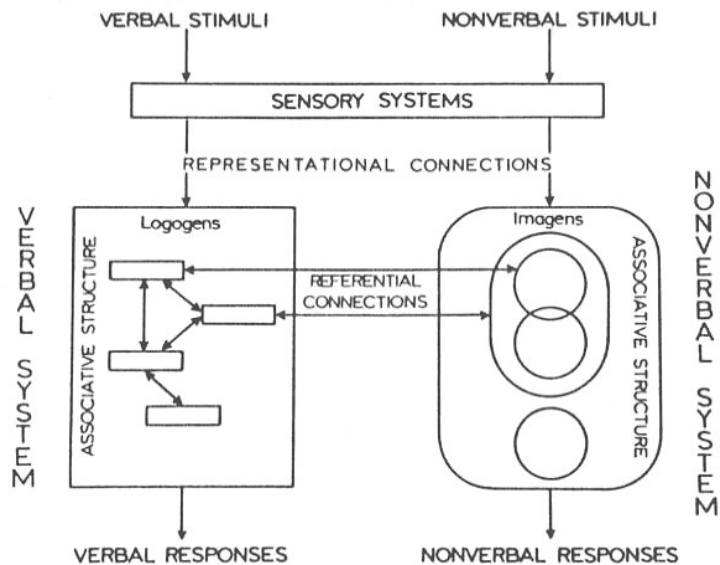
Figure 8: Verbal and nonverbal symbolic systems of dual coding theory.
(Taken from Paivio, 1991: 152).

Although the principles and assumptions of DCT are supported by a number of empirical studies, they are not without controversy. For example, the recall of concrete sentences was better than that of abstract sentences in general. According to DCT, the better recall of concrete sentences is attributed to the integrative memory induced by imagery. However, Marschark and Paivio (1977) found that when recall was successful, the memory of abstract sentences was also integrated (or holistic). Moreover, the picture superiority effect is questionable because it was found that the imaged words were recalled about as well as named pictures (Paivio, 1991). That is, the difference in recall between pictures and words was dependent on the experimental instructions. Finally, the propositional approach criticizes the assumption of modality-specific mental representations suggested by DCT. As I have mentioned in the previous section, the propositional approach assumes that all kinds of information are represented in a unitary form—as propositions—in the human mind. Despite these criticisms, DCT seems to have survived and has been regarded as the dominant theory for explaining the effects of pictorial aids on learning.

2.4 Mental model

The term "mental models" was first used by Craik (1943) in the sense that "the mind constructs "small-scale models" of reality to anticipate events, to reason, and to underlie explanation." (Wilson and Keil, 1999: 525). Though the term "mental models" is widely used, the definitions of mental models are vague and quite diverse in different research fields. According to Johnson-Laird (1989), a mental model is to be understood as a representation of a body of knowledge that meets the following conditions:

1) The structure of a mental model corresponds to the structure of the situation it represents;

2) A mental model can comprise elements that correspond to perceptible entities or abstract notions;

3) Unlike other forms of mental representations, a mental model does not contain variables. Instead, it employs tokens representing a set of individuals.

Two broad classes of mental models have been proposed by Johnson-Laird (1983): physical and conceptual:

1. Physical models represent physical things or states of affairs and are accessible to empirical observation. Physical models contain a finite set of entity or property tokens, and a finite set of relations between them. Types of physical models are as follows:

- Relational models represent entities and their properties.
- Spatial models represent spatial relations between entities with properties.
- Temporal models represent changes in entities with properties.
- Kinetic models represent ongoing change and movement of entities (like mental simulation).
- Dynamic models represent alteration and movement in regard to causal contingencies.
- Imaginal models are an observer-centered representation of the visual characteristics of an underlying spatial or kinetic model.

2. Conceptual models represent abstractness and truth or fiction. They are able to explain phenomena such as negation, conjunction and disjunction. Types of conceptual models are, for example, as follows:

- Monadic models represent entities with their properties, and statements in terms of the existence and identity of entities.
- Relational models represent a finite set of abstract relations between entities.
- Metalinguistic models represent entities that represent linguistic expressions.

- Set-theoretical models represent specific or vague quantification. Numbers that cannot be easily visualized can be represented by a corresponding propositional label.

In the context of discourse comprehension, mental models are regarded as "dynamic cognitive representations of the contents of an utterance on the part of the recipient" (Rickheit and Sichelschmidt, 1999: 24). To mentally set up a corresponding structure in light of the situations described by the verbal discourse, it is assumed that mental models may consist of analogical components such as "quasi-pictorial images" (cf. Rickheit and Sichelschmidt, 1999). Nevertheless, unlike mental images, mental models are not bound to specific sensory modalities and are capable of representing abstract notions (Schnotz, 2002). In addition, mental models are assumed to contain new information that is not explicitly uttered in the discourse but is inferred by the recipients.

Unlike the propositional approach according to which text comprehension is carried out by sequentially connecting lists of propositions, the mental model approach assumes that text comprehension is based on the construction of a mental model of the facts described by the text. In other words, mental models may go beyond the text base or the propositional representation of a text, respectively. In the construction of a mental model, the information given in the text is integrated with the recipient's knowledge which initially plays an important role in text comprehension. That is, due to the recipient's prior knowledge, certain expectations are already imposed on the way in which he or she interprets the text. In addition, it is assumed that the processes of mental-model construction are incremental. The initially-built mental model is assumed to be constantly modified and elaborated in the course of text processing (cf. Schnotz, 1988).

There is some empirical evidence supporting the mental model approach. For example, some studies showed that subjects had difficulty in understanding a text properly when the topic from the proceeding text was suddenly changed or when referential connections between sentences were not clear, so that subjects failed to establish a coherence between the sentences. (Lesgold et al., 1979; Collins et al., 1980; Sanford et al., 1980; Sanford and Garrod, 1981, 1982; Schnotz, 1985). In those studies, the difficulty of text comprehension was typically revealed by the prolonged reading time measured in the passage of the text, where subjects encountered the problems just mentioned. The theorists of the mental model approach argue that the prolonged reading time indicates that subjects had to draw inferences with the help of their knowledge in order to understand the text because the information provided in the text was not sufficient for text comprehension. In contrast, the propositional

approach fails to explain why the reading time was prolonged and how recipients should draw inferences to solve the comprehension problems.

Moreover, evidence supporting the view that a mental model can comprise analogical components corresponding to the structure of the objects and events it represents came from the studies conducted by the proponents of the analogical approach to imagery (Kosslyn, 1980; Shepard and Metzler, 1971; Shepard and Cooper, 1982; Moyer, 1973; Moyer and Landauer, 1967; Paivio, 1975). Since I have explained the view of this approach in Section 2.2, I will not repeat it here. Some other studies provided evidence indicating that subjects could remember the superficial structure of an ambiguous text very well without really understanding the meaning of the text. In contrast, subjects could remember the meaning of an unambiguous text very well but not the superficial structure of the text. (Bryant and Trabasso, 1971; Trabasso et al., 1975; Mani and Johnson-Laird, 1982; Perrig und Kintsch, 1985). This indicated that subjects could construct an adequate mental model of a text only when the text provided unambiguous information. When the text information was ambiguous, the construction of the mental model was hindered because there were many possible mental models that could be built into the same text at the same time. Subjects, however, could not know which mental model was correct. Therefore, they were unlikely to build an adequate mental model (cf. Schnotz, 1988).

All in all, the mental model approach seems to be more capable of explaining text comprehension in comparison to the propositional approach because it includes analogical and dynamic mental representations and also takes the influence of recipients' knowledge with regard to information processing into account. In some respects, the concept of mental models is quite similar to that of schema. The differences between these two concepts seems to be that mental models refer to a more concrete form of mental representation than schemata because mental models can be regarded as a kind of "filled" schemata. Mental models do not have variables, whereas schemata do.

2.4.1 Mental models and multimedia learning

There are different assumptions concerning how people constructed a mental model when they process multimedia learning materials comprising texts and pictures. The cognitive theory of multimedia learning proposed by Mayer and his colleagues (Moreno & Mayer, 2000a; Mayer and Moreno, 2002a, 2002b) suggests that people initially build a mental representation of text information, which is termed *word base*, and a mental representation of picture information, which is termed *image*

base. And then, a verbal mental model for texts and a pictorial mental model for pictures are constructed separately. Subsequently, the verbal and the pictorial mental models are supposed to be integrated in their working memory. The prior knowledge stored in long-term memory is assumed to be actively involved in the integration processes as well (for more details see Chapter 5, Section 5.2).

In contrast to Mayer's theory, Schnotz and Bannert (1999) proposed an integrative model of text and picture comprehension which provides a more plausible and elaborated explanation concerning how text and picture information is mentally represented during information processing. An outline of this model is depicted by Figure 9. According to this model, the mental representation of texts and of pictures follows quite different principles because texts and pictures are based on different sign systems. The mental representation of texts is *descriptive* in nature, whereas that of pictures is *depictive* in character. The interaction between descriptive representations is based on *symbol processing*, whereas the interaction between depictive representations is based on *structure mapping*. The processes of processing text and picture information are both "…based on an interaction of bottom-up and top-down activation of cognitive schemata that have both a selective and an organizing function." (Schnotz, 2002: 108).

In the processing of text information, a recipient builds a mental representation of the text surface structure, and generates a propositional representation of the semantic content, based on which a mental model is then constructed. In the processing of picture information, a recipient builds a visual mental representation of the picture, and represents the picture's semantic content by means of a mental model and a propositional representation of the subject matter shown in the picture. "When a mental model has been constructed, new information can be read from the model through a process of model inspection. The new information gained in this way is made explicit by encoding it in a propositional format. The new propositional information is used to elaborate the propositional representation. In other words, there is a continuous interaction between the propositional representation and the mental model." (Schnotz, 2002: 110). Moreover, it is assumed that an interaction between the text surface representation and the mental model, and between the visual representation of the picture and the propositional representation may occur (see the dotted arrows in the diagram).
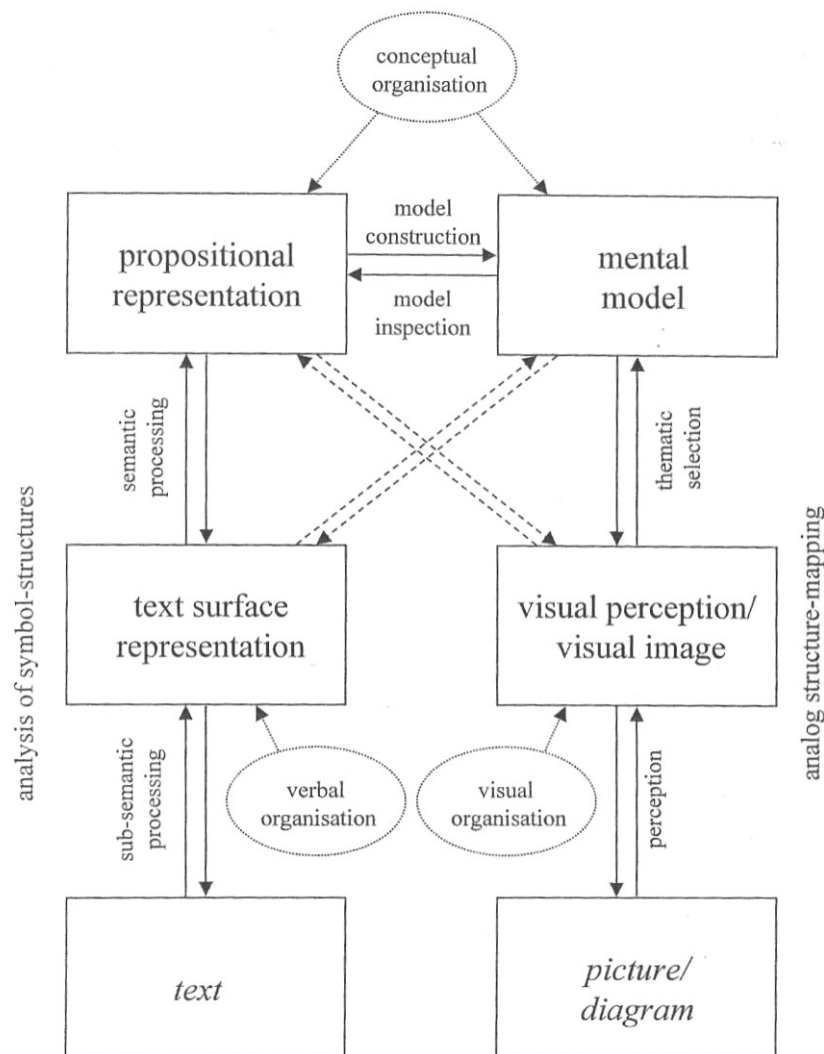
Figure 9: Schematic illustration of an integrative model of text and picture comprehension. (Taken from Schnotz, 2002: 109).

To sum up, an essential idea proposed by the integrative model of text and picture comprehension is that both texts and pictures can generate internal mental representations that are descriptive as well as depictive. During information processing, a number of interactions occur between text processing and picture processing at different processing levels, so that the propositional representation of texts can affect the construction of a mental model of pictures and vice versa. In my view, this is one point that distinguishes this model from Mayer's theory as well as from DCT. In Mayer's theory, the constructions of a verbal and a pictorial mental model are carried out separately. The theory does not allow for interactions between the verbal and the pictorial mental model during the processes of construction. In addition, the theory does not exactly explain how the integration of a descriptive and a depictive mental model can be performed. Similarly in DCT, the mental

representation of verbal as well as pictorial information is initially carried out separately. Besides, DCT includes neither the top-down cognitive processes of information processing nor the notion of mental model construction. According to DCT, a representational unit of the verbal system may trigger the corresponding unit(s) of the nonverbal (i.e. the visual) system through the referential connection, and vice versa. Therefore, using texts and pictures together should facilitate learning because the verbal and the pictorial coding have an additive function in memory. However, DCT fails to explain why using texts and pictures together does not always prove to be beneficial for learning, and why different visual displays can lead to different learning results. In contrast, this can be elucidated by the integrative model of text and picture comprehension.

The study by Schnotz and Bannert (1999) clearly demonstrated that the way in which information is visualized influences the way in which learners form their mental model. In this study, subjects were asked to study a learning material concerned with time differences on the earth. The learning material comprised texts and diagrams. Two visualization forms were compared in terms of their effects on learning. One visualization form depicted the earth's surface as a "flying carpet" that moves along the time axis (see Figure 10), whereas another visualization form showed the earth as a sphere (or circle) that rotates within a shell of different time states (see Figure 11). The results showed that subjects who learned with the "flying-carpet" diagram outperformed subjects who learned with the "circle" diagram in time difference tasks. In contrast, subjects who learned with the "circle" diagram outperformed subjects who learned with the "flying-carpet" diagram in regard to circumnavigation tasks. These results were in line with the predictions of the integrative model of text and picture comprehension: 1) The mental representation of pictorial information is generated as a mental model which preserves the structure of the depicted information. Different forms of visualization lead to different mental models. 2) A mental model may facilitate the performance of one kind of task while impairing the performance of another kind of task. It depends on what information can be read from the mental model. "If a learner tries to solve a task and the respective model has an inappropriate structure, it is either impossible to read off the required information, or the model allows to read off different propositions that contradict each other." (Schnotz and Preuß, 1999: 149). As the study showed, the mental model constructed for the "circle" diagram was not appropriate for solving the time difference tasks, and the mental model built for the "flying-carpet" diagram was not suitable for solving the circumnavigation tasks.
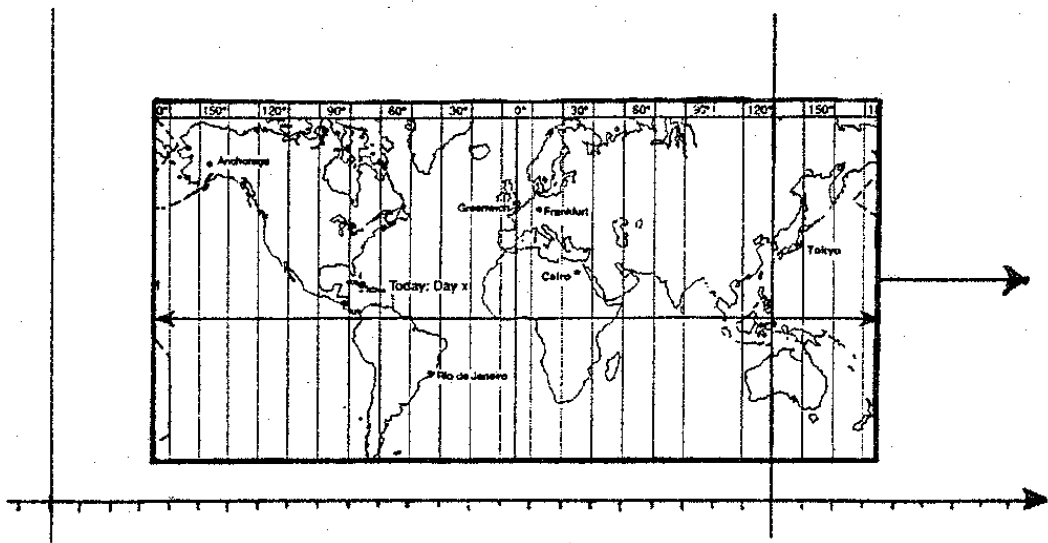
Figure 10: The "flying carpet" diagram
(Taken from Schnotz and Preuß, 1999: 145).
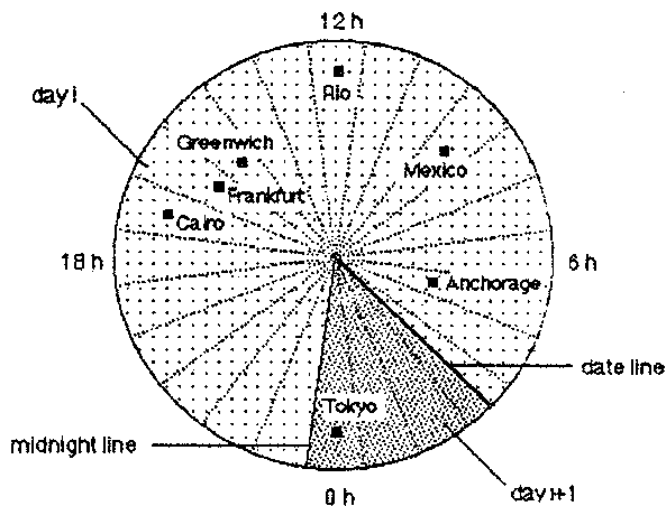
Figure 11: The "circle" diagram
(Taken from Schnotz and Preuß, 1999: 145).

In my opinion, the model proposed by Schnotz and Bannert (1999) provides a better explanation of text and picture comprehension than do DCT or Mayer's theory because the model is able to explain why different visualization forms have different effects on learning, and how knowledge is acquired and organized in learners' memory during the construction and elaboration of their mental models.

2.5 Summary

This chapter provides an overview of theories that are concerned with how verbal and pictorial information is represented in the human mind. According to the different theoretical approaches, mental representations of verbal and pictorial information can be conceived as propositional networks or mental models. The fundamental issue here is whether all kinds of information are represented as propositions or as modality-specific codes. The propositional approach claims that there are only propositional representations in human memory. However, the debate on mental imagery indicates that it is still unclear whether mental images are depictive or descriptive (propositional) in nature. Dual coding theory suggests that there are modality-specific mental representations that are interconnected via referential links between different subsystems in memory. The mental model approach tries to incorporate propositional, analogical, and schema approaches. However, the definitions of mental models still remain vague and are quite diverse in different research disciplines.

3. HUMAN MEMORY AND INFORMATION PROCESSING

The structures and the functions of human memory have been investigated for a long time in different disciplines such as psychology, neuropsychology, physiology as well as cognitive science. In spite of the long-standing research and the employment of sophisticated instruments for measuring mental activities, scientists to date have not been able to unravel all the mysteries concerning human memory. There have been many different approaches to modeling human memory with respect to the types of information that is stored, the way in which information is perceived, encoded, stored, and retrieved, or simply the duration of information storage in memory. In this chapter, I shall briefly introduce the structure and the different categories of human memory while mainly concentrating on the structure of working memory and elucidating its functions in information processing.

3.1 Categories of human memory

Human memory systems have been categorized in many different ways. Various conceptual dichotomies such as long-term/short-term memory (Miller, 1956; Brown, 1958; Peterson & Peterson, 1959; Broadbent, 1958; Atkinson and Shiffrin, 1968; Cowan, 1988, 2000), primary/secondary memory (Waugh and Norman, 1965), semantic/episodic memory (Tulving, 1972, 1983, 1993, 1995, 2001), implicit/explicit memory (Graf and Schacter, 1985; Schacter, 1987; Graf, 1994; Duffy, 1997), and declarative/nondeclarative memory (Squire, 1987; Squire and Zola, 1996; Eichenbaum, 1997), etc. have been used to classify human memory. Tulving (1995) has combined these dichotomies and proposed a more general scheme of human memory, which contains at least five major categories of memory system, whereby each system may contain several subsystems (see Table 1).

      Procedural memory (or nondeclarative memory) involves knowledge about how to perform an action. "The operations of procedural memory are expressed in the form of skilled behavioral and cognitive procedures independently of any cognition" (Tulving, 1995: 840). The skillful performance of many motor or non-noetic tasks such as driving a car, riding a bike, or getting dressed and so on relies mainly on procedural memory. **Priming** is a kind of perceptual learning, which

is expressed in enhanced re-identification of objects that one has already encountered before. "A perceptual encounter with an object on one occasion primes or facilitates

System	Other terms	Subsystems	Retrieval
Procedural	Nondeclarative	Motor skills Cognitive skills Simple conditioning Simple associative learning	Implicit
Perceptual representation system	Priming	Structural description Visual word form Auditory word form	Implicit
Semantic	Generic Factual knowledge	Spatial Relational	Implicit
Primary	Working Short-term	Visual Auditory	Explicit
Episodic	Personal Autobiographical Event memory		Explicit

Table 1: Major categories of human learning and memory (Tulving, 1995: 841)

the perception of the same or a similar object on a subsequent occasion, in the sense that the identification of the object requires less stimulus information or occurs more quickly than it does in the absence of priming" (Tulving, 1995: 841). Priming and procedural memory are essential to human beings in order to survive in their environment. According to Tulving, these two memories are probably phylogenetically very early forms which also develop early in human infants.

      Primary memory is commonly referred to as short-term memory or working memory. It is responsible for the registration, organization, and storage of incoming information for a short period of time. **Semantic memory** refers to the factual information or the general knowledge of the world. It represents our knowledge system and enables "cognitive operations on the aspects of the world beyond the reach of immediate perception" (Tulving, 1995: 841). Semantic memory and episodic memory are sometimes termed declarative memory or propositional

memory. **Episodic memory** contains recollections of personally experienced events in the past. According to Tulving (1995: 841), "Episodic memory enables individuals to remember experienced events as embedded in a matrix of other personal happenings in subjective time. It depends on but also transcends the range of the capabilities of semantic memory." Conscious awareness, which is also referred to as auto-noetic consciousness, plays an important role in describing the memory of past happenings.

The differentiation between **implicit** and **explicit memory**, according to Markowitsch (1999), does not refer to different memory systems but to different forms of memory or the way in which these two kinds of memory are retrieved. Implicit memory refers to the expression of what a person has learned without necessarily recollecting when, how, or where learning had occurred. In contrast, explicit memory specifies the expression of what the person is consciously aware of as a personal experience (Douglas and Brian, 1992; Tulving, 1995). The retrieval of procedural, priming, and semantic memories are based on implicit operations while that of primary and episodic memories is carried out by means of explicit operations.

To account for the relationship between memory systems and memory processes (encoding, storage, and retrieval), Tulving has proposed the SPI model. SPI stands for serial, parallel, and independent; encoding is serial, storage is parallel, and retrieval can be independent. This model assumes that the relations among different systems are process specific. "Different systems are dependent on one another in the operations of interpreting, encoding, and initial storing of information. Once encoding has been completed, different kinds of information about the same initial event are held in various systems in parallel… Access to different kinds of information about a given event is possible in any system independently of what happens in other systems" (Tulving, 1995: 844).

Aside from the memory categories proposed by Tulving, there are some other important categories of human memory, which are based on different criteria such as the duration of retaining information or the way in which information is received. Regarding memory as a function of duration, we can subdivide memory into a **short-term memory** system and a **long-term memory** system. The former can retain information only for a short period of time (a matter of seconds or, at most, a few minutes) and the latter can store information permanently. In considering the function of memory as a system for storing and retrieving information received through our senses, we can classify the memory systems according to different senses (**visual** or **auditory memory**, for example).

The storage of visual and auditory information in human memory has been well investigated in a number of laboratory experiments, while the storage of

olfactory, tactile, or gustatory information has not been well explored. However, it may be assumed that, for each kind of sensory information, there is a corresponding memory system. When people receive different sensory input from their environment simultaneously, different sensory registers work in parallel and hold information in the same form in which it is received for a few milliseconds. Neisser (1967) termed the memory systems responsible for storing visual and auditory input over a matter of milliseconds as **iconic** and **echoic memory** accordingly. Following Baddeley (1997), however, these two kinds of short-lived memories should be regarded as a part of the processes involved in perception. Both of the memories seem to be able to prolong the initial stimulus to enable later processing to be carried out by the corresponding short-term visual and auditory memory systems. Further processing involves the manipulation and the integration of the sensory-based information with information from other sources and the information stored in long-term memory. This should be performed by means of the working memory system (see Section 3.3). It should be noted that even with long-term storage, the representation of sensory-based information in memory may still retain sensory characteristics. Such memory involves the recollection of faces, scenes, voices and tunes.

The question of how human cognitive systems process and store information over shorter or longer periods of time is a core issue in any discussion on learning. Short-term memory, which is also termed working memory by some theorists, is particularly relevant for the acquisition of new information. A slight difference between short-term and working memory, however, is that short-term memory focuses on the duration of information storage while working memory focuses more on the processing of new information. In the following subsections, I shall describe these two memory systems in more detail.

3.2 Short-term memory

Based on our daily experiences, we all know that the amount of information we can keep in mind at a time is actually very limited. For instance, if you ask somebody's phone number or e-mail address, it is very difficult to remember it without writing it down. If you cannot take notes in this situation, you will probably try to repeat it a couple of times (aloud or sub-vocally) as a precaution against forgetting. This strategy is termed *rehearsal*. If rehearsal fails, the information will be wiped out from memory. Indeed, some information can be remembered only for a short period of time whereas other information is retained for life (your own name,

for example). The memory that lasts briefly is therefore termed short-term memory (STM), while its counterpart is termed long-term memory (LTM).

3.2.1 Debate on memory as a unitary or dichotomous system

There has been debate about whether it is necessary to regard STM and LTM as two separate memory systems. Theorists who support the view of memory as a dichotomous system have argued that information in STM can be rapidly forgotten if rehearsal is prevented. That is, without rehearsal, memory traces automatically fade away after a short period of time. The *trace decay* phenomenon was demonstrated in several classical studies (Brown, 1958; Peterson and Peterson, 1959). In one of these studies, Peterson and Peterson (1959) showed their subjects consonant trigrams such as 'khv', followed by numbers such as '567'. The subjects were asked to repeat the number and then to count backwards from it in threes (567, 564, 561, 558…) until receiving a signal, whereupon they should repeat the consonants. In this study, the rehearsal of consonants was prevented by the counting task, which led to rapid forgetting. Based on this result, Peterson and Peterson argued that short-term forgetting results from trace decay, whereas long-term forgetting results from interference. In the research literature, two types of interference have been discussed: *proactive interference* and *retroactive interference*. Proactive interference occurs when new learning is disrupted by old habits, while retroactive interference occurs the other way round. According to the interference theory, "forgetting reflects the disruption of the memory trace by other traces, with the degree of interference depending on the similarity of the two mutually interfering memory traces" (Baddeley, 1997: 32-33). The forgetting demonstrated in the study by Peterson and Peterson (1959) was not attributed to interference because remembering consonants and counting are quite different tasks. Consequently, they suggested that STM and LTM should be two separate systems.

However, some theorists have argued that STM should be regarded as a portion of LTM that is constantly activated (Melton, 1963; Anderson, 1983; Ericsson and Kintsch, 1995). That is, STM and LTM depend on the same unitary system. This view is based on the assumption that the loss of information in STM does not result from trace decay but from interference. In addition, a long-term learning effect can be observed in short-term memory tasks as well. In a task involving the immediate recall of sequences of random numbers, a sequence of random numbers was repeatedly shown in every three trials. Though subjects were not aware of this fact, the probability of recalling that particular sequence of numbers gradually increased. Melton's (1963) findings showed that LTM is also involved in short-term memory

tasks; however, that does not necessarily falsify the dichotomous-system view. Actually, several studies have yielded evidence against the unitary-system view. According to Baddeley (1995, 1997), the major arguments against the unitary-system view of memory are as follows:

1) In a free recall task, where subjects are asked to memorize a list of unrelated words, the first few items and the last few items are remembered particularly well, if recall is tested immediately (the primacy and recency effect). If recall is delayed, the recency effect disappears while recall of the first few items is unaffected. This result suggests that the recency items are held in a temporary short-term store, while earlier items are recalled from the long-term store (Glanzer and Cunitz, 1966; Glanzer, 1972).

2) When subjects attempt to recall strings of consonants, the errors they make are acoustically similar to the target items, which suggests that short-term storage is acoustically based (Sperling, 1960; Conrad, 1964). When recalling a group of words that are similar in either sound or meaning, the acoustic similarity effect still shows in immediate recall. However, after a filled delay, a semantic similarity effect occurs while the acoustic similarity effect disappears. This indicates that the short-term store relies on acoustic or phonological coding while the long-term store is more dependent on semantic codes (Baddeley, 1966a, 1966b).

3) Neuropsychological evidence indicates a double dissociation between the impairment of STM and LTM. There are patients whose STM is intact, while their LTM is grossly impaired (Milner, 1966; Baddeley and Warrington, 1970). Other patients, however, show the opposite impairment of memory (Shallice and Warrington, 1970). This is powerful evidence in favor of the view that STM and LTM are separate systems.

3.2.2 Measurement of short-term memory span

The main function of STM lies in temporary storage and manipulation of information. Since STM has a very limited capacity, many studies have investigated the constraints on information processing that result from these capacity limits. The seminal paper regarding the span of STM was written by Miller (1956), who concluded that *the span of absolute judgment* (the sensory channel capacity for making unidimensional judgments such as tones, auditory loudness, and saltiness, etc.) contains about 7 different stimuli, while the span of immediate memory comprises 7 plus or minus 2 chunks of information. Chunking is a strategy people use to group or organize information into familiar units or chunks, so that more

information can be remembered at a time. For example, a string of letters 'ACRAZYDRCAT' is difficult to remember, but it becomes easier if it is chunked into 'A CRAZY DR CAT'. Nevertheless, the range of a chunk remains unclear, which, in my view, renders any reasonable assessment of the absolute amount of information which can be held in STM impossible. Recently, Cowan (2000) has argued that, under stricter and better-controlled experimental conditions (where rehearsal and long-term memory cannot be used to combine stimulus items into chunks of an unknown size, nor can storage mechanisms that are not capacity-limited, such as sensory memory, allow the capacity-limited storage mechanism to be refilled during recall), the 'pure STM capacity limit' comprises only about 4 chunks on average. In my opinion, regardless of whether the span of STM is maintained to comprise 4 or 5 to 9 chunks, we should not take these 'magical numbers' too seriously. These numbers simply reflect the fact that the capacity of STM is limited. I question the applicability of such an approach because, in real life, the amount of information that can be held in STM at a time certainly depends on the characteristics of the information per se, on its integration into context, and how or to what extent chunking is employed as an information processing strategy. Certainly, all these factors will affect the magical number. Meanwhile, as long as the range of any single chunk of information cannot be determined precisely, magical numbers are of little help in finding out how much information can actually be held in STM.

3.2.3 Short-term memory model

STM has so far been characterized as a memory system that is separate from LTM and that has limited capacity for briefly storing information. Atkinson and Shiffrin (1968) proposed a memory model to further elucidate the function of STM storage in information processing (see Figure 1). According to this model, there are three types of memory store: sensory registers, a short-term store, and a long-term store. Information is first simultaneously processed by a number of sensory registers that forward information into the short-term store. The short-term store is regarded as a limited capacity working memory that is assumed to enable temporary storage and manipulation of information and to carry out some control processes such as rehearsal and retrieval from the long-term store. The longer an item is maintained in the short-term store, the greater the probability that it will be transferred into the long-term store. In other words, the more frequently an item is rehearsed, the more likely it will be recalled. It should be noted that information processing in this model is executed in a serial manner. The short-term store serves as a kind of 'gateway'

between the sensory registers and the long-term store. Without the short-term store, information cannot be transferred into or out of the long-term store.

```
┌─────────────────┐
│  Environmental  │
│      Input      │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│ Sensory registers│
│  ┌───────────┐  │
│  │  Visual   │  │
│  └───────────┘  │
│  ┌───────────┐  │
│  │ Auditory  │  │
│  └───────────┘  │
│  ┌───────────┐  │
│  │  Haptic   │  │
│  └───────────┘  │
└─────────────────┘
         │
         ▼
┌───────────────────────┐          ┌──────────────────┐
│   Short-term store     │          │                  │
│        (STS)           │ ───────▶ │ Response output  │
│ Temporary working memory│          │                  │
│ ┌───────────────────┐ │          └──────────────────┘
│ │ Control processes:│ │
│ │ Rehearsal         │ │
│ │ Coding            │ │
│ │ Decision          │ │
│ │ Retrieval strategies│ │
│ └───────────────────┘ │
└───────────────────────┘
       │    ▲
       ▼    │
┌─────────────────┐
│ Long-term store │
│      (LTS)      │
│    Permanent    │
│  memory store   │
└─────────────────┘
```
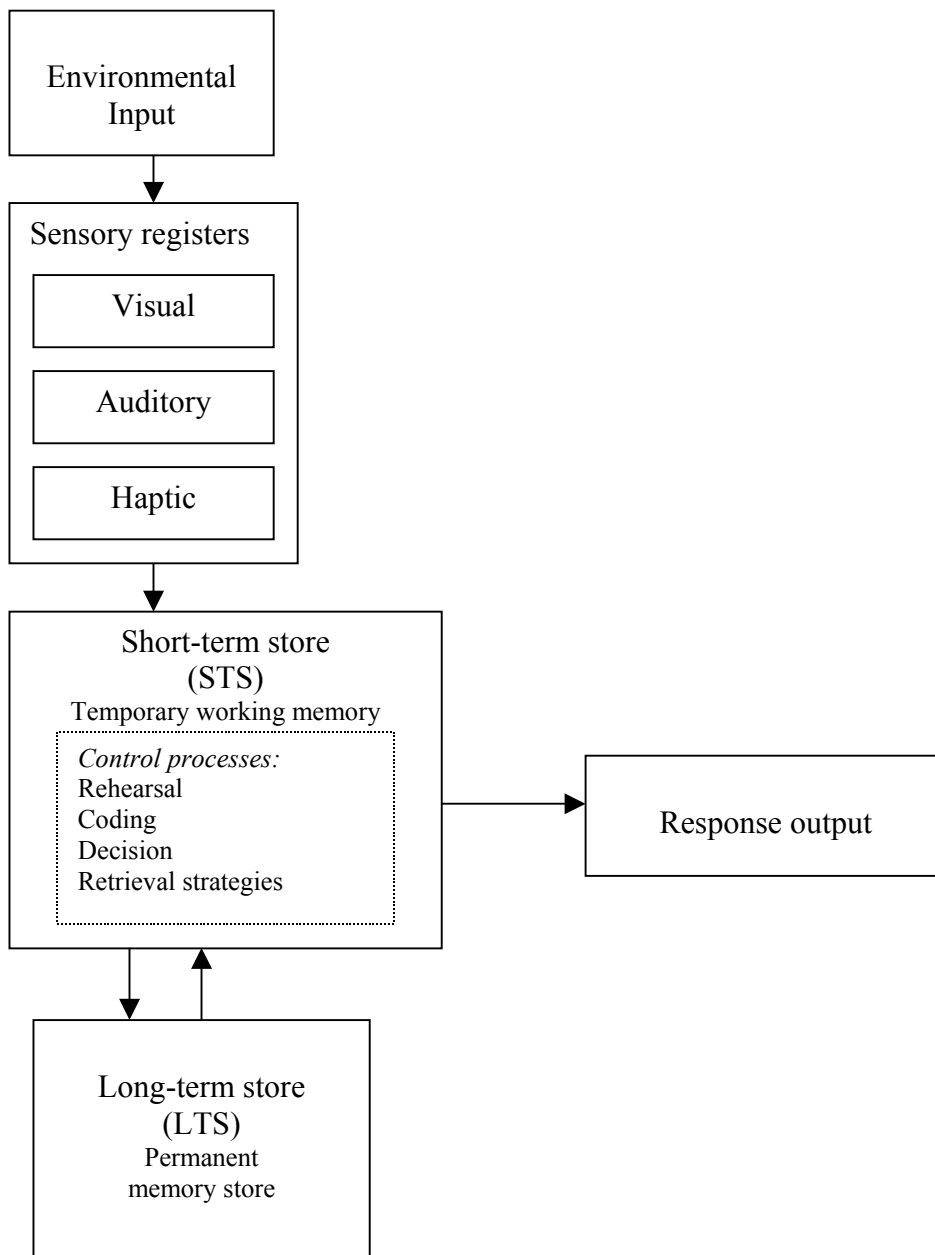
Figure 1: The short-term memory model of Atkinson and Shiffrin (1968)
(Taken from Baddeley, 1997: 44)

There is, however, evidence against Atkinson and Shiffrin's model. The assumption that long-term learning is contingent on the duration of maintaining information in the short-term store has proved to be poorly supported in a number of studies. In addition, the model is not able to explain the double dissociation between the impairment of STM and LTM, since any deficit in relation to STM will

inevitably lead to a deficit in relation to LTM as well. Furthermore, according to the model, the short-term store is regarded as a capacity limited unitary system. In the light of this assumption, a trade-off should be observed when performing a concurrent task. That is, the performance in relation to one task should be severely impaired by the other, since the two tasks are supposed to share the same resource from the unitary short-term store.

This assumption was challenged by Baddeley and his colleagues (Baddeley and Hitch, 1974; Baddeley, 1986). With a dual-task paradigm, the subjects were required to rehearse digits (ranging from 0 to 8 places in length), while concurrently performing a reasoning task (verifying a series of sentences such as A follows B – BA (true), B is not preceded by A – AB (false)). The results showed that the concurrent digit load did to some extent impair performance in the sentence-verification task. Nevertheless, this impairment, according to Baddeley, was far from dramatic. With a concurrent load of 8 digits, the subjects' performance did not deteriorate much, which provides evidence countering the predictions based on the view that working memory is a single unitary store. On the other hand, the accuracy of retrieval from LTM proved to be independent of the concurrent digit load (Baddeley et al, 1984). Based on these findings, Baddeley argued that "the limit of digit span may be set by one of a number of subsystems, leaving other components of working memory relatively unimpaired." Short-term memory, or working memory, according to Baddeley and Hitch's view (1974), should not be a unitary system but rather a multi-component system.

3.3 Working memory

The development of the concept of working memory has a long history (see Logie, 1996; Baddeley, 2001). The term 'working memory' was first used by Miller et al. (1960). It refers to the memory that is supposed to control the capacity of information processing. It has access to consciousness and is responsible for the temporary storage of information, for decision making, and for the execution and coordination of plans (cf. Richardson, 1996). Although working memory has been defined in many different ways, there is a common denominator among the various definitions: working memory plays a central role in information processing, and it has a limited capacity for temporary storage and manipulation of information.

The working memory model proposed by Baddeley and Hitch (1974), a simplified representation of which is given in Figure 2, assumes that working

memory consists of three components: a controlling attentional system (termed the central executive) that supervises and coordinates (at least) two subsidiary systems.

Visuo-spatial   **Phonological**

sketch pad   **loop**
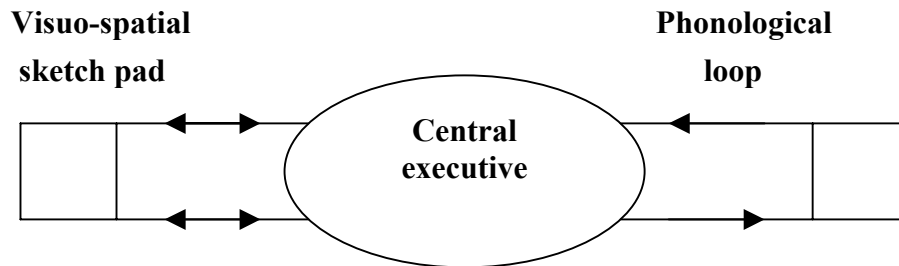
Central executive

Figure 2: The working memory model of Baddeley and Hitch (1974)

The functions of the two subsystems–the phonological loop and the visuo-spatial sketchpad–are well explored, while the function of the central executive is less clearly expounded. In the following subsections, I shall explain the functions of the two subsystems and discuss the role of the central executive.

3.3.1 The phonological loop

According to Baddeley (1997), the phonological loop consists of two components, *a phonological store* that is able to hold speech-based information for about 2 seconds, and *an articulatory control process* based on inner speech. The articulatory control process can transform printed material into phonological codes and refresh the information held in the phonological store by subvocal rehearsal. It is assumed that the phonological loop plays an important role in language comprehension, first or foreign language acquisition, acquiring vocabularies, and learning to read. Empirical evidence pertinent to the phonological loop is provided by a number of studies. The evidence can be summarized as follows (Baddeley, 1997):

1. The phonological similarity effect

    The phonological similarity effect shows in the impaired performance of immediate serial recall of items that are similar in terms of the sound or articulatory features. The impairment in recall is believed to result from interference based on phonological coding. However, whether the similarity effect occurs at the level of sound, phonemes, or articulatory commands remains unclear.

2. The unattended speech effect

    It is found that when visually presented numbers are accompanied by irrelevant speech (spoken in a foreign language), recall performance for the

numbers is worse than when no speech is presented. The same effect is also found when auditory words or nonsense syllables, are presented in the background, whereas presenting noise does not cause this effect. This suggests that the access to the phonological store by speech or speech-based sound is automatic, which therefore damages the recoding of visually presented digits.

3. The word-length effect

   The word length effect is manifested in the fact that immediate memory span is affected by the spoken duration of the words presented. Memory span is shorter when sequences of words have long vowels and are spoken slowly. Apparently, memory span reflects the number of items that can be uttered in about two seconds.

4. Articulatory suppression

   The operation of the phonological loop is disrupted by overt or covert repeated articulation of an irrelevant item. For example, when subjects are required to repeat the word *the* while trying to remember sequences of digits, the span of digits they can retain is reduced substantially. "This is assumed to occur because the articulation of an irrelevant item dominates the articulatory control process, hence preventing it from being used either to maintain material already in the phonological store, or convert visual material into a phonological code" (Baddeley, 1997: 58).

3.3.2 The visuo-spatial sketch pad

The visuo-spatial sketch pad (VSSP) is responsible for setting up and manipulating mental images. Similar to the phonological loop, the VSSP can be fed either directly through visual perception or indirectly through the generation of a visual image. There has been, however, debate on the characteristics of information stored in the VSSP. First, there is controversy regarding whether the mental representation of visual input is analogous to its original form or if it is a kind of propositional representation which is descriptive in nature. Second, the debate is concerned with whether the mental imagery has a visual or a spatial basis. Since the representational format issue has been discussed in Chapter 2, I shall focus on the issue concerned with the visual or spatial basis of representations here.

Brooks (1967) conducted a pioneering study to tackle this issue. He presented subjects with a 4×4 matrix and specified one of the squares as the starting square. The subjects were then asked to repeat back sequences of sentences (see Figure 3).

		3	4
	1	2	5
		7	6
		8	

Spatial material	**Nonsense material**
In the starting square put a 1.	In the starting square put a 1.
In the next square to the *right* put a 2.	In the next square to the *quick* put a 2.
In the next square *up* put a 3.	In the next square to the *good* put a 3.
In the next square to the *right* put a 4.	In the next square to the *quick* put a 4.
In the next square *down* put a 5.	In the next square to the *bad* put a 5.
In the next square *down* put a 6.	In the next square to the *bad* put a 6.
In the next square to the *left* put a 7.	In the next square to the *slow* put a 7.
In the next square *down* put an 8.	In the next square to the *bad* put an 8.

Figure 3: Example of stimulus material used in Brooks' experiments on the
visuo-spatial sketchpad (Taken from Baddeley, 1997: 75)

There were two conditions: under one condition, sentences were used that required some imagery-based strategy, while under the other condition the sentences required rote verbal rehearsal. Under the spatial condition, words such as 'left', 'right', 'up', and 'down' were used in the instruction, whereas under the non-spatial condition, those words were substituted accordingly by 'slow', 'quick', 'good', and 'bad' to ensure that the encoding was carried out by verbal rehearsal. The sentences were presented in either an auditory or visual manner. The results showed that auditory presentation was helpful for the spatial condition, while visual presentation was beneficial for the non-spatial condition. Brooks assumed that sentences remembered on the basis of visual imagery use the same resource for visual perception, while sentences maintained by rote verbal rehearsal employed the same systems that are also used in auditory perception.

In a further experiment, Baddeley (1975, 1980) used the dual-task method to examine whether the processing strategy evoked by the Brooks matrix task is visual or spatial in nature. He combined the Brooks task with either a pure visual task that involved judging the brightness of a large screen, or with a pure spatial task that involved tracking the position of a pendulum via auditory signals. The results clearly showed that the spatial matrix task was mostly impaired by the tracking task, whereas the verbal matrix task was mostly disrupted by the visual task, which indicates that the processing strategies elicited by the Brooks task rely on spatial encoding rather than on visual encoding.

There is another line of evidence indicating that the VSSP probably comprises two separate components: one visual and one spatial. Neurological research investigating the function of a monkey's brain suggests a "what" and a "where" pathway in the monkey's visual system. The human brain is believed to be structured in a similar way (see Figure 4). The "what" pathway involves a set of brain areas going from the primary visual area (the occipital lobe) into the temporal lobe, which is responsible for recognizing particular visual patterns, shape, color, and speed of movement. The "where" pathway includes the brain areas going from the primary visual area into the parietal lobes. Those areas are concerned with the location of an object in the visual world (Posner and Raichle, 1999). Moreover, neuropsychological evidence shows double dissociation between the impairment associated with processing visual and spatial information. For example, there are patients who are able to recognize objects, but cannot locate them, whereas some other patients are capable of localizing an object, but not of recognizing it.

According to the evidence mentioned above, it seems plausible to consider the VSSP as a system comprising two separate components. In my opinion, it is still hard to say whether the visual imagery is visual or spatial in nature. The strategies elicited by the Brooks task, for example, could be primarily spatial because the construction of the imagery relies on the spatial configuration. Nevertheless, it is hard to believe that there was no visual component involved in this process at all.

"where" pathway
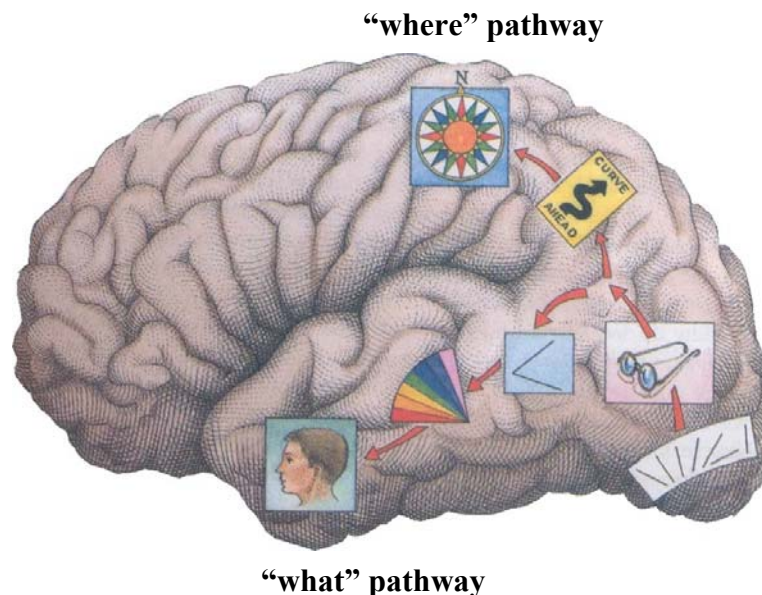
"what" pathway

Figure 4: The "what" and "where" pathways in the visual system
(Taken from Posner and Raichle, 1999: 15)

On the other hand, the mental image stored in the VSSP at the end could be visual (i.e. a visualized matrix pattern or a curve going through the eight positions in the matrix). Therefore, mental imagery in some cases is actually both visual and spatial. In addition, I would like to point out that we should be careful about interpreting the results of experiments employing the dual-task method, since the characteristics as well as the appropriateness of the concurrent task determine the degree of interference with the primary task. This in turn will certainly influence our judgment on the nature of mental imagery. Thus, results of the dual-task method will sometimes be inconclusive. In light of the Brooks task, does the fact that the brightness judgment task did not disrupt the spatial matrix task really exclude the possibility that the mental imagery could have a visual basis? I personally doubt it. Indeed, some theorists (including Baddeley) also recognize this problem.

Finally, the role of the VSSP in verbal learning is examined. According to the dual coding theory of Paivio (1969, 1986, 1991), concrete words, in contrast to abstract words, are easier to remember because they are highly imaginable. If the words to be learned can be encoded both visually and verbally, the recall of those words will be better than if the words are encoded only verbally. Visual imagery mnemonics is the method employed to enhance learning and memory by creating mental images to represent the words and imagining those images interacting. Baddeley et al. (1975) investigated whether the VSSP is necessary for setting up images when subjects use visual imagery mnemonics to learn a list of words. If so, a concurrent spatial disrupting task should reduce the positive effect of highly imaginable words on recall performance. Their results indicated that the concurrent spatial task impaired the recall of both concrete and abstract words to roughly the same degree. Baddeley et al. concluded that the imaginability effect is not dependent on setting up an image in the VSSP, but rather is contingent upon long-term semantic memory.

Logie (1986) demonstrated the unattended picture effect on learning word lists using either a visual imagery strategy or a verbal rehearsal strategy. He showed his subjects colored patches appearing at regular intervals on a screen. The subjects were asked to learn word lists while looking at the screen but ignoring whatever they saw. The results showed that there was a significant decrement in learning performance in the case of the subjects who used a visual imagery strategy, but there was no such effect for the subjects who used a verbal rehearsal strategy. This suggests that visual information has obligatory access to the VSSP, which disrupts the use of a visual imagery strategy.

To sum up, the VSSP is assumed to set up and manipulate depictional mental representations. This system controls the use of visual imagery strategies but is not

responsible for the imaginability effect on recall. Evidence from neuropsychological studies suggests separate visual and spatial components of imagery (cf. Baddeley, 1997).

3.3.3 The central executive

The central executive was first postulated by Baddeley and Hitch (1974) as an attentional controller that supervises and coordinates the subsidiary systems in working memory. The function of the central executive was, however, not very well explored at that time; it was only characterized as "a limited capacity pool of general processing resources" (Baddeley, 2001: 855). Later, Baddeley (1986) adopted a concept from the SAS (supervisory attentional system) model proposed by Norman and Shallice (1986) (see Figure 5). According to the model, human actions are controlled by a series of schemata and habits, so that well-learned skills or routine tasks such as driving a car are carried out automatically. The SAS, a capacity limited attentional system, comes into play when a conflict occurs between the automatic action plan and the stimuli in the environment. It is assumed that the SAS is able to plan a solution by combining information from long-term memory with existing stimuli. The evidence of the SAS is supported by the data shown as absentmindedness in normal subjects and the disturbance of attentional control in patients with frontal lobe deficits.
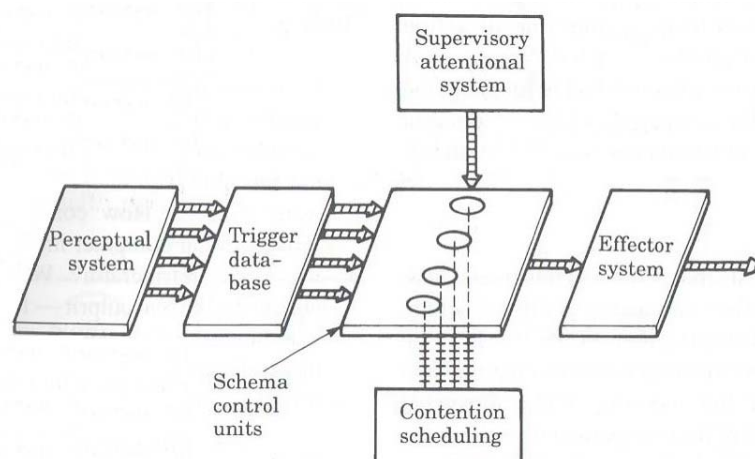
Figure 5: A simplified version of the Norman and Shallice (1986) model

The concept of the SAS used as the basis of the central executive function has been further improved by Baddeley and his colleagues (Baddeley, 1996). In an attempt to fractionate the central executive, which was only vaguely conceived of as

a capacity limited attentional controller, they conducted a series of studies to examine the executive function of normal subjects and patients with Alzheimer's disease or dysexecutive syndromes. In the experiments, a dual-task paradigm was employed where subjects had to perform a pursuit tracking task, which relies on the visuo-spatial sketchpad, plus a digit span task, which depends on the phonological loop. It turned out that in comparison to normal subjects including elderly and young people, the performance of the patients with Alzheimer's disease was significantly worse. However, when the two tasks were performed independently, there was no evidence that those patients' performance was differently affected by increasing task difficulty. Since Alzheimer patients are characterized by an impairment of episodic memory and attentional deficits, the difference in single and dual task performance suggests a separable executive capacity to allocate attention to coordinate two tasks (see Baddeley et al., 1991, 2001; Logie et al., 2000). It is important to bear in mind that, with normal subjects, a performance decrement in dual tasks is quite normal; it indicates that the capacity to coordinate information from the slave systems is limited (Bourke et al., 1996; Baddeley, 1997).

In addition to the capacity to coordinate two tasks, Baddeley (1996, 2001) suggested that the central executive also has the capacity to: 1) switch retrieval strategies used for the random generation task; 2) to selectively focus attention on one stimulus and inhibit the disrupting effect of others, which is based on the assumption that "anything that limited attentional capacity would impair performance" (Baddeley, 2001: 856); and 3) to hold and manipulate information from long-term memory, which is reflected in the working memory span (Daneman and Carpenter, 1980). Yet it remains an open question whether the central executive should be regarded as "a single coordinated system that serves multiple functions, a true executive, or a cluster of largely autonomous control processes—an executive committee" (Baddeley, 1996: 26). In a recent article, Baddeley (2001) has suggested that the central executive involves the capacity to focus attention as well as the capacity to divide and switch attention (Baddeley et al., 2001), but it has no capacity for storing information (Baddeley and Logie, 1999) and might not be much involved in retrieval from long-term memory (Baddeley et al., 1984; Craik et al., 1996). As to the last assumption, retrieval from LTM was tested by means of a dual-task experiment. Subjects were required to perform a demanding secondary task while learning or retrieving lists of words. The concurrent load from the secondary task did affect learning, but it had little effect on recall.

In Baddeley's working memory model (1986), the central executive is supposed to be capable of combining the information from working memory with that from LTM. Yet the functions of the central executive described so far have not

yielded any concrete information about how the central executive and LTM interact. There are some other phenomena that cannot be well explained by the current model provided that the central executive is purely an attention system without any storage capacity of its own (Baddeley, 1996; Baddeley and Logie, 1999). Firstly, some patients with dense amnesia have been able to use a chunking strategy to perform well in a task involving immediate recall of prose. Secondly, the model does not explain how and where the central executive combines the verbal and the visual information from the two subsystems. To overcome these problems, Baddeley (2000) has modified his model by adding a fourth component— *the episodic buffer*—to the working memory system (see Figure 6).

The episodic buffer is assumed to be the place where information from the subsystems of working memory and that from LTM are integrated. "It is assumed to be episodic in the sense that it holds integrated episodes or scenes and to be a buffer in providing a limited capacity interface between systems using different codes" (Baddeley, 2001: 858). The integration of different codes from the two subsystems and from LTM is explained by assuming that the buffer uses a kind of common code. It is capable of chunking information and storing it coherently in a multimodal fashion. Moreover, the episodic buffer is supposed to depend heavily on the central executive because there is no direct link between the buffer and the phonological loop as well as the visuo-spatial sketchpad. The integration of information from the two subsystems and that from LTM is still mainly controlled by the central executive. In my opinion, the buffer does exist because it provides *a workspace* where the integration of information as well as the influence of LTM in the process of information processing (e.g., chunking) may take place and also provides *a temporary storage* for the integrated information. With the assumption of an episodic buffer, the problems of the previous model can be solved.

It is further assumed that information is retrieved from the buffer through conscious awareness. "This allows multiple sources of information to be considered simultaneously, creating a model of the environment that may be manipulated to solve problems and plan future behavior" (Baddeley, 2001: 858). The episodic buffer, in my view, takes over some functions that previously had been implicitly ascribed to the central executive. Therefore, the buffer may actually be regarded as a fraction of the central executive which carries out the information processing but leaves other executive functions to the central executive. "The executive is now assumed to be a purely attentional system whose role extends beyond memory function, whereas the episodic buffer is assumed to be purely mnemonic in character" (Baddeley, 2001: 858).
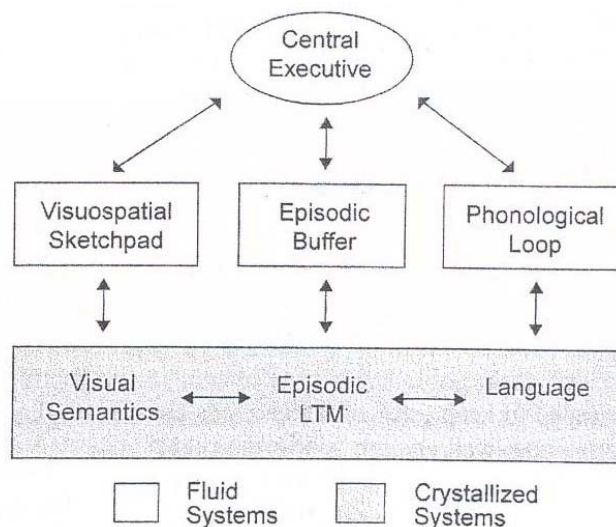
Figure 6: The revised working memory model
(Taken from Baddeley, 2000: 421)

Another noteworthy update to Baddeley's working memory model is that there are direct links between the phonological loop and the long-term verbal memory as well as between the visuo-spatial sketchpad and the long-term visual memory. Evidence that supports this assumption is based on studies which show that knowledge or previous experience stored in LTM is also involved in the processing of information within the subsystems of working memory. For example, non-words that resemble English in their phonotactic structure are easier to remember than those differing from English (Baddeley, 1996; Adams and Gathercole, 1995; Baddeley et al., 1998), which suggests that the working of the phonological loop is sensible to LTM retrieval. Moreover, the links between the two subsystems and LTM indicate that the transmission of information between them can be carried out directly without a bypass through the episodic buffer or the central executive. That is, the interface function between working memory and LTM is not exclusively taken over by the central executive as suggested in the previous model, nor is it solely controlled by the episodic buffer.

As an addendum to my outline of Baddeley's model of working memory, I would like to point out that there are several alternative accounts of working memory to be found in the literature (see Miyake and Shah, 1999). However, I have concentrated on Baddeley's model because it is widely accepted, well supported both by psychological and by neurophysiological evidence, and is relevant for the purpose of my study. Its recent developments emphasize the role of LTM in information processing, a point that has largely been ignored before. In my opinion, the greatest challenge in memory research is to find the boundary between working memory and

LTM. This is an issue that needs to be solved in the future. In the following section, I shall address the role of LTM in learning.

3.4 Long-term memory and learning

LTM refers to the memory system which holds information for a lifetime. Information stored in LTM comprises diverse kinds of knowledge, which we have acquired and keep on acquiring, such as semantic knowledge, episodic knowledge, and procedural knowledge. However, LTM does not only serve as a knowledge base where information can be stored and retrieved, but it also plays a crucial role in information processing. In fact, the knowledge stored in LTM is organized in a particular way, which may influence how we acquire new knowledge. In the following, I shall introduce the most important approaches to semantic memory and elucidate the role of LTM in learning.

3.4.1 Models of semantic memory

3.4.1.1 Network models

Several classical empirical studies have shown that when subjects were asked to learn a list of words and then recall as many words as they could, they did not recall the words according to their order in the list, but tended to recall the words clustered by content, for example man-woman, bread-butter, etc. Moreover, subjects' recall performance was enhanced when the words to be learned were organized in categories (Jenkins and Russel, 1952; Deese, 1959; Tulving, 1962; Bower et al., 1969). Research on semantic memory was in the beginning focused on investigating how the concepts of single words are structured. An example of the classical approach to semantic memory is *the hierarchical model* proposed by Collins and Quillian (1969; 1972). The model assumes that the concepts in LTM are organized in a hierarchical network consisting of nodes and links, with each node corresponding to a concept, and each link designating the relationship between two nodes (see Figure 7). For example, the concept LIVING BEING is associated with subordinate categories ANIMAL and PLANT. The concept ANIMAL in turn is associated with mammal, and bird, etc., while mammal is associated with cat, dog, cow, etc. Empirical evidence supports this assumption; people can verify the sentence "A robin is a bird" more quickly than the sentence "A robin is an animal,"

conceivably because the pathway between 'robin' and 'bird' is shorter than that between 'robin' and 'animal' in the network. However, evidence to the contrary can also be found. Firstly, people can verify "A robin is a bird" faster than verifying "A penguin is a bird." The hierarchical model is not able to explain why more typical instances of a category are verified more quickly than the less typical ones. Secondly, "A chicken is an animal" is verified more quickly than "A chicken is a bird." The hierarchical model is not able to explain why some levels in the network seem to be more accessible than others. Thirdly, it takes longer to falsify "A bat is a bird" than to falsify "A bat is a plant," though the pathway between 'bat' and 'bird' is shorter than that between 'bat' and 'plant'. As a result, the concepts in LTM are not organized in a strictly hierarchical fashion.
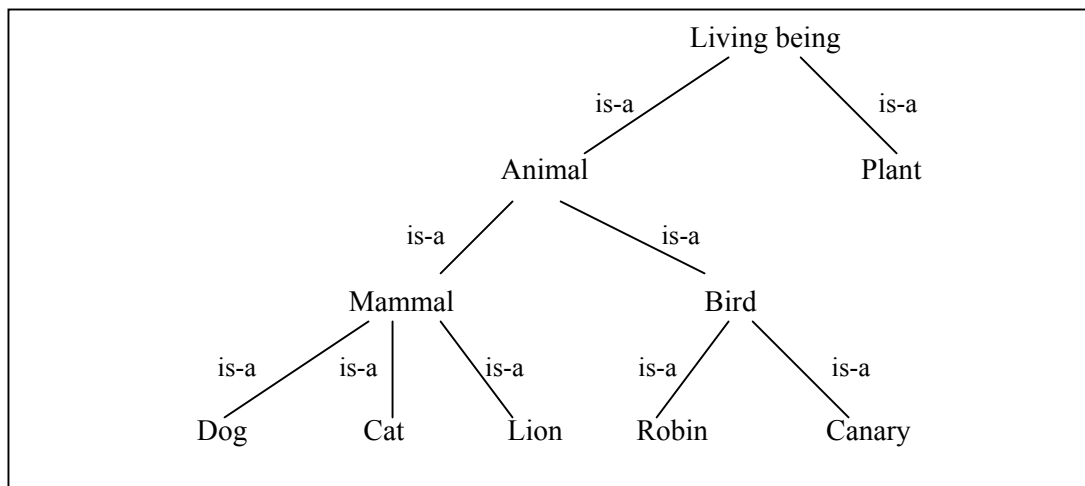
Figure 7: Hierarchical network model (after Douglas & Brian, 1992: 221)

An alternative model that attempts to solve the problems just mentioned is *the feature comparison model* of Smith, Shoben, and Rips (1974). According to this model, a concept consists of a set of semantic features. The essential features of the concept are called defining features while others that are less defining but generally true are called characteristic features. It is assumed that the more similar two instances in a category are, the more characteristic features they share. Although the feature comparison model is able to account for the phenomena that are incompatible with the hierarchical model, a central problem of this model lies in how those features are defined. Moreover, the model postulates two stages of decision processes when judging whether an item (e.g., a robin) belongs to a category (e.g., a bird). The first stage is supposed to roughly compare all the features without considering how defining the features are. The judgment is dependent upon the degree of the feature overlap. If the feature overlap is intermediate, a second stage is required to compare the defining features. In my opinion, it is difficult to define all the semantic features as well as to set up the defining features of a concept. I also believe that it is not

plausible that the defining features are not compared in the first stage. Since they are essential to a concept, they should be included in the first stage of comparison.

Collins and Loftus (1975) pro-posed a refined network model, which is quite flexible and allows for different types of relations between concepts in comparison to the hierarchical model (see Figure 8). In this revised model, there is no strict hierarchy in the network. Instead, memory is determined by the interconnection between the concepts. Apart from the 'is-a' links, there are some other types of links such as 'is-not-a', 'can', and 'can-not'. On the other hand, the links also differ in terms of 'the importance'. The most crucial links are supposed to be traversed more quickly, which can explain why the response to the prototype of a category is faster than that to the atypical instances.
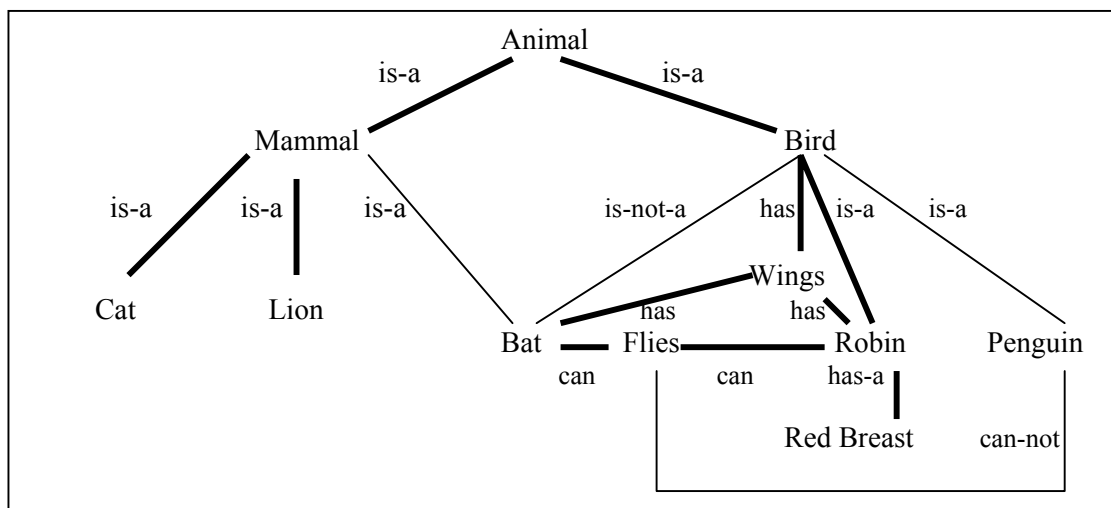
Figure 8: An example of a spreading-activation network model
(based on Douglas & Brian, 1992: 229)

Through these alterations, the effect of semantic relatedness, which refers to the prolonged reaction time when determining the relationship between two concepts sharing many common features, is explained by a *spreading-activation* search mechanism. It assumes that when two concepts are activated, the activation spreads throughout the network until the two initial concepts are linked. There are several intersections where the initial activation meets in the network. Evidence from all these intersections is then summed until a threshold is reached to give a response. It takes more time to falsify the sentence "A bat is a bird" because 'bat' and 'bird' have some features in common, which leads to more intersections (with positive or negative evidence) they need to be evaluated, and hence, require more time. In addition, spreading activation can account for the priming effect of the lexical decision tasks. When subjects are asked to decide whether the letter strings are

words, the reaction time of a letter string (e.g., doctor) is shorter if a semantically-related word (e.g., nurse) has been shown in the previous trial. This is because the first word has partially activated the concepts related to the second word through its spreading activation, which in turn facilitates the activation of the second word. As we can see, the spreading activation network model is able to explain a wide range of data, and it also allows the use of prior knowledge and the computation of information that was not stored (Douglas and Brian, 1992). Nevertheless, the model has been criticized for being very complex and for neglecting the interaction between the concepts and the real world (Baddeley, 1997).

In the last decade, powerful network models have been based on different principles as the more traditional semantic networks have been developed. These models are termed *PDP* (*parallel distributed processing*) or *connectionism models* (McClelland and Rumelhart, 1986; McClelland et al., 1986). In a connectionist network, the nodes or units are connected by weighted links. The strength of the flow of activation from the given unit to another one is a function of its weight. A simple example of a PDP network is depicted by Figure 9.
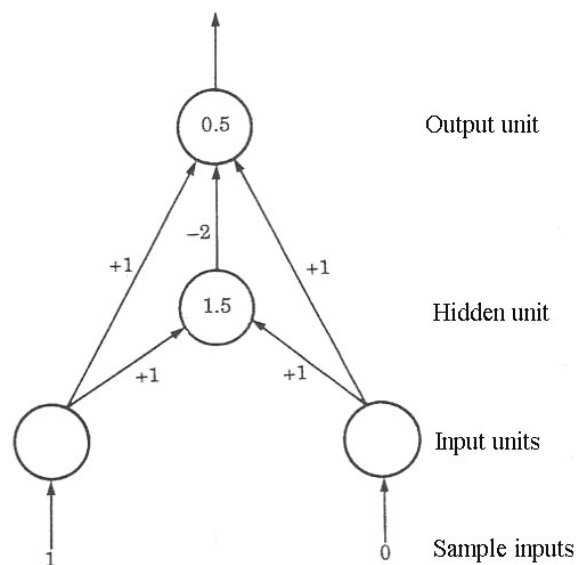
Figure 9: An example of a PDP network with a hidden unit
(Rumelhart and McClelland, 1986; taken from Baddeley, 1997: 260)

In contrast to semantic network models, a concept in a connectionist network model does not correspond to a particular node but is represented as a pattern of activation over the whole network. That is, concepts are distributed throughout the network, i.e. encoded as a set of connection weights. The advantages of connectionist networks are as follows: 1) The network allows partial inputs to retrieve partial outputs. 2) The network is capable of learning. When an error occurs, the network

can adjust the connection weights to produce the correct output (backward error propagation). 3) The same set of weights can be used for remembering specific information and learning abstractions (Douglas & Brian, 1992). All in all, the connectionist approach is more powerful than the semantic networks approach in modeling human memory and learning because of the parallel distributed processing mechanism. However, the development of connectionism does not ensure progress in our understanding of how the human brain processes information. As a matter of fact, there is a basic problem concerning cognitive adequacy. From an engineering point of view, as long as the network can produce the correct output, it does not matter whether the way in which the network operates is consistent with the way in which the human brain operates. From the viewpoint of cognitive science, however, it most certainly does (Baddeley, 1997). Baddeley (1997: 272) suggested that in the future "it will probably be necessary to blend connectionist approaches with more rule-based models, using the empirical methods of experimental psychology to evaluate and shape such developments."

3.4.1.2 Schema theories

There is another family of theoretical approaches which follows from the assumption that semantic memory comprises structures that are much more comprehensive than the simple concepts proposed by network models. These approaches suggest that people remember information in terms of existing structures which is termed schema. According to Bartlett (1932), "A schema is a structure that organizes large amounts of information into a meaningful system… A schema is a stereotype specifying a standard pattern or sequence of steps associated with a particular concept, skill, or event. Schemata are types of plans we learn and use during our environmental interactions" (Schunk, 1996: 168). As I have mentioned in Chapter 2, the study by Bartlett (1932) showed that when people tried to recall the story, they often distorted or ignored the parts of the story that were not compatible with their past experiences. This indicates that people actively use the schemata stored in their memory to reorganize or reconstruct the events (effort after meaning).

An essential notion of schema was proposed by Piaget (1952). In Piaget's view, a schema is "a completely coordinated set of physical actions and cognitive functions, a set that worked together in its totality to respond to every perceived experience that could be related to the schema." (Piaget, 1952: 237). Schemata are assumed to develop only for the situations, events or patterns that occur repeatedly. Two functions are ascribed to schemata: 1) Assimilation: The new experience is changed to fit the schema and its altered features are then incorporated into the

schema. 2) Accommodation: When assimilation fails, the schema has to adapt itself to accommodate the situation it is trying to assimilate. In the course of learning, new information is checked against a schema, which may be specified, extended or modified to accommodate the new information.

Both Bartlett's and Piaget's notions of schema still lack specificity in terms of what schemata exactly contain and how they are exactly developed and structured. Minsky (1975) suggested that a useful schema theory can only be established if the following issues have been addressed: 1) how people select an appropriate schema from their memory to deal with a given situation; 2) how schemata are interrelated to each other and are retrieved as needed; 3) how a schema is modified and created; and 4) how the memory store changes as a result of learning (cf. Marshall, 1995). None of these issues have been fully addressed to date. However, schema theories play an important role in regard to learning and memory.

For example, schema theories explain why experts can more efficiently acquire new knowledge related to the domain in which they specialized than can novices. The reason for the difference between experts and novices lies in the amount of their prior domain knowledge (or schemata).

> "Perhaps the largest source of individual differences in memory performance is difference in knowledge in a particular domain. It is much easier to remember something if we have a framework in which to embed that new knowledge… There is clear evidence that the ability to acquire new facts about some domain depends a great deal on what one already knows. For example, Spilich, Vesonder, Chiesi and Voss (1979) found that people who knew more about baseball were much better able to remember descriptions of baseball games…. For the facts that were not relevant to the game, the groups showed no difference in recall. Thus, having prior knowledge allows one to understand (and remember) the relevant information better…, this previous knowledge allows one to interpret new information more easily to make it meaningful, to incorporate it into what one already knows, and to retrieve it easily using prior retrieval schemes." (Douglas & Brian, 1992: 208-209).

Ausubel (1963, 1968, 1977) also pointed out that learning is more effective when new information bears a relation to knowledge in memory. Hence, the amount and the accessibility of prior knowledge in terms of established schemata should influence learning results and learning efficiency.

3.4.2 Implications for learning

The models of semantic memory all suggest that the semantic knowledge in LTM has a well-organized structure, which is to be described not as a strict hierarchy, but probably more as a network of interrelated concepts. The empirical evidence supports the ideas that well organized material is easier to learn than poorly-organized material, and that learners tend to organize the learning materials in terms of their pre-existing knowledge. Basically, the strategies people employ to enhance

learning mainly follow the principle that the information to be learned should be accommodated into their schematic world knowledge. Once a connection between the new and the old information is established, the probability of retaining the new information is increased. However, if there is no apparent link between the new and the old information, inferential and elaborative strategies can be employed to enhance recall. Mnemonics, for example, is a strategy for improving memory performance by associating the information to be remembered with something familiar, for instance a visual image or a verbal feature such as rhyme or rhythm. Moreover, the advantage of associative learning is supported by Paivio's (1991) dual-coding theory, which suggests that information can be better remembered if it is encoded both verbally and visually because activating one form of a concept in memory will activate the corresponding one as well. The spreading-activation network models excellently simulate the idea that the more connections between a given concept and the others, the more likely it is that the concept will be recalled.

Taken together, the implications for learning are, in my view, the following:

1. The basic principle of learning (meaningful material) is accommodating the new information under pre-existent schemata.

2. The pre-existent schematic knowledge of experts is qualitatively and quantitatively different from that of novices.

3. Materials that are organized in such a way that their structure can easily be mapped to the structure of the relevant schemata (of a particular learner) are easy to learn.

4. Poorly organized materials require additional processing such as inferences and elaboration to enable schema consumption; these processes make for additional cognitive workload.

5. Modality plays a major role in these additional processing requirements because working memory is a capacity-limited, modality-specific system.

3.5 Summary

This chapter began by introducing different categories of human memory and the characteristics of each category of memory. Various theories of how information is processed in human memory were subsequently presented. There has been debate on whether working memory and long-term memory are two separable memory systems. Evidence from experimental psychology and neuropsychological studies is in favor of a separation between the two systems. The processing of new information in working memory was mainly discussed with respect to Baddeley's working

memory model. With respect to learning, both working memory and long-term memory determine the performance of learning. In light of the way in which knowledge is stored and organized in long-term memory, it is suggested that associative and organized learning are more effective.

4. LEARNING WITH TEXTS AND PICTURES

In Chapter 2 and Chapter 3, respectively, I explained the theories concerned about mental representations of verbal and pictorial information and how human memory systems encode, store, and retrieve multicodal and multimodal information. When putting those theories into actual educational practice, however, these theories were not always found to be true. For instance, based on the dual coding theory, presenting pictures together with texts should be beneficial for learning, but this has not been a consistent finding. Therefore, it is necessary to examine under which conditions the theories are applicable, and to discern why the theories do not always hold true. In this chapter, I shall deal with the issue of learning with text and pictures. Principles and reasons concerning how, which, when, and why pictures facilitate learning will be explained.

According to the dual coding theory (Paivio, 1986, 1991) as well as the integrative model of text and picture comprehension (Schnotz and Bannert, 1999), presenting pictures together with text can facilitate learning because pictures help learners to construct mental models that are essential for comprehending the information to be learned and thereby enlarge the retrieval possibility of this information. This assumption was confirmed in many empirical studies. However, in some studies pictures did not prove to be beneficial. The reason for this lies in that the effects of pictures on learning are influenced by many other factors such as the nature of the information to be learned, the instructional method, the way in which text and pictures are presented, the learner characteristics, etc. In addition, the method for assessing learning outcomes should be considered because, depending on the assessment method, the same pictures might or might not have a positive effect on learning.

Mayer (1993) explained learners' cognitive processes in processing learning material with the help of Atkinson and Shiffrin's (1968) memory model (cf. Chapter 3). He postulated four essential cognitive processes of learning in human memory systems: selecting, organizing, integrating, and encoding (see Figure 1). Selecting refers to paying attention to the relevant information in the instruction. Organizing establishes the relationship between relevant information pieces. Integrating relates the incoming information with knowledge in long-term memory. Finally, encoding refers to the process of storing the new information in long-term memory. Furthermore, Mayer proposed a research framework comprising some other factors that affect learning from text and pictures (see Figure 2). In the following

subsections, I shall address the issue—when, why, and how pictures facilitate learning—according to this framework.
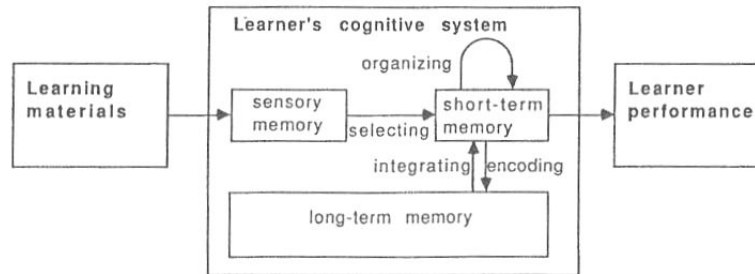
Figure 1: The cognitive approach to research on learning from text and
pictures (Taken from Mayer, 1993: 260)
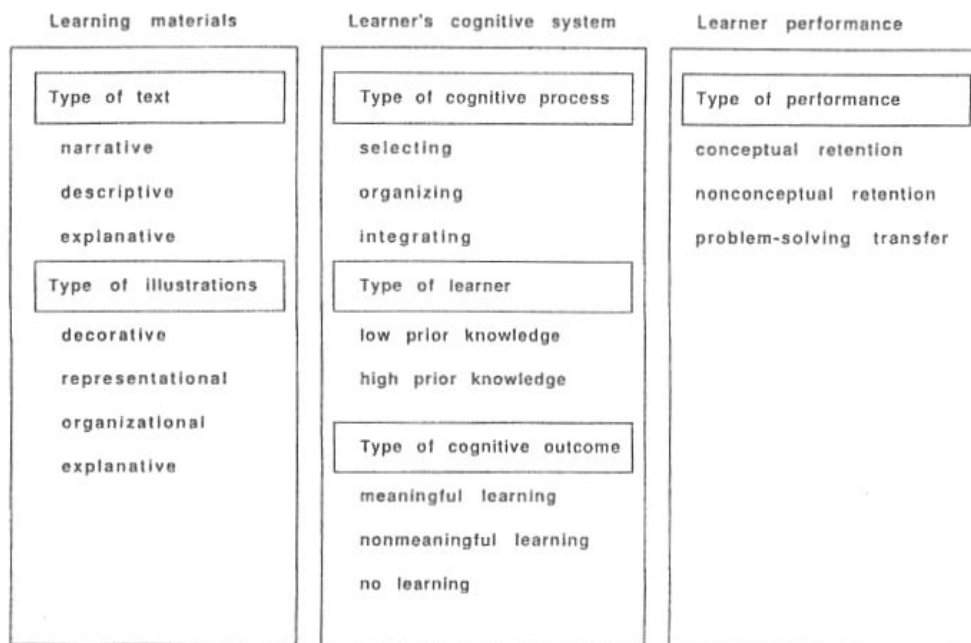
Figure 2: A framework for research on learning from text and pictures
(Taken from Mayer, 1993: 264)

4.1 Types and functions of pictures

The functions of pictures are contingent upon the type of text they accompany. Pictures used in storybooks, according to Fang (1996), may help readers to "1) establish the setting; 2) define/develop the characters; 3) extend/develop the plot; 4)

provide a different viewpoint; 5) contribute to the text's coherence; and 6) reinforce the text." (Carney and Levin, 2002: 6). In the context of education, pictures in textbooks serving in text processing are categorized by Levin (1981, Levin et al., 1987) as follows:

- *Decoration*: Pictures are simply used to decorate the page. They have little or nothing to do with the text content.
- *Representation*: Pictures are employed to represent part or all of the text content, which is the most frequently used picture type.
- *Organization*: Pictures give a structural framework of the text content (e.g., a map).
- *Interpretation*: Pictures elucidate the difficult text content.
- *Transformation*: Pictures serve as mnemonic aids.

Mayer (1993) proposed four types of pictures, which are basically derived from Levin's categorization. He also investigated the percentage of use of these picture types and how these pictures influence learners' cognitive processes (see Table 1). It should be noted that Mayer's classification only pertains to pictures in textbooks serving to teach scientific concepts, for example, how a car's braking system or a bicycle tire pump works. Therefore, the functions described here are not applicable to the pictures used to depict concepts other than this kind.

Type of illustrations	Percentage of surveyed illustrations	Definition	Cognitive processes affected
Decorative	23%	fill space on page without enhancing message of passage	none
Representational	62%	portray a single element	selecting
Organizational	5%	depict relations among elements	selecting & organizing
Explanative	10%	explain how a system works	selecting, organizing & integrating

Table 1. Four types of illustrations (Taken from Mayer, 1993: 265)

Peeck (1993) suggested that pictures are able to decorate the page, excite the learner, explain difficult concepts, expand the written narrative, and affect intellectual skills and processes. Finally, Carney and Levin (2002) regarded pictures as adjunct aids for the processing of text information. They help learners to perceive, understand, and remember the information to be learned.

4.1.1 Static versus dynamic visual displays

When considering the presentation modes of pictures, we can roughly classify them into two categories: static and moving pictures, or in broad terms: static and dynamic visual displays. The former term refers to still graphics such as diagrams, charts, illustrations, etc. that do not involve any motion, whereas the latter term refers to the displays of any type of pictorial and graphical movement including "continuous motion as in films, television, and interactive video discs, graphical animations in computer-based instruction (CBI), such as simulations of chemical processes, electronic behaviors, and pop-in arrows, labels, etc." (Park and Hopkins, 1993: 427).

Theoretical assumptions concerning the superiority of dynamic visual display (DVD) over static visual display (SVD) are based on that DVDs can direct learners' attention to the relevant information in the instructions, and that "DVDs provide a convenient means of visually representing imaginable events, actions, and ideas which change over time. They can make complex cognitive tasks more concrete and easy to understand by providing motion and trajectory attributes." (Park and Hopkins, 1993: 430). In contrast, SVDs are less efficient in presenting the processes of dynamic events changing over time. Consequently, it is assumed that "representing concepts and tasks involving motion with animation triggers the student's automatic ability of the visual system to induce apparent motion and directly encode them into the imaginal subsystem of memory, while static representation requires the student's ability and effort to form mental images of the task's dynamic nature by connecting and integrating discretely presented information" (Rieber and Kini, 1991, cited in Park and Hopkins, 1993: 430).

Some empirical studies demonstrated the superiority of DVDs in conveying instructions in procedural tasks such as knot tying or machine gun disassembling (Silverman, 1958; Roshal, 1961; Spangenberg, 1973). However, some other studies could not find any positive effects of DVDs employed to instruct procedural skills. Park and Hopkins (1993) reviewed 27 studies that investigated the effects of DVDs versus those of SVDs and only found 15 of them demonstrated the benefits of DVDs. Their explanation for the inconsistencies was two-fold. First, the DVDs did not successfully direct learners' attention to the relevant information. Second, individual differences between learners such as prior knowledge, experience, learner aptitude, etc. determined whether DVDs were useful or not (more details about this point are given in Section 4.4). Based on the review of research findings regarding the effects of DVDs, they recommended using DVDs in relation to the following conditions: 1) demonstrating sequential actions in a procedural task; 2) simulating causal relationships among the components of complex systems; 3) demonstrating visually

invisible system functions and behaviors such as blood flow in the human body or current in electronic systems; 4) providing a visually motional cue, analogy or guidance to show time-dependent processes; and 5) directing attention to the essential information to be learned (cf. Park and Hopkins, 1993: 444).

4.2 How and why pictures facilitate learning

With the description of the functions of pictures, the reasons why pictures may facilitate learning are already partially explained. Here I shall summarize the various reasons that I found in the literature. Pictures may facilitate learning because they can:

- help learners to focus their attention on relevant information in the learning material (Mayer, 1989, 1993; Levin and Mayer, 1993);
- motivate learners to study the accompanying text (Peeck, 1993; Schnotz, 2002);
- induce more elaborate processing of text information depicted by the illustrations (Peeck, 1987, 1993);
- help to clarify and interpret difficult text content (Bransford and Johnson, 1972; Bock, 1978);
- reduce interference and/or decay in memory of the material concerned and facilitate retrieval (Rusted, 1984);
- help to construct mental models (Weidenmann, 1988; Hegarty and Just, 1989; Mayer and Gallini, 1990; Glenberg and Langston, 1992; Shah et al., 1999);
- help to build connection between verbal and nonverbal information and therefore increase the retrieval potential for the illustrated text content (Paivio, 1986; Mayer and Anderson, 1991);
- present information in a more concrete and concise way (Levin and Mayer, 1993);
- help to organize the text information (Levin, 1981, Levin et al, 1987; Mayer, 1993);
- serve as mnemonic aids to enhance the memory of the learned material (Levin, 1981);
- effectively increase recall when they depict the type of information for which the text itself invites processing (Waddill et al., 1988; Waddill and McDaniel, 1992).

4.3 When pictures help learning

As I mentioned earlier, pictures are not always useful for learning in practice because whether or not pictures foster learning is contingent upon many factors that interact with each other. According to Mayer (1993), pictures can only exert a positive effect on learning when three conditions are met. First, the type of text and pictures must match each other well. Second, they must be used to instruct the learners who can benefit from the support of pictures. Third, an appropriate test must be employed to assess learning performance. It goes without saying that both text and pictures must be of good quality in presenting and explaining the information to be learned. Explanative text, for example, requires explanative pictures. Other types of pictures such as decorative, representational or organizational pictures are certainly not appropriate for this purpose, and therefore no positive effect can be expected. Learner characteristics as well as the assessment method, on the other hand, are quite important factors that I would like to address separately.

4.4 Learner characteristics

4.4.1 Prior knowledge

The first learner factor that should be considered is the amount of prior knowledge. In Chapter 3, I have explained the role of prior knowledge in learning. The more knowledge a learner has about the relevant domain, the more efficiently he or she can process new information without resorting to pictorial aids. Consequently, it is assumed that learners with high prior knowledge might not benefit from pictures as much as learners with little prior knowledge because as experts, unlike novices, they can easily connect the new information to their prior knowledge and construct appropriate mental models simply on the basis of text information.

This assumption was confirmed by the study of Mayer and Gallini (1990). In a series of three experiments, students were required to learn how a brake system, a pump system, and electric generators work. The instructions were given by means of four different conditions: only text, text plus static illustrations of the device with labels for each part, text plus static illustrations of the device with labels for each major action, and dynamic illustrations (depicting the 'on' and 'off' states of the device) with labels for each part and each major action. Students' learning outcomes were assessed on the basis of conceptual recall, non-conceptual recall, problem-solving tests, and verbatim recall. The results showed that learners with low prior

knowledge were sensitive to the instructional conditions. Their performance in problem-solving transfer was enhanced when the instruction was conveyed by means of dynamic illustrations with labels for parts and actions. Learners with high prior knowledge, in contrast, were not sensitive to the instructional conditions in terms of their performance in problem-solving transfer.

A study by Kunz et al. (1989, reported in Najjar, 1996) found that there was a positive correlation between the use of pictures in text and the comprehension level of students with low prior knowledge. Similar results were found in Mayer and Sims' study (1994). They argued that "students who possess domain-specific knowledge may not need visual aids with text because they can generate their own familiar analogical representations as they read or listen to an explanation of a system" (p. 400). Rieber and Kini (1995) pointed out that novices are often poor in allocating their attentional resources and in organizing the information appropriately. Accordingly, pictures could help them to focus attention on the relevant information and to organize the material properly (Bennett and Dwyer, 1994).

ChanLin (1998) investigated the effects of different visual treatments (no graphics, static graphics, and animated graphics) on acquiring procedural and descriptive knowledge with students having either a high or low prior knowledge level. The results showed that high prior knowledge (HPK) students outperformed low prior knowledge (LPK) students in acquiring both descriptive knowledge and procedural knowledge. In terms of learning descriptive knowledge (e.g., facts about the basic unit of growth hormone gene and its composition), the effect of visual treatments on the performance of LPK students was not significant. However, they tended to learn better with graphics (either static or animated) than without graphics. HPK students, on the other hand, only benefited from animation. When acquiring procedural knowledge (e.g., processes for building a polypeptide chain), LPK students who received different visual treatments did not differ in their performance, while HPK students learned better only when they received static graphics. This study provides evidence against the assumption that LPK learners stand to benefit more from visual aids than HPK learners and against the counterevidence that animation stands to facilitate the comprehension of a procedural text more than a descriptive text (Large et al., 1996, reported in ChanLin, 1998).

In my view, the graphics or animation most likely did not help LPK students in learning because the information to be processed (especially in terms of the procedural knowledge) was very complicated per se. Consequently, LPK students' performance could not simply be improved by any visual treatment (Large, 1996). The unexpected result that HPK students acquired descriptive knowledge more easily with animation and procedural knowledge more easily with static graphics could be

explained by the different nature of animation used in this study. ChanLin assumed that the animation for conveying descriptive knowledge served mainly as a mnemonic device to enhance memorization. Although it only demonstrated the information described in the text without any explanatory function, it still helped HPK students to better remember the information. The animation for depicting procedural knowledge could be redundant for HPK students, and thus, it might interfere with memorization. Another possibility is that the pace of the animation might not correspond to HPK students' internal pace for processing the information depicted in the animation, which could be an obstacle to learning. In contrast, static graphics allow students to process the information at a variable pace, and thereby facilitate learning.

4.4.2 Spatial ability

In addition to prior knowledge, learners' spatial ability may also determine whether they can benefit from pictures. Spatial ability refers to the capacity to process visual information in general. It has been further defined in various ways. In the broadest sense, spatial ability involves "spatial sense, spatial perception, visual imagery, spatial visualization, visual skills, spatial reasoning, mental rotations, and visual processes" (ChanLin, 2000: 230). Spatial visualization is the ability to mentally manipulate objects in two or three dimensions and to imagine the changing configurations of objects stemming from this manipulation (Mayer and Sims, 1994), which is supposed to be an important indicator of conceptual performance in scientific and mathematical learning (ChanLin, 2000). According to Sternberg (1990), spatial visualization concerns the ability to visualize shapes, the rotation of objects, and how pieces of a puzzle would fit together (cited in Mayer & Sims, 1994). High spatial-ability (HSA) people are able to mentally perform visual-spatial operations easily and rapidly while low spatial-ability (LSA) people cannot.

In relation to learning materials containing text and pictures, HSA learners are generally assumed to benefit more from the supplement of visualization because they are capable of extracting the meaning of pictures. On the contrary, LSA learners would find it difficult to understand the pictorial information. This assumption was confirmed in a study by Mayer and Sims (1994). When animation was presented simultaneously with a narration explaining either how a bicycle tire pump or the human respiratory system works, only HSA students could benefit from this type of presentation condition but LSA students could not. Mayer and Sims explained this result on the basis of a dual-coding theory (see Figure 3). When visual and verbal information is in working memory at the same time, HSA learners can devote more

cognitive resources to establish referential connections between propositional and depictional representations of the state of affairs conveyed in the learning material, while LSA learners must allocate more cognitive resources to building representational connections between the pictorial information and its mental representation. Hegarty and Sims (1994) found similar evidence indicating that visual representations seem to be more helpful to subjects with relatively good visualization.
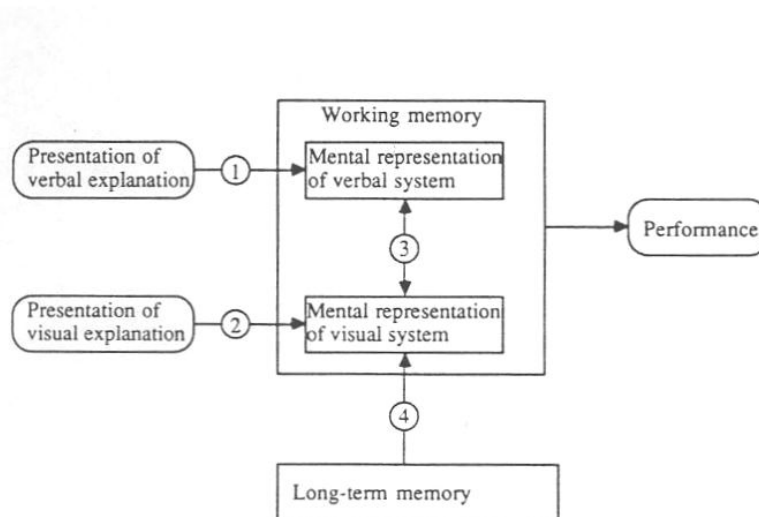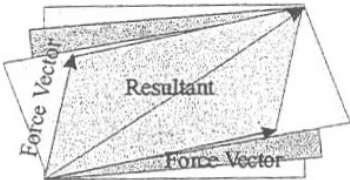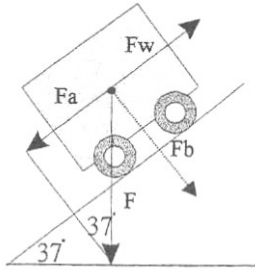
Figure 3: The dual-coding model of multimedia learning
(Taken from Mayer and Sims, 1994: 390)

Nevertheless, counterevidence was found in a study by Hays (1996). He examined whether different graphic presentations (no graphics, static graphics, and animated graphics) exert a different effect on the comprehension of learners with LSA and of learners with HSA. The results indicated that, in terms of short-term comprehension, LSA subjects receiving animation in the instruction performed better than did other LSA subjects receiving other graphic presentation-conditions. The level of their short-term comprehension was as high as that of the HSA subjects who received animation in the instruction. Long-term comprehension, however, was independent of the subjects' spatial ability. This indicates that LSA subjects can benefit as much from animation as can HSA subjects.

ChanLin (2000) investigated how different visual treatments (no graphics, static graphics, or animation) combined with (visual) text influence students with HSA or LSA in acquiring descriptive and procedural knowledge in the field of physics. The descriptive knowledge consisted of a recital of facts or the description of objects or events. An example of the learning material concerning descriptive knowledge is depicted in Figure 4. The procedural knowledge was concerned with

the construction of the problem-solving procedures. Figure 5 provides an example of this learning material. The subjects' grade level, as well as prior physics and mathematics scores were used as covariates.

"Force Vector" refers to the amount of force with a direction. When two forces are acting on an object, a single force vector then forms, which is called "Resultant". One resultant can be formed from many different pairs of force vectors. Thus, if you make a parallelgram out of it, you could say that the resultant is equivalent to the diagonal and the force vectors become the sides.
 	
Figure 4. Example of using graphics for descriptive content (taken from ChanLin, 2000: 232)
	When a cart weighing 20 kg is pushed upward along a slanted surface with an angle of 37° from horizontal, how much force (Fw) is needed to move the cart upward? (Suppose the abrasion from the slanted surface is 0).
	Rule 1: When the neighboring sides of a parallelogram are perpendicular to each other, the parallelogram is rectangular.
	Rule 2: When the angles of a triangle are 37°, 53°, and 90°, then the ratios of the sides opposite those angles will be 3:4:5.
	
	Solution:
	Step 1: The resultant F (20kg) can be formed by the following pair of force vectors:
	  F_a is paralleled downward to the slant surface.
	  F_b is perpendicular to the slant surface.
	Step 2: F is 20 kg, and the angle formed by F_a and the resultant F is 53° (90°-37°) (According to rule 1 and step 1)
	Step 3: Apply rule 2 to get F_a
	  F_a: F_b: F = 3: 4: 5,
	  F_a = 3/5F = 3/5 * 20 = 12 kg
	Step 4: If Fw is more than 12 kg, then you will be able to pull the cart upward.
	
	Figure 5. Example of using graphics for procedural content (taken from ChanLin, 2000: 232)

The results showed that the visual treatments only had a significant effect on the acquisition of procedural knowledge but not on the acquisition of descriptive knowledge. Students who were shown graphics or animation outperformed those who had received only text. Spatial ability, however, did not have a significant effect on acquiring descriptive or procedural knowledge in this study. ChanLin assumed that the effects of spatial ability might be tempered by factors within the training and instructional setting. "The use and training of various spatial strategies also serves as

a function of gaining experiences of solving problems in descriptive and procedural knowledge" (ChanLin, 2000: 236).

Separate analyses were carried out to examine the effects of visual treatments among students within each level of spatial ability. The results showed that, when acquiring descriptive knowledge, HSA students who received text and animation performed better than other HSA students who received only text in the instruction, whereas the performance of LSA students was independent of visual treatments. In terms of the acquisition of procedural knowledge, LSA students who received static graphics outperformed the other LSA students who received only text in the instruction, whereas HSA students were insensitive to the visual treatments.

All in all, the relationship between spatial ability and learning performance seems to be inconclusive. In my view, spatial ability alone might not be a reliable indicator for predicting a learner's ability to benefit from visual aids because of interactions with some other factors such as the amount of prior knowledge, the instructional method, the way in which text and pictures are presented, or even the method used for assessing performance. Mayer and Sims (1994), for example, suggested that low-experience, HSA students are the most likely to benefit from animation synchronized with a narration. Hays (1996) pointed out that people differ in the use of their spatial ability to create internal visualization. "If one does not use spatial abilities to create spatial visualization and form spatial representations of a particular concept, then one will only have a verbal/linear representation of the concept" (p. 149), which may not be sufficient for a complete comprehension. Young students might not know when and how to best use their spatial ability, while HSA learners might not use their spatial ability to transfer the perceived information into long-term memory. Peeck (1993) indicated that "merely asking or telling learners to pay attention to pictures in illustrated text is unlikely to induce more, or more intensive picture inspection than what would otherwise occur on the basis of the nature of the text, pictures, learner characteristics, and their interaction" (p. 233). Consequently, it is necessary to assume that people do not automatically use their spatial ability to process pictorial information, unless they are appropriately instructed and well controlled by learning activities (Peeck, 1993).

In my view, although we can not use spatial ability to predict learners' performance, spatial ability still affects the efficiency in learning pictorial information to a certain extent. As long as the effect of spatial ability is significant, we should take this factor into account if we intend to investigate the effects of multimedia presentations on learning.

4.4.3 Other characteristics

Some theorists assume that *reading ability or comprehension skill* may determine the extent to which illustrations promote a person's learning performance. However, the theories at this point are inconsistent. One view is that poor readers are less capable of processing textual information. Illustrations might help them to comprehend the information to be learned. Another view is that poor readers might have difficulty in finding relevant information in the pictures and in integrating information from text and pictures (Peeck, 1993). It appears that whether or not a learner can benefit from visual aids is based on his/her capacity to extract meaning from pictures, an ability commonly termed *visual literacy*. Beyond reading ability, *age* itself is assumed to be correlated with one's visual literacy. Young children seem to be less adept, in comparison to adults, at directing attention to the essential parts of pictures, and they are less efficient and systematic in inspecting pictures (Peeck, 1987). Moreover, Dretzke (1993) found that "…with increasing age it becomes more difficult for adults of relatively lower verbal ability to process effectively the information presented in complex pictorial interaction." (Dretzke, 1993: 499, cited in Carney and Levin, 2002).

4.5 Measurement and assessment methods

There are different types of tests that can be employed to measure learning outcomes, such as retention tests (including immediate recall, delayed recall, free recall, and cued recall), multiple-choice tests, operational or problem-solving tests, and testing in the visual mode. Peeck (1993) suggested that testing should not only be limited to a certain type of test but should also incorporate different testing procedures in order to assess the effect of using visual aids more thoroughly.

As to the nature of learning outcomes, there are various dimensions to be assessed. In general, the assessment, according to Levin (1989), involves understanding, remembering, and applying. In other words, it must measure how well learners have comprehended and remembered the information they learned, as well as the learners' ability to apply the new information to solve problems. In a way, similar to Levin, Mayer (1993) suggested that learning performance can be measured in terms of the recall of the main elements and relations in the text (i.e. conceptual retention) as well as the recall of isolated facts or recognition of the verbatim wording of sentences (i.e. non-conceptual retention), or the answers to open-ended questions that require inferences (i.e. problem-solving transfer). Since different

methods may yield different patterns of results, which in turn can influence the interpretation of the role of text and pictures in learning, it is recommended that researchers exercise caution when choosing appropriate methods for testing learning performance.

Learner characteristics and assessment methods are the factors that can have an effect on learning outcomes that is independent of the effect of how learning material is presented. When examining the effects of multimedia presentations on learning, such factors must be considered in order to eliminate their disruptive effects.

4.6 Summary

This chapter has dealt with the issue of learning with texts and pictures. I began by introducing different types of pictures in terms of their functions as well as the relevant presentation modes. Furthermore, I have summarized the findings from the literature concerning how, why, and when pictures foster learning. Based on the framework for research on learning from texts and pictures proposed by Mayer (1993) in conducting research on this topic, I have elucidated the role of learner characteristics and methods for assessing learning performance. Since a number of factors in parallel can determine whether using pictures when giving instruction is beneficial to learning, researchers should take all of these factors into account in order to investigate the pure effects of using pictures on learning.

5. COGNITIVE ASPECTS OF PROCESSING MULTIMEDIA INFORMATION

In this chapter, I will introduce the current theories of multimedia learning which place emphasis on cognitive processes, specifically *the cognitive load theory* (Sweller et al., 1990; Sweller and Chandler, 1991, 1992) and *the cognitive theory of multimedia learning* (Mayer, 1993, 1997; Moreno and Mayer, 2000; Mayer and Moreno, 2002a, 2002b). As we shall see, these theories are mainly derived from the dual coding theory and from theories of human memory. Since this thesis focuses on the effects of multimodal presentations on processing multicodal information, a series of studies that have investigated these effects will be discussed in the light of the two theories. However, due to the diversity of the factors that determine the effectiveness of multimedia-based learning, the current theories warrant further study.

5.1 Cognitive load theory

The cognitive load theory (CLT) proposed by Sweller and his co-workers (Sweller et al., 1990; Chandler and Sweller, 1991; Sweller, 1993, 1994; Sweller et al., 1998) is an instructional theory that considers information processing in human memory as a crucial determinant of the effectiveness of learning. The main assumption of CLT is that since new information must be processed in working memory before it can be stored in long-term memory and since the capacity of working memory is limited, learning will be impaired if the information-processing capacity of working memory is overstrained. It is therefore advisable to design instructional materials in such a way that they can be processed within the capacity of the learner's working memory.

CLT suggests a series of measures that can be implemented by instructional designers to promote learning. For example, the use of *goal free problems* and *worked examples* rather than traditional problem-solving tasks may reduce working memory load and facilitate *schema acquisition* and *automation* by learners (Sweller, 1988; Sweller, 1993; Marcus et al., 1996). Since these recommendations do not directly deal with multimedia learning, I shall not address them further. Instead, I will introduce different types of cognitive loads and the various measures used to eliminate them.

5.1.1 Levels and types of cognitive load

It is assumed that the level of cognitive load is determined by causal as well as assessment factors (see Figure 1) (Paas and van Merrienboer, 1993, Kirschner, 2002). Causal factors can be learner characteristics (e.g., cognitive abilities), task or environmental demands (e.g., task complexity or noise from the environment), and/or interactions between the two. Assessment factors consist of three measurable aspects: mental load, mental effort, and performance. Mental load is engendered by the task and environmental demands. Mental effort refers to the cognitive exertion learners devote to the task. Learners' performance is affected by mental load, mental effort, and causal factors.
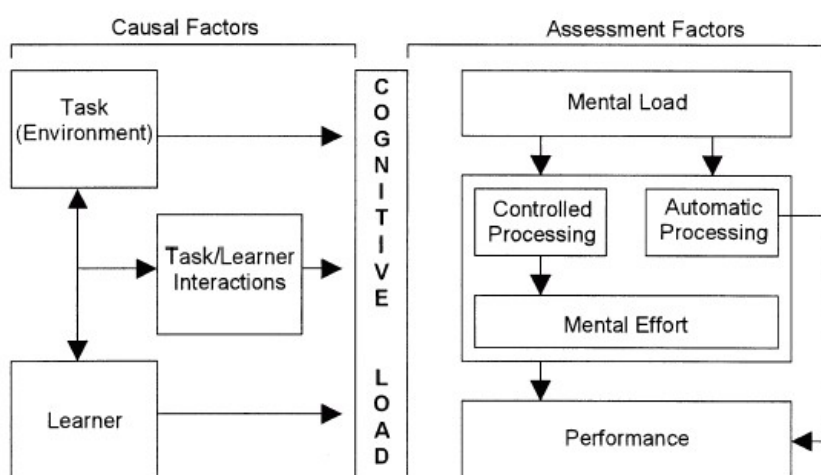
Figure 1: Factors determining cognitive load
(Taken from Kirschner, 2000: 4)

CLT discriminates between three types of cognitive load induced by instruction: *intrinsic cognitive load, germane cognitive load, and extraneous cognitive load*. The intrinsic cognitive load refers to the inherent complexity of the information to be processed, or the degree of interconnection among the elements of information. It is assumed that complex information involves a high level of element interactivity, whereby a single element of information cannot be learned independent of other elements. The germane cognitive load results from the construction of schemata and their storage in long-term memory. The extraneous cognitive load is imposed on the recipient by the manner in which information is presented (Sweller and Chandler, 1994; Sweller et al., 1998; Kirschner, 2002). Recently, Valcke (2002) has proposed a subset of germane cognitive load, termed meta-cognitive load, which

is concerned with the meta-cognitive monitoring abilities of one's own learning processes (see Figure 2). In addition, Valcke has pointed out that prior knowledge should be taken into account in CLT because it influences the efficiency of schema acquisition and the ability to engage in meta-cognitive monitoring.

According to CLT, intrinsic and extraneous cognitive load should be minimized while germane cognitive load should be increased. This could be accomplished by optimizing the learning materials, provided that the total cognitive load does not exceed the information-processing capacity of working memory (Bannert, 2002; Kirschner, 2002). Diminishing intrinsic and extraneous cognitive load prevents learners from diverting their attention to irrelevant information, whereas increasing germane cognitive load encourages learners to consciously deal with schema construction.
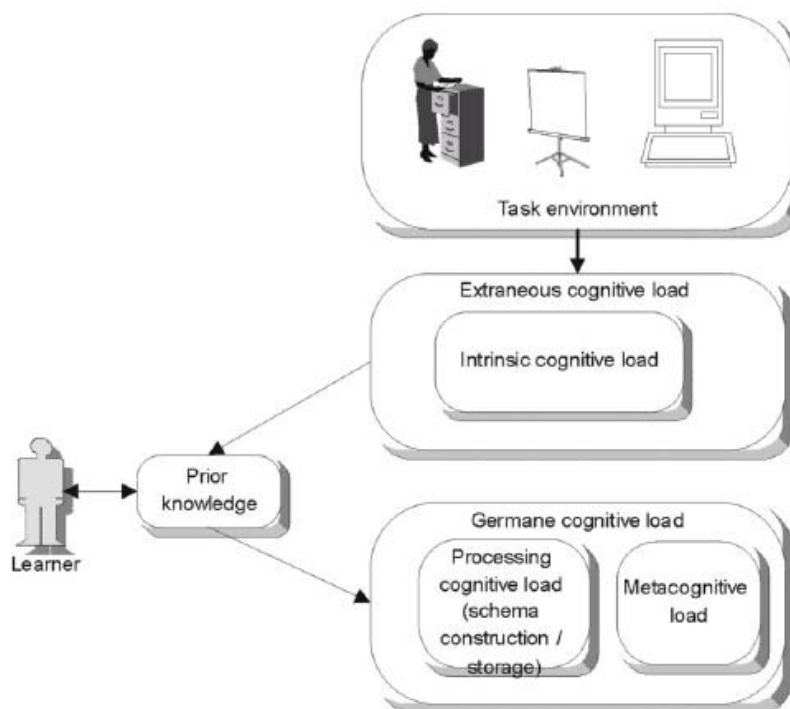
Figure 2: An updated model for CLT suggested by Valcke
(Taken from Valcke, 2002:150)

5.1.2 Measures for reducing intrinsic cognitive load

Measures to reduce intrinsic cognitive load have been investigated by Pollock et al. (2002). They found that complex learning material which is high in element interactivity can only be understood if elements are incorporated in a schema. Due to the limitation of working memory, processing all elements simultaneously to

construct the schema induces working-memory overload. However, the element interactivity can be artificially reduced by presenting material as isolated elements. The results of their study showed that subjects learned more efficiently when the complex material was presented as isolated elements in the first phase and as a whole in the second phase than when the material was presented as a whole in both phases.

5.1.3 Measures for reducing extraneous cognitive load

It is assumed that the effects of extraneous cognitive load are more likely to be observed only when the intrinsic cognitive load is high. If not, the extraneous cognitive load is unlikely to give rise to severe problems in learning, since the overall demand on working memory in this case is not extensive (Sweller, 1993, 1994). CLT proposes several methods to reduce or eliminate the extraneous cognitive load that are relevant to the influence of multimedia presentations on learning. Three effects connected to this kind of cognitive load are the split-attention effect, the redundancy effect, and the modality effect. In the following paragraphs, I shall introduce these effects and the measures by which they may be reduced.

5.1.3.1 The redundancy effect

According to CLT, redundant information in instruction induces extra, unnecessary cognitive load because learners must devote some attentional resources to process the redundant information and therefore fewer cognitive resources are left to process the relevant information. The redundancy effect generally occurs when different sources of simultaneously presented information are comprehensible in isolation and when each source conveys similar information but in a different format (Kalyuga et al., 1999; Kalyuga, 2000). For example, when a textual and a graphical instruction present roughly the same information, and both of them are intelligible in isolation, it is advisable to remove one of the sources to avoid the unnecessary cognitive load (Cooper, 1998).

Kalyuga, Chandler, and Sweller (1999) found that subjects who received a diagram combined with auditory text as instruction outperformed those subjects who received the same diagram with the same text presented both in auditory and visual format at the same time. The redundant textual information did have a negative effect on learning. Kalyuga (2000) further assumed that this redundancy effect might not occur if the redundant sources of information are presented successively with some delay in between because successive presentation does not force learners to process different sources of information at the same time. That is, learners' cognitive

resources are not diverted to processing irrelevant information. Moreover, Kalyuga et al. (1998, 2000) found that when considering learners' levels of expertise, experienced learners required only diagrams while less experienced learners required diagrams with additional integrated textual information to understand the learning material. Therefore, whether a source of information is redundant or not is contingent upon the learners' requirement. Mayer, Heiser, and Lonn (2001) compared the learning performance of students receiving computer-based instruction containing either animation with narration or animation with narration and on-screen text. Again, students trained by means of animation and narration performed better in the problem-solving transfer test than did students who received redundant text in instruction.

In my view, it is normal for the information presented in the form of text and pictures to be similar because they refer to each other. Nevertheless, it is difficult to say how similar the textual and pictorial information should be to consider one of them redundant. What about the mnemonic function of pictures, when pictures are used to help with memorizing while simply presenting the same information as that given in the text? Surely, whether or not a particular source of information is redundant can only be judged by the learners but not by the instructors, since the judgment by the instructors might differ from that by the learners. Levels of expertise or of prior knowledge could be good indicators used to predict learners' needs. However, they could be a danger when the instructor's' estimates of the learners' ability are incorrect.

5.1.3.2 The split-attention effect

The split-attention effect occurs when learners must divide their attention among different sources of information that must be mentally integrated to achieve comprehension. According to CLT, when text and pictures are presented that are not intelligible in isolation and are separated in space or time, learners must often search for the relevant information in the pictures that corresponds to the text (or vice versa) in order to establish the connections between the textual and pictorial information. This kind of search imposes extra cognitive load on working memory and is believed to impair learning. To circumvent the split-attention effect, it is suggested that text should be physically integrated into the pictures or presented auditorily (Sweller et al., 1990; Chandler and Sweller, 1992; Sweller and Chandler, 1994; Cooper, 1998; Kalyuga et al., 1999). In the first case, the physical integration of text and pictures can avoid unnecessary visual search between text and pictures, so that the attentional resource can be focused on learning. The same idea is used for the second case. If

text is presented auditorily, CLT assumes that it does not only avoid unnecessary visual search, but it can also increase the effective capacity of working memory to enhance the efficiency of information processing, which relates to the modality effect that will be addressed later.

Sweller and his colleagues (Sweller et al., 1990; Chandler and Sweller, 1991, 1992) conducted a series of experiments to investigate the effects of split or integrated format of instruction on learning. Using instructional materials concerned with electrical engineering, biology, or geometry, they presented instructions comprised of either only diagrams or both text and diagrams, whereby the text and diagrams were not comprehensible in isolation. When the split-format was used, text was shown either over, below, or next to the diagram. When the integrated-format was used, text was displayed directly in the corresponding positions in the diagram. The results of their experiments were consistent with the hypothesis that subjects performed better when text was integrated with diagrams. A series of studies by Mayer and his co-workers demonstrated the split-attention effect elicited by temporal separation of text and pictures. They found that subjects who received animation synchronized with narrations as instructional material performed substantially better in problem-solving tests than did subjects who received animation with narrations presented either before or after the animation (Mayer & Anderson, 1991; Mayer & Sims, 1994; Mayer et al., 1999). This finding suggests that when animation and narration are presented concurrently, learners are more capable of constructing referential connections between verbal and pictorial information because both sources of information are held in working memory at the same time. An interesting result of the study by Mayer et al. (1999) was that presenting narration before or after animation was proved to impair learning when the narration as well as the animation contained much information. Nonetheless, when the narration and the animation were divided into small portions, the effect of the successive presentation did not differ significantly from that of the concurrent presentation. Mayer et al. concluded that "it is not necessary that corresponding visual and verbal segments be presented concurrently (i.e. in physical contiguity) but rather that they be held in working memory at the same time (i.e. in cognitive contiguity)" (Mayer et al., 1999: 643).

It is noteworthy that the physical integrated-format of instructions did not promote learning when the accompanied text information was redundant. The study by Chandler and Sweller (1991) demonstrated clearly that when a diagram was self-explanatory (e.g., a diagram explaining the flow of blood around the heart, lungs, and body, which is labeled with the names of the essential components; see Figure 3), integrating explanatory texts into the diagram was superfluous. However, when using numbers instead of labels in the diagram (see Figure 4), so that it became

unintelligible without the explanatory texts, the positive effect of the integrated format of instruction (see Figure 5) could be restored.
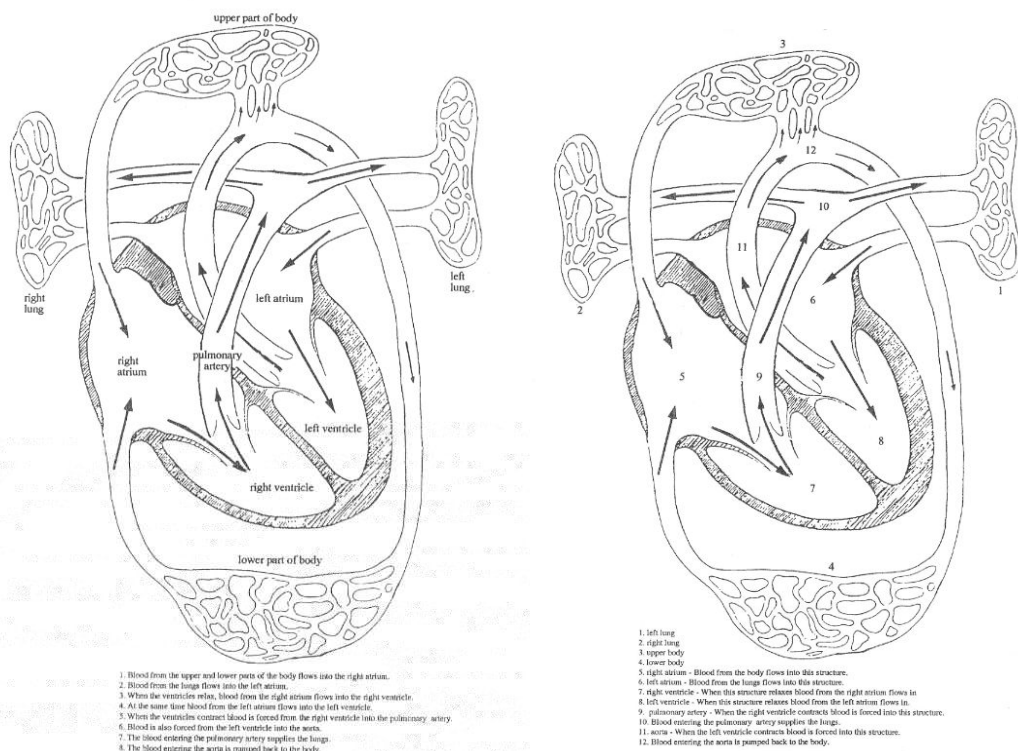

Figure 3: The diagram was comprehensible
without the text beneath it.
(Taken from Chandler & Sweller, 1991: 319)

Figure 4: The diagram was not
understandable without the text beneath it
(Taken from Chandler & Sweller, 1991: 325)
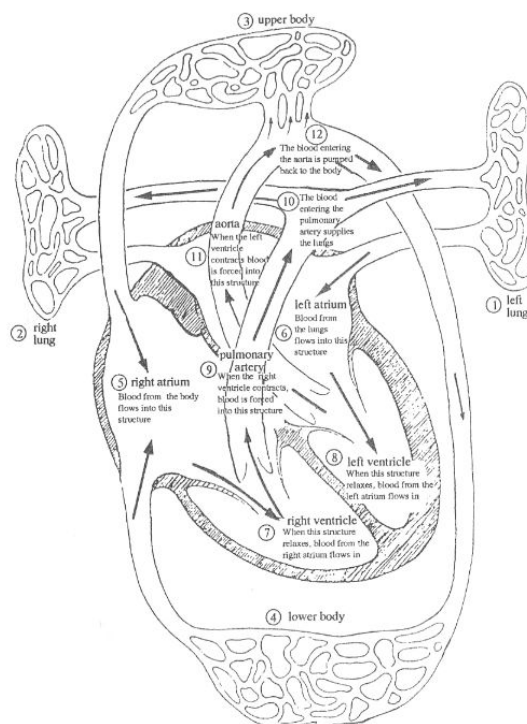
Figure 5: The text was integrated in the diagram. The circled numbers were not shown in the experiment. (Taken from Chandler & Sweller, 1991: 326).

Moreover, the split-attention effect is also found when people must mentally integrate multiple sources of information described in a text. Chandler and Sweller (1992) indicated that the conventional structure of reports of experiments, which has a strict sequential format − introduction, method, results, and discussion − induces split-attention effect and impedes understanding. "The cognitive effort required to mentally integrate various sections of a research report is essential only because of the conventional structure used. With a different structure in which various sections are physically integrated, the requirement to mentally integrate material may be reduced" (Chandler and Sweller, 1992: 240). In another study (Chandler and Sweller, 1996), a split-attention effect could be observed when concurrently using a manual and a computer to learn how to use a software package. Students studying the manual alone outperformed those who studied the manual while practicing each step of the instructions on the computer (reported in Cooper, 1998). To put the findings together, split attention will occur "whenever a learner needs to simultaneously attend to two or more sources of instruction or activities" (Cooper, 1998: Section 5.9).

In my opinion, physical integration of text and pictures can certainly help learners to match the textual information with the pictorial one and thereby eliminate unnecessary visual search. The problem is that the inherent complexity of information to be processed will not be reduced by this manipulation. On the other hand, there will be a technical problem in terms of presenting text fragments in a picture if the picture contains many objects densely arrayed in space. If electronic instead of printed documents are employed in the instruction, mouse-over (i.e. a pop-up message balloon appears when the mouse pointer is moved over an element) or mouse-out (i.e. the message balloon disappears when the mouse pointer is moved out of the element) events are a possible choice to be considered to avoid this problem. Alternatively, learners may be allowed to decide whether the labels (or pop-up messages) should stay after they are displayed. Whether these alternative presentations would have the same positive effect needs to be examined. Kalyuga et al. (1999) proposed another two measures to overcome the split-attention effect. One was to color elements of a diagram in the same colors as corresponding textual elements. The results of their study showed that subjects receiving color-coding format did perform considerably better than did subjects in the control condition. The other one was to use dual-modality presentation, which will be subsequently discussed.

5.1.3.3 The modality effect

The modality effect observed in the framework of CLT is based on the assumption that working memory, as suggested by Baddeley (1986, 1992, 1997), comprises at least three components: a central executive and two subsidiary slave systems, namely the phonological loop and the visuo-spatial sketchpad. According to the CLT, presenting both text and pictures visually induces the split-attention effect, whereas presenting text auditorily together with pictures may circumvent this problem by increasing the effective capacity of working memory, since both auditory and visual channels are used to process the information. "In a split-attention situation, increasing effective working memory by using more than one modality produced a positive effect on learning, similar to the effect of physically integrating separate sources of information" (Kalyuga et al., 1999: 353).

There are several studies that demonstrate the modality effect. The study by Mousavi, Low, and Sweller (1995) showed that students who learned geometry worked examples substantially better when the instructional material was presented in audio-visual format in comparison with that presented in visual-only format. Similar results were found in the study by Kalyuga et al. (1999) using instructional material concerned with solder and light-switching circuitry. A series of studies by Mayer and his colleagues (Mayer and Moreno, 1998, 2002b; Moreno & Mayer, 1999; Mayer et al., 1999) also found that subjects who received animation (showing the process of lightning formation or how a car's braking system works) with concurrent narration outperformed the subjects who received animation with concurrent on-screen text in terms of recalling the content, in matching named elements in the illustration, and when engaged in problem-solving tests. They concluded that displaying animation with on-screen text gives rise to the split-attention effect. The advantage of presenting animation with concurrent narration was that learners could attend to and hold both verbal and pictorial information at the same time, which facilitated the integration of both sources of information.

Nevertheless, some studies indicate several limitations of the modality effect. Tindall-Ford, Chandler, and Sweller (1997) found that the modality effect was only obtained for instructions with high element interactivity but not for those with low element interactivity. They argued that material with low element interactivity which imposes low cognitive load did not show the modality effect because increasing effective working memory was irrelevant when the information to be processed did not overstrain the capacity of working memory. Furthermore, Kalyuga et al. (1999) assumed that auditory text would not be effective if it is too long or complex, which

might overburden working memory because "auditory information is fleeting and difficult to retrieve once heard" (Kalyuga et al., 1999: 368). In contrast, visual text has the advantage of being permanent and thus can be referred to repeatedly. In relation to the redundancy effect, auditory text should not be effective if the same text is presented visually at the same time (Kalyuga, 1999, 2000).

Some interesting findings were obtained in the study of Jeung, Chandler, and Sweller (1997). They conducted experiments to examine under what conditions dual modality presentation (DMP) using an audio-visual format for instructions is superior to single modality presentation (SMP) using a visual-only format. The results showed that when a diagram presented together with (auditory or visual) text was visually high-demanding, DMP was not superior unless a visual aid (flashing) was used to guide learners' visual attention. On the contrary, when the diagram was visually easy to process, DMP was beneficial. Jeung et al. (1997) concluded that the increased effective working memory capacity provided by DMP would only enhance learning if mental resources were not devoted to extensive visual based search in order to coordinate auditory and visual information.

In my view, there is paradox in the modality effect. First, presenting information in different modalities at the same time, according to CLT, is supposed to circumvent the split-attention effect. However, there is virtually no reason to assume that only attending to visual text and pictures would cause split-attention while processing auditory and visual information simultaneously would not. Second, the results presented in the study by Jeung et al. imply that working memory was overburdened by the complex pictorial information or extensive visual search, respectively; thus working memory was less able to process the auditory information. The only plausible explanation for this is that the central executive of working memory is actually responsible for the integration of verbal and pictorial information, but not one of the two subsystems (cf. Chapter 3) because the performance of the phonological loop would not be impaired by the visuo-spatial sketchpad if the two systems were not coordinated by the central executive. However, Mousavi, Low, and Sweller (1995) argued against the view that the integration of verbal and pictorial information is controlled by the central executive. Instead, they suggested that "more working memory resources are available for coordination when a dual-presentation mode is used because more information is likely to be held in both auditory and visual working memory rather than just in one" (Mousavi et al., 1995: 332). Their argument, in my opinion, is a fallacy in that it equates the amount of information held in the two subsystems as the amount of information that can be processed by working memory at the same time. If their

argument were true, DMP would always be more efficient than SMP, which is not the case as shown in the study by Jeung et al.

Finally, I would like to point out a well-established fact that there is a trade-off between the performances of two tasks carried out concurrently because the attentional resources are limited. Yet, the trade-off phenomenon disappears when the two tasks are performed successively. In my view, SMP does not impose the same amount of cognitive load on working memory as does DMP. With SMP, information is actually processed serially, since one cannot read the text and view the pictures at the same time. Therefore, under certain circumstances, DMP could be even worse than SMP because as more information is processed simultaneously, it becomes more likely that the working memory (or the central executive) will be overloaded. In other words, in certain situations, DMP contributes not to overcoming, but rather to inducing the split-attention effect.

5.2 Cognitive theory of multimedia learning

The cognitive theory of multimedia learning proposed by Mayer and his colleagues (Mayer, 1993, 1997; Moreno & Mayer, 2000a; Mayer and Moreno, 2002a, 2002b) is a combination of dual coding theory, cognitive load theory, and constructivist learning theory. Therefore, the theory assumes that: a) humans have separate systems for representing verbal and non-verbal information (adopted from dual coding theory); b) working memory has two independent subsystems – an auditory and a visual working memory (stemming from Baddeley's working memory model); c) both auditory and visual working memory are very limited in capacity (cognitive load theory); and d) meaningful learning occurs when learners actively select and organize the information in each memory into coherent representation, build connections between them, and integrate the information with the already existing knowledge in long-term memory (adopted from the constructivist learning theory and the dual coding theory). A schema of this theory is given in Figure 6.

The theory provides several principles for designing instructional material, which partially overlap with the principles suggested by CLT. Since I have addressed these principles in the previous section, I shall only briefly discuss them here and instead focus on the cognitive aspect of processing multimedia information that this theory proposes.
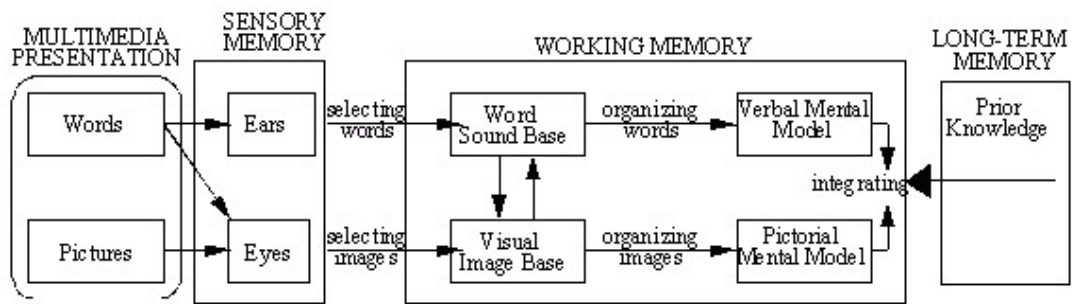
Figure 6: Depiction of a cognitive theory of multimedia learning
(Taken from Moreno and Mayer, 2000a)

According to the theory, the principles for instructional design are as follows (Mayer, 1997; Moreno, Mayer, 2000a; Mayer and Moreno, 2002a, 2002b):

1. *The multiple representation principle*: It is better to present text and pictures together rather than text or pictures alone (Mayer, 1989; Mayer and Anderson, 1991, 1992).

2. *The spatial contiguity principle*: Text and pictures should be presented close to each other rather than separately (Moreno and Mayer, 1999).

3. *The temporal contiguity principle*: Text and pictures should be presented simultaneously rather than successively (Mayer and Anderson, 1991, 1992; Mayer and Sims, 1994; Mayer et al., 1999).

4. *The coherence principle*: Irrelevant words, sounds or videos should be excluded. Instructions should be concise (Mayer et al., 2001; Moreno and Mayer, 2000c).

5. *The redundancy principle*: It is better to present animation and narration rather than animation, narration, and on-screen text (Mayer et al., 2001).

6. *The modality principle*: When presenting text and pictures together, it is better to present text auditorily (Mayer and Moreno, 1998; Moreno and Mayer, 1999).

7. *The personalization principle*: It is better to give explanations in conversational rather than formal style, i.e. using personal pronoun such as "I" and "you" in the text. The study by Moreno and Mayer (2000b) demonstrated that students performed better when explanations were presented in conversational style rather than formal style.

8. *The split-attention principle*: Presenting animation and on-screen text simultaneously induces the split-attention effect, whereas it would not occur if text were presented auditorily (see Figure 7). Moreno and Mayer (2000a) assumed that when animation is combined with visual text, "students try to

74

represent both the animation and the on-screen text in visual working memory. Although some of the visually-represented text eventually may be translated into an acoustic modality for auditory working memory, visual working memory is likely to become overloaded." (p. 3).

I believe that the argument depicted by the split-attention principle, similar to the one proposed by CLT, is based on an incorrect interpretation of the processes of information processing in working memory. It is incompatible with the views of the dual coding approach. Although visual text is perceived by the eyes, it does not mean that the text has to be processed or stored in the visual working memory. Simply because visual text comprises verbal information by nature, there is no reason to assume that it should be processed by the visual memory, which is in fact responsible for processing non-verbal information (i.e. pictorial information). The mental representation of visual text certainly does not comprise visual images because people do not treat a text simply as strings of letters without meaning. Consequently, the argument concerning the split-attention principle does not seem to be plausible.
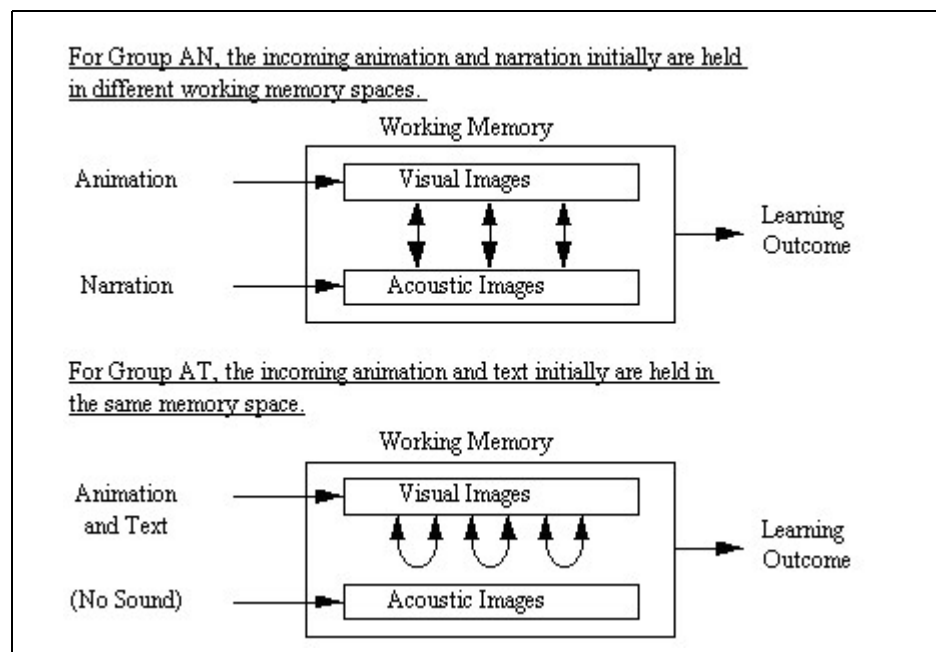
Figure 7: The split-attention principle interpreted by Moreno & Mayer (2000a). The information-processing processes of Group AN (animation +narration) and Group AT (animation + on-screen text) are contrasted.

5.3 Summary

In this chapter, I introduced and discussed two current theories concerning cognitive aspects of processing multimedia information. Both the cognitive load theory as well as the cognitive theory of multimedia learning have proposed several principles for instructional design, which are widely used in educational practice. Despite the empirical evidence, there are some fallacies in both theories, especially in terms of the split-attention and the modality effects. Further research is certainly required to clarify these issues.

6. EYE-MOVEMENT RESEARCH

A promising way of investigating human visual cognition is to observe people's eye-movement behavior while they are performing cognitive activities. Research on reading and scene perception, for example, necessitates eye-movement data to infer underlying cognitive processes. To examine the effects of multimedia presentations on information processing, it is useful to track the viewers' eye movements to find out how they allocate their attention to integrate the various sources of information.

In this chapter, I will introduce some important eye-movement variables usually measured in eye-tracking experiments and their connections with cognitive processes. Afterwards, I shall report on some studies that are concerned with how people coordinate text and picture information with their eyes.

6.1 Essential eye-movement variables

The most frequently measured eye-movement variables are the positions of fixations, the number of fixations, the fixation duration, the saccade length, and the pupil sizes. A fixation occurs when a person looks at something with his/her eyes keeping still. The region of this "point of regard" covers the foveal (the central 2° of vision), the parafoveal (5° on either side of the fixation point), and the peripheral (beyond the parafovea) areas. The clearest vision is in the fovea, and acuity degrades outward toward the periphery. Due to the constraints of sight acuity, it is necessary to move our eyes to the object that we want to see more clearly. When the eyes move from one fixation point to another, this rapid movement is referred to as a saccade. A saccade length is therefore the (spatial) distance between two successive fixations. The fixation duration refers to the amount of time focused on a specific location. The length of a saccade or a fixation duration depends on many factors, such as the viewer's intention, the task the viewer has to perform (reading, visual search, scene perception, etc.), the physical (color, size or shape of the objects or characters) and semantic features of the visuals to be regarded or the text to be read, etc. Generally speaking, the fixation duration usually lasts for 200 to 300 ms. A typical saccade length of reading is 2° and that of scene perception is 5° visual angles on average (cf. Rayner, 1998). An overview of typical values for fixation duration and saccade length in reading, visual search, scene perception, music reading, and typing is given

in Table 1. Rayner (1998) pointed out that the values mentioned below were taken from a number of sources and could vary due to an assortment of factors.

Task	Mean fixation duration (ms)	Mean saccade size (degrees)
Silent reading	225	2 (about 8 letters)
Oral reading	275	1.5 (about 6 letters)
Visual search	275	3
Scene perception	330	4
Music reading	375	1
Typing	400	1

Table 1: Approximate values of mean fixation duration and saccade length in different tasks (Taken from Rayner, 1998: 373)

Pupil size, or the diameter of the pupil respectively, is associated with the intensity of cognitive activation. Several studies provide clear evidence that task-evoked pupillary dilation is an indicator of mental effort, and is a function of the processing load required to perform the cognitive task (Kahneman & Beatty 1966, Kahneman 1973, Beatty 1982).

The eye-movement variables mentioned above are crucial elements for investigating a viewer's cognitive processes while he/she is performing a specific task. In the following section, I will give more information regarding the mental processes underlying these eye-movement variables.

6.2 Eye movements and cognitive processes

Generally speaking, people tend to look at the location to which their attention is directed. Therefore, the positions of eye fixations can reveal the focus of attention. However, based on our daily life experiences, we know that this is not always the case. We can easily shift our attention without moving our eyes. Does this mean that eye movements cannot reveal our mental processes? Certainly not, because when we are performing a cognitive task, such as reading, inspecting a picture or working at a computer, etc., we must focus our attention on the task and therefore direct our eyes to the place where we can get the required information. In this case, gaze positions and attention are closely related.

The control of eye movements, including where to look, for how long and where/when to move our eyes next, is based on "bottom-up" as well as "top-down" mental processes. The former is driven by the physical characteristics of the visual display (data driven), so that specific color or size of objects as well as motion, for example, attracts our eyes. The latter, in contrast, is driven by higher-level cognitive processes (schema driven), so that it manifests a viewer's intention or viewing strategies. Both the top-down and the bottom-up mental processes underlying the eye-movement behavior are important for examining strategies or difficulties in learning. In terms of the top-down processes, we can observe when and where people direct their attention to information and, therefore, we can infer how they process this information. For the bottom-up processes, we can discern the components in the visual display – independent of its relevance to the information to be processed – that best attract people's visual attention, which provides useful guidelines for designing multimedia learning material.

6.2.1 Eye movements in reading and language processing

Research on eye movements in reading and language processing indicate that eye tracking is a promising technique for investigating moment-to-moment cognitive processing activities. The number of studies concerning eye movements in language processing and reading is so vast that it is beyond the scope of this thesis to review them in detail. Instead, I will describe the fundamental eye-movement behavior in relation to reading and the processing of auditory language.

Remembering that viewing patterns vary from case to case as well as from person to person, eye-movement patterns in reading are, in comparison to those of scene perception, relatively predictable. The general characteristics of eye movements in reading have been examined in many studies. When people are reading sentences, the fixation duration on average is about 250 ms. When reading English text, saccade length is on average 7 to 9 characters in size (typically 7-9 characters downstream from left to right) (Liversedge and Findlay, 2000). The number of characters that a reader can process in the fovea is termed the visual span, whereas the perceptual span comprises the number of characters that a reader can process beyond the foveal region. "The perceptual span is asymmetrical about the point of fixation, being extended towards the direction in which the reader progresses through the text. For English readers the perceptual span is about four characters to the left and 15 characters to the right of fixation." (Liversedge and Findlay, 2000: 10, cf. Rayer et al., 1980; Rayner et al., 1982). Fixation duration on a word depends on the lexical, syntactic, semantic, and of course, the visual features of the word. For

example, longer, low-frequency or semantically ambiguous words are fixated longer than others. Word skipping, on the other hand, is also quite a normal phenomenon. Contextual constraint, word frequency, and word length are the most influential factors for word skipping. Words that are highly constrained by the preceding context are skipped more frequently than words that are not constrained. High-frequency words are skipped more often than low-frequency words (Rayner & Well, 1996; Rayner, 1998). Short words are more likely to be skipped than long words (Brysbaert & Vitu, 1998).

Sometimes regressions are made when readers encounter semantic or syntactic difficulties in a sentence such as "The horse raced past the barn fell." Readers are said to have been 'garden-pathed' as they initially analyzed the syntax of a sentence incorrectly. In this case, regressive saccades are carried out to re-read the text in order to amend the inappropriate analysis. The number of regressions that readers make is strongly affected by text difficulty (Rayner, 1998). Moreover, it is more likely that readers will regress to a word on the current line than to words on previous lines (Duffy, 1992; Ehrlich & Rayner, 1983). When readers regress further back in text, they usually make quite accurate saccades to the area of text where they had problems, which indicates that readers often have good spatial memories for the text region they had trouble with (Kennedy, 1983, 1992; Kennedy & Murray, 1987).

The *eye-mind hypothesis* of Just and Carpenter (1987) suggests that readers try to interpret each word of a text immediately as they encounter it, rather than use a wait-and-see strategy. The interpretation of each word occurs while the word is being fixated. When listening, the strategy of immediate interpretation is employed as well. "The clearest evidence for the immediacy strategy and the eye-mind hypothesis is that the time spent looking at a word is strongly influenced by the characteristics of that word" (Just and Carpenter, 1987: 41). As mentioned earlier, fixation duration is positively related to word length but inversely related to word frequency. Thus, readers spend relatively more time reading long or low-frequency words. In addition, gaze duration (the sum of consecutive fixation durations on an item) is unusually long on a word that is semantically or syntactically difficult or anomalous, indicating that the semantic and syntactic analyses of each word occur while the word is being fixated (Just and Carpenter, 1987).

The immediacy strategy reflected in the eye movements when people are listening to spoken language is also observed in the studies of Tanenhaus et al. (1995, 1996). In their studies, subjects were required to follow auditory instructions to manipulate real objects. According to their findings, subjects made a saccadic eye movement to the target object instantly after hearing sufficient information in the verbal instruction, which indicates the incremental processing of spoken language.

On the other hand, visual context clearly influenced the processing of the auditory instruction. When displaying another candidate object that has a similar name to or some common features with the target object, it took subjects more time to identify the target. When there was no alternative candidate referent presented in the visual context, subjects were able to identify the target object before hearing the end of the word. Moreover, the same syntactically ambiguous phrase was interpreted differently in accordance with the different visual contexts. The studies of Tanenhaus et al. (1995, 1996) demonstrate that the auditory verbal input is very rapidly integrated with the information given in the visual context, and that the visual context imposes a strong influence on the earliest moments of language processing.

6.2.2 Eye movements and scene perception

Eye-movement patterns in scene perception are not nearly as predictable as eye movements in reading. The reasons for the difficulty in predicting viewing patterns in scene perception are: a) The array of information in a picture is distributed in the display. It is not simply a sequential alignment of information presented in a text. b) It is hard to define any grammatical structures of the information depicted in a picture and therefore we are not able to predict what a viewer will look at first and how or in what kind of viewing sequences he/she will scan the picture. c) A picture can display different visual information, such as color, shape, texture, orientation, motion, etc. simultaneously, with each having a different effect on eye movements. Because of these difficulties, the findings of studies concerned with eye movements in scene perception are often controversial.

In general, people can already get the gist of a scene during the first fixation. Subsequent fixations on the picture are probably strongly guided by this initial impression. The time required for getting the gist of a scene is probably within the first 100 ms following the onset of the scene (Biederman et al., 1974, Loftus, 1983). For such a rapid acquisition, a large amount of peripheral processing is required. Nelson and Loftus (1980) examined the useful field of view in terms of the acquisition of substantive information during picture viewing. They pointed out that the majority of essential information within a fixation on a scene was drawn from the foveal area (about 2 degrees of visual angle) while less useful information was being acquired from fairly far into the periphery. The probability of correct picture recognition was about 60% in the peripheral region, which is not as high as that in the foveal region (85-90%) but is still above the chance level.

Some studies indicate that viewers tend to quickly get the gist of a scene and then move their gaze to the informative area. The term 'informative area' refers to

something that is unexpected, surprising, distinctive or has a low probability of occurring, such as features that do not fit the impression of the scene or the viewer's schema for whatever the scene depicts (Mackworth & Morandi, 1967; Loftus & Mackworth, 1978). One example from the study by Goodman and Loftus (1981) depicted a fishing scene. In the normal version, the fisherman was pulling a fish out of the water, while in the unusual version the fisherman was pulling a portrait out of the water. In comparison to the fish, the portrait attracted many more fixations. Loftus (1983) supposed that this kind of improbable feature in a scene would be fixated earlier, more often, and with longer duration.

Furthermore, Loftus et al. (1982) examined how fixation duration was distributed during scene perception. In an experiment, subjects saw a single picture for two seconds in a free-viewing situation. The (unpublished) data (reported in Loftus, 1983) showed that the first fixation on a picture was shorter than subsequent fixations. Fixation duration of the second or third fixation was the longest, which probably corresponds to searching or viewing the unusual feature. From the fourth fixation on, fixation declined monotonically. Some other recent studies also confirmed the finding that important or interesting objects in a scene are fixated more and for longer than less important ones, but they did not replicate the finding that semantically inconsistent objects were fixated earlier than consistent objects (Kristjanson & Antes, 1989; Christianson et al., 1991; Rayner, 1998).

The findings of Loftus et al. (1982) can describe the temporal distribution of fixations during scene perception, but they are not able to predict how viewers move their eyes to scan the information from the scene. Difficulty in forecasting the scan patterns in scene perception lies in the scan paths being more or less idiosyncratic. Studies about the effects of viewers' intentions on their viewing patterns yield some information regarding the area of the scene to which viewers' attention is allocated. A study by Yarbus (1967) clearly demonstrated how a viewer's intention influences the way in which he/she inspects a picture. Subjects were shown pictures like the one in Figure 1. When a subject was asked to estimate the ages of people, most fixations were located on the faces of the people (see Figure 2). When a subject was asked to judge the material circumstances of the family, many fixations were on furniture and women's clothing (see Figure 3). By contrast, when a subject was asked to examine the picture without any special intention, the viewing patterns differed from those with special viewing intentions (see Figure 4).

Another issue of eye movements in scene perception is the relationship between memory performance and the number of fixations made on the picture. Boynton (1960) pointed out that when people have limited time to look at a picture, good searchers make many brief eye fixations, whereas relatively poor searchers
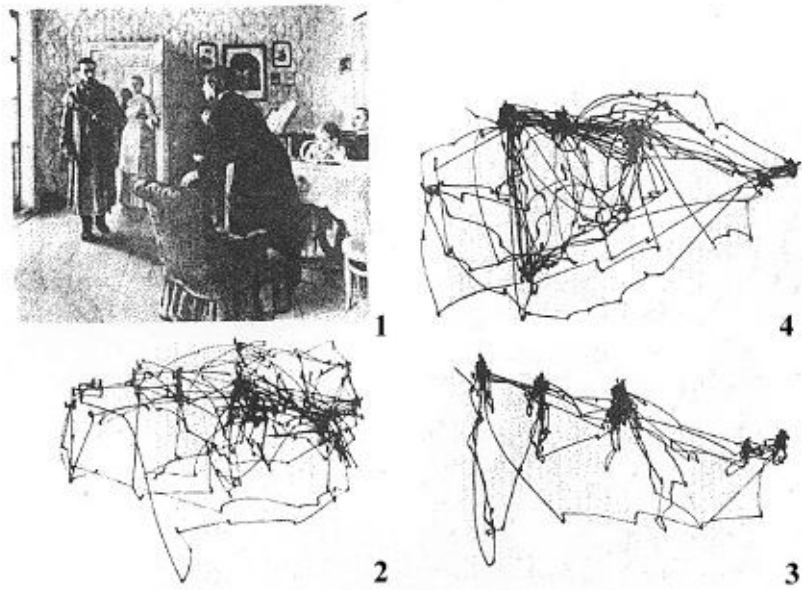
Figures 1 to 4. (Taken from Yarbus (1967)

make fewer fixations with longer duration. Loftus (1972) found that the more fixations a person makes on a picture, the better the recognition memory in relation to that picture. According to Loftus's point of view, information about the picture is transferred to memory in discrete chunks, with each chunk corresponding to an eye fixation. Thus, the greater the number of information chunks, the greater the likelihood of correct recognition. However, Tversky (1974) found that more fixations on words was associated with better verbal recall, while fewer and longer fixations on the picture was associated with better picture recognition. The contradictory findings of the studies by Loftus and Tversky are probably due to the different stimuli used in their experiments. Tversky (1974) assumed that longer fixation duration yields more information per fixation. "Such a scanning pattern, where breadth is sacrificed for depth, might be efficacious in the recognition task, whereby success required discrimination between pairs of pictures with many features in common. If the stimuli are highly differentiated and easily distinguishable, like those stimuli used in Loftus's study, cursorily skimming each stimulus may instead be advantageous" (Tversky, 1974: 278).

6.2.3 The connection between mental workload and eye movements

"Pupillometrics" is a term invented by Hess (1965) to describe a research field (started in 1960) that encompasses the effects of psychological influences, perceptual

processes, and mental activities upon the pupil size. The method used to measure pupil response is referred to as pupillometry. Both positive and negative affect states exert their influence on pupillary dilation. Extreme fear, for example, evokes a manifestly enlarged pupil as a reaction to the fright, whereby the dilation persists even if intense light is shone into the eye. The pupil size actually changes constantly when we are awake. It is never stationary except during sleep. The pupil is highly sensitive and reactive to a number of factors, such as: 1) changes in environmental illumination; 2) closure of the eyelids; 3) accommodation to far and near vision; 4) general anesthesia, narcotic drugs, and other substances affecting the autonomic nervous system; and 5) specific types of neurological and brain damage, etc. (cf. Hess, 1972).

It has been known in psychophysiology that pupils will increase in size during mental activities. Several empirical studies have investigated the relationship between pupillary dilation and mental effort. The results of those studies indicate that the magnitude of pupillary dilation is positively related to the mental effort required to perform a cognitive task. Kahneman (1973) proposed three criteria for any physiological indicator of processing load: "1) It should be sensitive to within-task variations in task demands produced by changes in task parameters; 2) it should reflect between-task differences in processing load elicited by qualitatively different cognitive operations; 3) it should capture between-individual differences in processing load as individuals of different abilities perform a fixed set of cognitive operations" (Beatty, 1982: 276). In his capacity theory of attention, task-evoked pupillary response is justified as the physiological measure of processing load, which reveals the state of mental effort and arousal.

Kahneman and Beatty (1966) examined the task-evoked pupillary response in short-term memory tasks. One of the tasks was the digit span task, where strings of three to seven digits were aurally presented at the rate of 1 per second followed by a 2-second pause. After the pause, the subjects were asked to repeat the digit string at the same rate. The results showed that the pupillary diameter increased with the presentation of each digit, reaching a maximum in the pause preceding the report. During the report, the papillary diameter decreased with each digit spoken, reaching baseline levels after the final digit. The magnitude of the peak pupillary dilation was a monotonic function of the number of items held in memory. A similar pupillary function was also obtained when a string of items was recalled from long-term memory. Another short-term memory task varied in terms of the difficulty of the to-be-remembered items that required subjects to recall four digits, four unrelated nouns, or transform a four-digit string by adding one to each item. The results indicated that pupil diameter increased along with the items' difficulty.

Peavler (1974, reported in Beatty, 1982) investigated pupillary response in digit-span tasks in relation to the limited capacity of short-term memory. Five, nine or thirteen digits were randomly shown to subjects. During the presentation of the digit strings, pupil size increased as the number of digits was raised from 1 through 7. Interestingly, at the seventh or eighth digit, the pupillary response reached an asymptote. No further dilation could be observed afterwards. Since Miller (1956) had argued in his famous study that the capacity of short-term memory for processing strings of unrelated digits was restricted to approximately 7 plus or minus 2 items, the result of Peavler's study suggested that further increases in task demands could not evoke more pupillary dilation because the short-term memory was filled to capacity.

In addition to short-term memory tasks, some other studies examined the pupillary response during language processing or reasoning tasks. The results indicated that when subjects were required to judge pairs of words as similar or different in meaning, the pupils dilated twice as much in relation to the difficult target words. Ahern (1978, reported in Beatty, 1982) employed Baddeley's Grammatical Reasoning Task (Baddeley, 1968) to investigate pupillary response in sentence processing. Subjects were asked to verify a series of sentences which described the order of two successive letters–A and B–, for example, A follows B – BA (true); B is not preceded by A – AB (false). The grammatical complexity differed in the syntax: active-affirmative, active-negative, passive-affirmative, and passive-negative. In this task, pupillary dilation increased with the length and complexity of sentences. Just and Carpenter (1993) explored the intensity of processing load during sentence comprehension by measuring pupillary response. Simpler and more complex sentence types were compared. The two more complex sentence types tested in the experiments were 1) object-relative center-embedded sentences and 2) filler-gap sentences. An example of the first complex sentence type is: *The reporter that the senator attacked admitted the error.* By contrast, the simpler, subject-relative sentence type is: *The reporter that attacked the senator admitted the error.* Filler-gap constructions refer to sentences containing a *wh*-phrase (the filler) that is associated with the empty category where its constituent would occur in the canonical version of the sentence (immediately after the verb). An example of a more complex filler-gap sentences is: *The confused police didn't know which leader the rioters followed noisily down the street after the meeting.* The simpler sentence with a similar structure in which the constituents appear in a canonical order is: *The confused police didn't know whether the rioters followed the leader noisily down the street after the meeting.* The results of this study showed that the pupil began to dilate at the point in the sentence where a syntactic complexity was first encountered, reaching a maximal

diameter about 1.3 seconds later. The gaze duration at this point was raised as well, indicating the immediate response to the demand for the syntactic processing.

The relationship between mental workload and eye movements can also be revealed by the saccadic extent. In a study by May et al. (1990), the effect of auditory task load on the extent of the saccadic eye movement was examined. Subjects were asked to perform one-, two-, and three-channel counting tasks during free viewing. As the complexity of tone counting increased, the range of the saccadic extent significantly decreased. In addition, several studies have reported shrinkage of visual field as a result of increasing mental workload. Macworth (1965) proposed that the shrinkage of the functional field of view (a radius of about 2 to 4 degrees surrounding the point of fixation), serves to prevent an overload of the processing system when more information is available than can be processed. He referred to this narrowing of visual field as 'tunnel vision'. Aside from the effect of mental workload on the size of the visual field, Rantanen and Goldberg (1999) also investigated how the shape of the visual field changes as the mental workload increases. Subjects were required to count three different tones (presented in random order) of a certain frequency, while their visual field was measured. The size of visual field was reduced (by 14%) as the complexity of the tone counting task increased to its highest level, whereby the shape of the visual field became more irregular and smaller, with a more vertically shortened and horizontally elongated form.

6.3 Analysis of gaze positions during learning with text and pictures

Although many eye-movement studies have been conducted on reading or scene perception, studies concerned with how viewers integrate verbal and pictorial information presented by multimedia are still sparse. Only in the last decade have researchers performed studies dealing with this topic. Some of those studies engage in qualitative analyses of eye movements. The advantage of qualitative research is that a viewer's cognitive processes for integrating verbal and pictorial information can be studied on the basis of the viewer's gaze trajectories as a whole. However, there are also some disadvantages of this kind of research. First, the analysis of eye-movement data is very time consuming because the amount of data for each subject is quite large if the time spent on viewing the learning material is several minutes long. Hence, most of the experiments were conducted with only a few subjects. Second, the analysis of viewers' gaze trajectories is a very challenging work, especially when the material comprises a series of complex pictures (static or moving) and the text is presented in different modalities (visual or auditory). Specific

software is required for the analysis of such eye-movement data. Third, the viewing patterns differ from person to person. Consequently, the generalization of qualitatively-analyzed eye-movement patterns is a difficult task to achieve.

The quantitative studies, on the contrary, are carried out with more subjects, whose description of eye-movement behavior can be more generally applicable. Nonetheless, quantitative eye-movement data may only yield information regarding the allocation of subjects' attention to different visual components or to how the mental workload changes as the quantity as well as the quality of the information to be processed is altered, but can hardly demonstrate the temporal and spatial distribution of attention in real time. To investigate the cognitive processes of multimedia-based learning, from my point of view, both the quantitative and the qualitative aspects of eye-movement behavior should be taken into account.

6.3.1 Studies concerning eye movements in text and picture integration

Hegarty (1992) investigated how learners coordinate information from text and diagrams to construct a mental model of a mechanical system. Subjects were asked to process information regarding a series of pulley systems. The eye-fixation data showed that the comprehension process was largely text directed. Subjects usually read a small section of the text and then inspected the corresponding components in the diagram, suggesting that the construction of a mental model is an incremental process. Moreover, the eye-movement behavior showed that subjects tended to shift their gaze toward the diagram at the ends of sentences or clauses. It appears that they had built an initial representation of each clause they read, keeping those representations in a temporary buffer and then checking them against the diagram when the buffer was full. After the subjects read the sentences describing the kinematics of the pulley system, they spent more time inspecting the diagram, whereby they (according to Hegarty's assumption) probably tried to animate the components of the pulley system that were initially represented as static components in their mental models. When the subjects had finished reading the whole text, their final inspection of the diagram tended to be longer and more global than before. In contrast, for subjects who read the text and viewed the diagram alternately, the diagram inspection was shorter and more focused on the components about which they had most recently read.

In another experiment, Hegarty examined individual differences in processing text and diagrams. Subjects with high ability (spatial as well as mechanical abilities) were contrasted with those with low ability. The results indicated that the low-ability subjects inspected the diagram more frequently and read fewer clauses between

diagram inspections. In addition, Hegarty and Just (1989) found that when text was difficult to understand, subjects with high ability compensated for the difficulty in the text by taking more time to view the diagram, whereas low-ability subjects spent less time inspecting the diagram because they were less able to extract information from it. Similarly, when some information was missing from the text, low-ability subjects spent less time inspecting the diagram since they lacked prior knowledge of the domain which would have helped them process the diagram.

Faraday and Sutcliffe (1997) conducted four studies to investigate learners' attention and comprehension of multimedia presentations. In the first study, they employed an eye-tracking system to observe the viewing processes for different visual components, such as labels, and static and moving objects in a multimedia display. Subjects were asked to view an animation showing DNA repair by photo-reactivation, while their eye movements were measured. The text accompanying the animation was presented auditorily. According to Faraday and Sutcliffe, the subjects tended to shift their gazes to the object that had just appeared in the animation and tracked the object's path till the end of the movement. On the other hand, when a motion and a label were presented in different places at the same time, some of the subjects failed to attend to the label because their attention was totally directed toward the moving object, while some other subjects ignored the motion. In general, the presentation elements received more visual attention if they were referred to in the speech track.

In the second study, subjects with high- and low-domain knowledge were compared in terms of their comprehension of the same multimedia presentation. The comprehension was assessed by a free recall test. The number of propositions that were correctly recalled served as a measure of recall accuracy. The difference in recall between the two subject groups was not significant. High-domain knowledge subjects generally recalled a bit more than their counterparts. Propositions that were given only in speech or in animation were generally poorly recalled, whereas propositions in speech reinforced by labels were well recalled.

Faraday and Sutcliffe improved their design further by adding more text captions and additional speech cues to reinforce the captions or more complex motions, and by modifying certain parts of the animation by changing the color and shape of the objects involved. In the third study, they tested the recall of the re-authored version of the multimedia presentation only with low-domain knowledge subjects. Significant improvement in recall was found, in comparison to those low-domain knowledge subjects who viewed the previous version. To ensure that the positive effect of the re-authored version was not simply caused by adding extra information to the original version, they conducted a fourth study by testing the text-

speech only version with another subject group with low-domain knowledge. The text script originated from the third study and was accompanied by a matching speech track. The recall performance of text-speech only group was then compared with those of the second study and the third study. The results showed that the performance of the text-speech group was substantially worse than that in the third study, but did not differ from that in the second study. Faraday and Sutcliffe regard this result as a confirmation that the positive effect of the re-authored version did not only result from the additional information presented in the text captions and the speech track. The modification of the animation was also vital for the comprehension.

In my view, there are problems with the design of the fourth study. The fact that the recall performance of the text-speech only group was considerably worse than that in the third study is not necessarily due to the effect of the modified animation, but could also be attributed to the absence of animation. The fact that the text-speech only group did not outperform the group in the second study cannot be regarded as evidence in favor of the view that the additional verbal information did not support the recall performance, since subjects in the second study could view the animation as an additional resource. A better way of specifying the effect of redesigning the animation and adding verbal information would be to conduct a study using the same text and speech employed in the third study and the original animation in the fourth study, instead of using a text-speech only version.

Despite the minor shortcomings in their research methodology, the studies by Faraday and Sutcliffe still provide some design guidelines that might be useful for producing multimedia learning material, such as: 1) Use animation with care, since motion can capture learners' attention in a positive as well as a negative way. 2) Avoid displaying several presentation items at the same time, but gradually reveal labels and objects to control viewing order. 3) Speech cues should be synchronized to the corresponding labels, which may support the integration of information. Allow enough reading time after cueing a label. 4) Speech and animation can be used to emphasize information; 5) An object and its label should appear together to improve identification. 6) Animation should be cued by speech. Complex or important information given in speech could be reinforced by inserting a caption concurrently.

Furthermore, it is necessary to note that there are only a few studies that qualitatively analyze subjects' gaze positions while they are viewing moving pictures. Technically, it is actually difficult to get accurate gaze positions when the visual stimuli are not static because there will be a shift between the measured gaze positions and the real gaze positions if the eye fixations are plotted on a scene of the moving pictures. Therefore, it is very hard to qualitatively analyze gaze trajectories.

The eye-tracking system Faraday and Sutcliffe used in their experiment was a pupilometer system with a temporal resolution of 20 Hertz. The raw data of gaze positions were time-sliced into four 5 second parts, and the fixations close to each other were regarded as fixation clusters. The eye-movement data in relation to the six subjects were gathered together to construct a mixed trace. In my opinion, the accuracy of the measurement of gaze positions in their study cannot be very high because the time slice of each animation section was large. Therefore, the fixation graph they developed could only roughly outline where subjects' visual attention was located in the scene. Another problem is that when subjects use different viewing strategies to inspect the same visual stimuli, there are actually no common viewing patterns among the subjects. The meaningfulness of constructing a common scan path to represent subjects' viewing patterns is dubious.

Narayanan and Schrimpsher (2000) developed more sophisticated software for aggregating and analyzing learners' eye movements while they were learning the Quicksort algorithm displayed by a system referred to as HalVis (Hypermedia Algorithm Visualization), which contains five different views. Each view (or section, respectively,) provides information on different levels of the algorithm (detailed or fundamental) presented in text, graphics, animation, or a combination thereof. They used an ISCAN ETL-400 eye tracker with a temporal resolution of 60 Hertz. An eye fixation is identified if it is within a 20×20-pixel square and lasts for at least 100 ms. The software module GRIP (Gaze Recognizing and Information Processing) implemented in C++ is used for analyzing the eye-movement data. GRIP is able to deliver specified data about fixations on the display regions occupied by different visual components in various screens (such as text, static pictures or animation) when a learner is indeed interacting with those components. The software provides time stamps giving the chronological order of viewing. In addition, GRIP can aggregate and reduce the voluminous raw eye-tracking data. Successive fixations at the same location are added up into gazes. A jump is computed as the visual attention shifts from one visual component to another. Data, such as the total time a learner gazed at a visual component when it was visible or active, or the percentage of this time that was related to the total visible/active time of this component, etc. can also be computed.

The results of this study showed that the component the subjects viewed most frequently was concerned with the questions about the algorithm, which they could answer while also receiving feedback from HalVis. Besides the questions, it was animation that attracted the most attention. The textual explanation of the algorithm that was shown alongside the animation also received much attention. Gaze shifts from text to a graphical component (static or running animation) and vice versa could

often be observed. According to Narayanan and Schrimpsher, when gazes switch back and forth between different visual components representing the same information, this indicates that learners are trying to integrate different representations to construct their mental models. On the other hand, if the frequency of gaze shifts is high, it suggests that there is a comprehension bottleneck. In my view, a high frequency of gaze shifts between the visual components could also simply result from the characteristics of the visual components. Gaze shifts between a running animation and text, for example, need to be executed quickly if the animation and the text are presented at the same time. Another possible explanation is that the to-be-integrated information is either complex or large in terms of amount. More gaze shifts are mandatory, but that does not necessarily give rise to a comprehension problem. The learning performance should be evaluated together with the eye-movement data to find out whether there is actually a "comprehension bottleneck" and what could be its cause. Unfortunately, there is no information about learning performance mentioned in their paper.

Finally, I would like to point out that it is very time consuming and challenging work to analyze the voluminous eye-movement data produced in a multimedia-based learning scenario. The study by Narayanan and Schrimpsher (2000) has provided a means to overcome this difficulty. In addition, the method they employed for the quantitative analysis has attempted to yield more fine-grained descriptions of the eye-movement behavior during learning, which is carried out by analyzing the chronological order of viewing different visual components in different sections of the learning tool. Since each section contains different concepts of the topic to be learned, the observed eye movements can demonstrate how the learners integrate information from different sources to understand the specific concepts.

6.4 Summary

In this chapter, I introduced the eye-movement variables that are usually measured in the eye-tracking experiments. Those variables are: the number of fixations, the fixation duration, the positions of fixations, the saccade length, and the pupil sizes. The cognitive processes underlying the eye-movement behavior are elucidated with respect to reading, scene perception, and mental workload. Due to the voluminous data, recent eye-movement research that investigates how people mentally integrate multimedia information has employed more sophisticated software to compute the data in a manner that can be more easily analyzed and interpreted.

7. Experiment 1

7.1 Introduction

In Chapters 4 and 5, I reviewed a number of studies that are concerned with the effects of multimedia presentations (MPs) on learning. It appears that research investigating the issue has overall yielded inconclusive results. Thus, more research is required to systematically investigate the advantages or disadvantages of MPs in education. Several studies have examined the influence of MPs on learning scientific concepts, such as the working of some mechanical device, or the meteorological conditions for a thunderstorm or the like. Other studies have investigated the effects of using MPs for teaching mathematics, physics, biology, chemistry, and medicine, etc. However, with a change in the multimedia-application purpose, the effects of MPs on learning may sometimes be quite different.

The current research is a kind of basic research, the objectives of which are to investigate: 1) whether moving pictures are superior to static pictures in conveying information that involves a series of sequences of actions (i.e. procedural tasks), movement of objects, or changes in a state of affairs over time; and 2) whether presenting information in different modalities (e.g., audio-visual format) is indeed beneficial for information processing.

To find answers to these questions, I conducted an experiment in which subjects were asked to learn how to assemble a three-dimensional puzzle on a computer. I not only investigated the effects of different MPs on learning but also used eye-tracking techniques to examine subjects' information-processing strategies during learning. In Sections 7.2 and 7.3, I shall give detailed information about the variables, design, and hypotheses in regard to this experiment.

7.2 Variables

7.2.1 Independent variables

The information examined here was both verbal and pictorial in kind. Accordingly, the learning materials were presented using texts and pictures in different modalities simultaneously. To systematically examine the effects of different MPs on learning efficiency, a 2 (text mode) by 2 (picture mode) factorial design was used in this

experiment. The two levels of the factor *text mode* were 'written' and 'spoken' whereas those of *picture mode* were 'static' and 'moving'. The presentation modes of text and picture were cross-combined. In combination, there were altogether four experimental conditions of the learning materials:

- **Condition (written/static);**
- **Condition (written/moving);**
- **Condition (spoken/static);**
- **Condition (spoken/moving).**

Further two independent variables that determine eye-movement behavior were analyzed: *the number of propositions* in the texts, and *the number of objects* in the pictures. Finally, subjects' spatial ability and prior experience with similar 3D puzzles were also taken into account. They were used as covariates in the statistical analysis for assessing the learning efficiency.

Spatial ability

With regard to the learning task in this experiment, the specific ability facilitating perception and comprehension of the information to be learned would be the *spatial ability*. It refers to the ability to process spatial information, which should affect the efficiency of learning instructions involving assembly and sequences of actions. To measure subjects' spatial ability, I have used the Cube Folding Test developed by Meili (1955). In this test, subjects were given a number of drawings each of which depicted a cube or a block unfolded or unwrapped. Subjects' task was to imagine how to fold the pattern to form a cube or block, and then mark the square opposite the square marked with 'u' and mark those edges that would touch the marked edges on the pattern if folded. A sample item is given in figure 1. Subjects' achievements in the Cube Folding Test should be regarded as a valid measure of their spatial ability.

Prior experience with similar 3D puzzles

To roughly estimate subjects' prior experience with similar 3D puzzles, subjects were asked to rate the amount of their experience on a four-point scale (very much, much, a little, none).
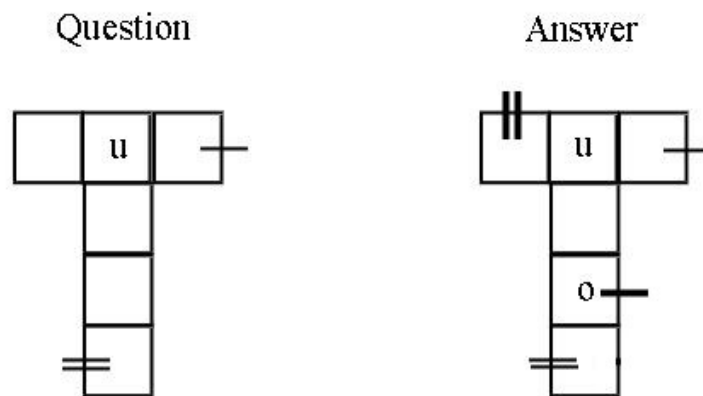
Figure 1: A sample item from the Cube Folding Test

7.2.2 Dependent variables

According to the conventional methods for measuring learning outcomes, subjects can, for example, be asked about the content of the learning materials. Multiple choice tests, free recall, and problem solving tests are frequently employed methods. The first two methods measure how well or to what extent subjects can remember the information after learning, whereas problem-solving tests measure to what extent subjects can apply the information they just learned to solve certain problems. For example, if subjects have learned how a mechanical system works, a typical problem-solving test might include questions concerning trouble-shooting. Since the solution to the cube puzzle is concerned with the description of sequential or successive actions, it is, in my opinion, not appropriate to ask subjects to verbally recall the content of those instructions. The easier and the more appropriate way to assess subjects' performance in this experiment is to ask them to perform the actions described in the instructions, i.e. to assemble the puzzle.

The main dependent variable used for assessing the learning outcomes was *learning efficiency,* which was defined as follows: Learning Efficiency = The score subjects achieved in the assembly task / (the learning time + the task time) (for more details see Section 7.5.1). The reason why I took learning time and task time into account was that most of the studies that examine the effects of MPs on learning have ignored time as a parameter in efficiency in information processing. However, a particular MP might lead to the best learning result at the cost of learning time. In my view, a really good MP should be able to present information in a way in which learners can easily process the information in the least amount of time. As to the task time, I assume that the better a subject has comprehended the information, the less time he or she will require to retrieve the information for performing the task.

In addition to learning efficiency, the time subjects spent to learn each instructional section (*section time*) and the frequency with which they repeated each section (*section frequency*) were analyzed. Moreover, as an empirical approach to learning strategies, I investigated subjects' eye movements in order to analyze the underlying cognitive processes regarding how subjects process the information in the learning materials.

Eye-movement variables and information processing strategies

With the help of eye-tracking techniques, subjects' information-processing behavior can be observed on-line. By measuring eye movements, it is possible to track subjects' focus of attention during stimulus inspection, and thus, I am able to reconstruct their learning strategies and learning processes. For example, a subject's eye movement trajectories can indicate at what time and in which order the subject has been paying attention to various elements within the learning materials.

All in all, the eye-movement variables measured in this experiment included: the number of fixations, the fixation duration, the fixation rate, the number of gaze changes between the text and the picture regions, and the percentage of time spent on the text vs. the picture region of the instructional display. I examined how these eye-movement variables were influenced by the experimental conditions, text mode, picture mode, the number of propositions, objects, and visits to the same section.

Generally speaking, the number of fixations is positively related to the amount of visual information to be processed and the time taken to inspect the stimuli. A high density of fixations can be taken to indicate that the subject is examining a particular area with particular care. Long fixation duration or large pupil size is a sign of great mental effort (Velichkovsky, 1995). Due to the limitations in terms of the function of the eye-tracker employed in this experiment, the measurement of pupil size is, however, problematic if I want to compare the pupil sizes between subjects. Since the SMI EyeLink Eye-Tracking System is a video-based eye-tracking system, the pupil sizes measured by this system are affected by the distance between the eye cameras and a subject's eyes. The closer the cameras are set to the eyes, the larger the pupil sizes will be. The variation in the distances between the eye cameras and the eyes cannot be controlled very well because the eye cameras are adjusted manually. Due to this limitation of the eye-tracking system, pupil sizes cannot yield plausible information about the intensity of mental workload, and therefore were not analyzed in this experiment.

Moreover, the fixation rate, which is defined as the number of fixations per second, can yield information about the speed of a subject's eye movements. The

number of gaze changes between (written) text and picture regions shows how often subjects switched their gaze positions between those regions in order to integrate the textual and pictorial information. Finally, the percentage of time subjects spent on reading the texts vs. inspecting the pictures gives information about how intensely subjects paid attention to each source of information.

7.3 Hypotheses

In regard to the influence of picture mode, I assume that moving pictures should be superior to static pictures in enabling subjects to construct an appropriate mental model of the assembly procedures. Subjects in moving-picture conditions should outperform subjects in static-picture conditions. As to the influence of text mode, I assume that the modality effect would hold because both the text and the picture information involved in the instructions is not highly complicated. When the effects of text modes and picture modes are combined, I assume that subjects in Condition (spoken/moving) should learn most efficiently because moving pictures can best demonstrate the procedures to be learned, and spoken texts can facilitate information processing. Subjects in Condition (spoken/static) should outperform subjects in Condition (written/static) because of the modality effect. Moreover, due to the advantage of moving pictures, subjects in Condition (written/moving) should outperform subjects in Condition (written/static).

Eye-movement hypotheses

The information processing strategies of the subjects should be revealed by their eye-movements. I assume that with static pictures and written texts, subjects' eyes will switch between viewing the text and the picture regions of the display more often than subjects who view video clips and written text. In the first case, subjects can easily integrate information across fixations directed at the text and the pictures, whereas in the second case, subjects would miss some information from the video clips while attending to the text. To avoid any loss of important information, subjects will probably view the videos and read the text separately.

Furthermore, the speed of eye movements, or the fixation rate, to be precise, should be affected by the spoken text: it should be slower than when subjects view the pictures without having any auditory input simultaneously. When processing spoken texts with static pictures, subjects' eye movements should be slower than when processing written texts with static pictures, because the way in which subjects

inspect the pictures should be influenced by the content of spoken text (provided that subjects pay attention to it).

The number of fixations should be positively related to the number of propositions in the texts and to the number of objects depicted in the pictures because the more information that is to be processed, the more fixations are required to establish the appropriate referential links. The fixation rate, which refers to the number of fixations per second, should be positively related to the number of objects as well because many brief eye movements are supposed to be more efficient for scanning a large number of objects. If so, the inverse should hold for fixation duration.

7.4 Method

7.4.1 Subjects

The subjects were 48 students from different departments at the University of Bielefeld. Their ages ranged from 19 to 39. They were paid for their participation in this experiment. All subjects had German as their first language, and not one of them was color-blind. Most of them did not have any experience with similar puzzle games.

7.4.2 Materials

The learning materials demonstrated the procedures for assembling a 3D cube puzzle. They were divided into 10 sections each of which was presented as a web page on a computer screen. The first section included a brief overview of the puzzle. In each successive section, one particular step in the solution to the puzzle was presented in detail. According to the sequential nature of the procedures, the 10 web pages were linked in a linear fashion, so that subjects could study the solution step by step. Subjects were free to jump to the previous section or to the next section by means of a mouse click on one of the two navigation buttons displayed in the bottom right corner of the screen. The display was set to 640×480 pixels in true color (24 bit).

Corresponding to the four presentation conditions, four different layouts of the materials were prepared. Examples are given in Figures 2 to 5:

1. Condition (written/static)

   For this condition, there were two photographs at the top of every page. Each photograph was 265×230 pixels in size. The photograph on the left side showed the state of the puzzle parts before they were assembled. The photograph on the right side showed the state after assembly. A small arrow between the two photographs indicated the order in which they should be viewed. Below the photographs there was a written text which explained the action shown by the photographs.

2. Condition (written/moving)

   For this condition, there was a video window 300×260 pixels in size at the top of the screen. A control bar with two buttons was below the video window. By clicking on these buttons, subjects could start, stop, or replay a video clip depicting the assembly of particular parts of the cube in close correspondence to the photographs used in Condition (written/static). The same texts as in Condition (written/static) were shown below the video window.
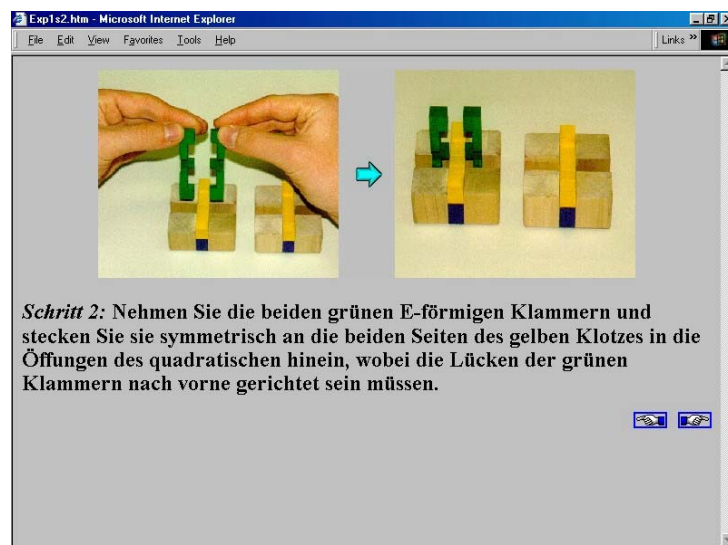

Figure 2: An example of the materials shown in Condition (written/static)

7. Experiment 1

Figure 3: An example of the materials shown in Condition (written/moving)

3. Condition (spoken/static)

    For this condition, the same photographs as in Condition (written/static) were
    used in the corresponding places. In contrast to Condition (written/static),
    however, the texts were presented auditorily. As soon as a page was loaded, a
    sound file presenting the spoken text would run automatically. In order to set the
    volume, or to replay or stop the spoken text, the subjects could use a control bar
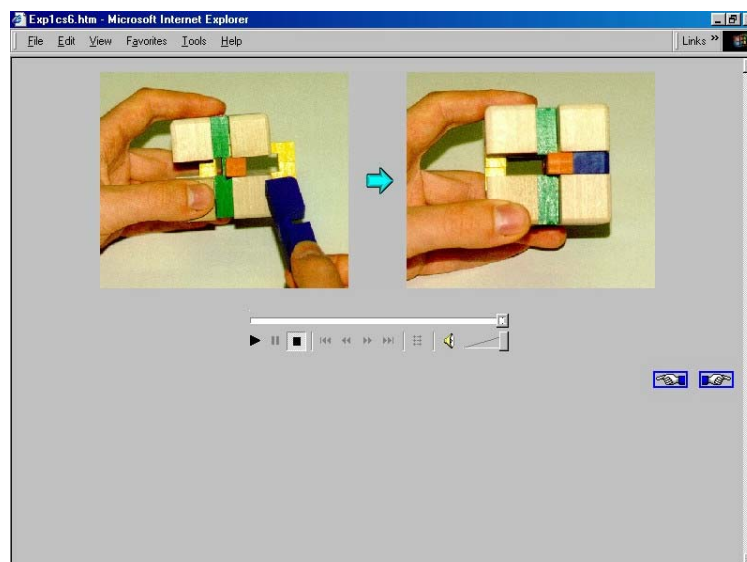   located below the photographs.

Figure 4: An example of the materials shown in Condition (spoken/static)

4. Condition (spoken/moving)

    For this condition, the same video clips as in Condition (written/moving) were
    employed. In contrast to Condition (written/moving), however, the text was

presented auditorily, synchronized with the video clip. The video clip with the spoken text would play automatically as soon as a page was loaded. In addition, subjects could replay or stop the video clip by means of a control bar with two buttons below the video window.

Figure 5: An example of the materials shown in Condition (spoken/moving)

7.4.3 Apparatus

An SMI EyeLink Eye-Tracking system was used to measure subjects' eye movements. The video-based system consists of two standard computers and an eye-tracker headset.

- *The first computer ("Subject PC")* is used for displaying stimuli. It has a Pentium 166 processor and comes with a 20″ CRT monitor. Each corner of the monitor is marked by an infrared LED (light emitting diode). The learning materials and the stimuli for calibration were displayed on the subject's monitor.

- *The second computer ("Operator PC")* is used for recording the data received by the cameras and for calculating any eye movement parameters. It is equipped with a Pentium 133 processor and a 17″ CRT monitor. On that screen, the operator can monitor the subjects' eye movements while the experiment is running, thus allowing the operator to decide about re-calibration in the course of the experiment. The subject PC and the operator PC are linked by an Ethernet-Link. The operator PC is responsible for recording data received by eye- and head-cameras, and for computing the gaze positions, saccade length, and fixation duration, etc.

- *The eye-tracker head*set (see Figure 6) supports three miniature digital infrared cameras. One camera (the "head camera") is directed at the subject's monitor; it receives the signals from the four infrared LEDs. Two more cameras (the "eye cameras") that are directed at the subject's eyes yield digital images of the pupils, which can be analyzed by the hardware and software in the operator PC.

Figure 6: The eye tracker headset of the SMI EyeLink System (Pomplun (1998))

With the help of the LEDs in the corners of the subject's monitor and the head camera attached to the eye tracker headset, the position of the subject's head can be calculated relative to the monitor. The eye cameras attached to the headset yield information about the position of the subject's pupils, so that, in effect, the pupil positions can be computed relative to the positions of the LEDs on the monitor. The spatial precision of the gaze position measurement lies within the range from 0.7° to 0.9°, and the temporal resolution of the system is 250 Hz. A scheme of the SMI EyeLink eye tracking system is depicted in Figure 7.

Prior to eye tracking, the system has to be calibrated. The calibration yields reference data for the computation of gaze positions. Calibration proceeds as follows: a dot is presented successively at nine different positions on the subject's monitor, and the subject has to fixate this dot as accurately as possible every time it appears on the screen.
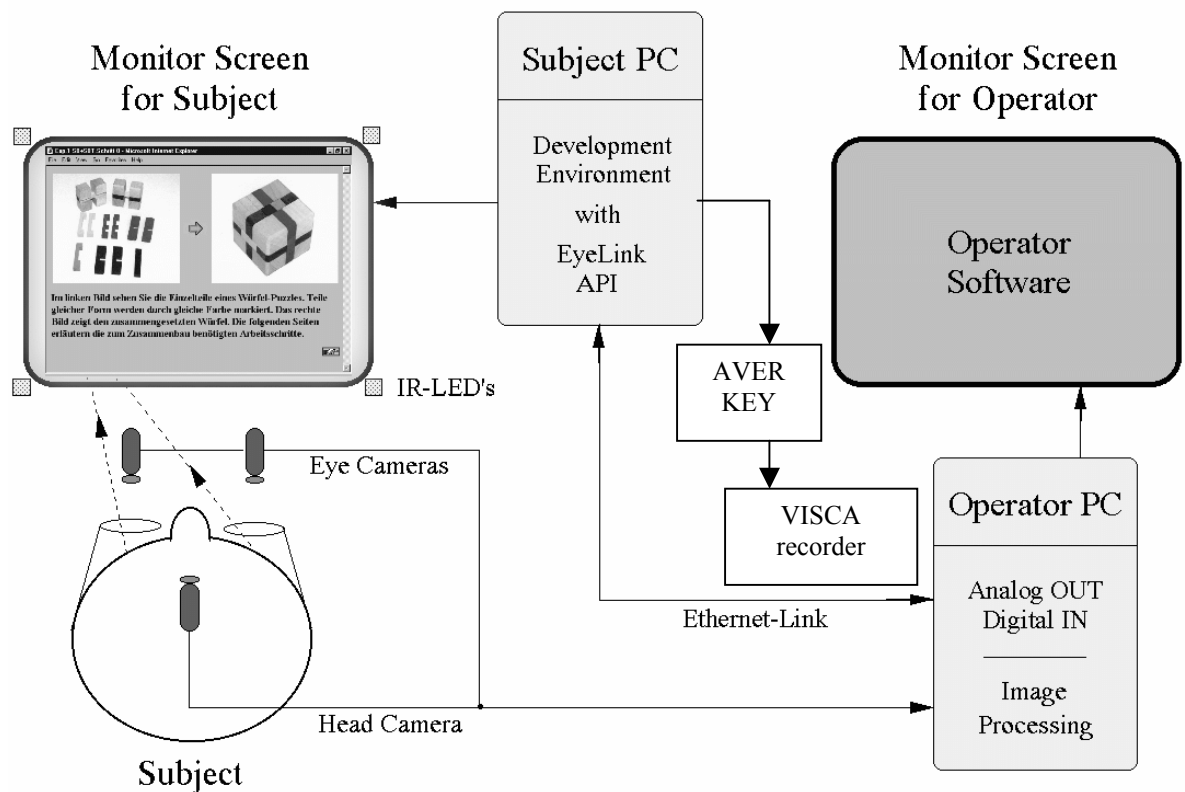
Figure 7: Scheme of the SMI EyeLink System (adapted from Pomplun (1998))

In order to reconstruct the subjects' gaze trajectories 'a posteriori,' that is, to determine when and where subjects had been looking, their eye movements had to be projected onto the stimuli they viewed at that time. In doing so, their eye movements were synchronized with the stimuli by means of a video recording system for tapping the sequences of the stimuli and a software specially developed for this purpose. An AVER KEY interface was employed to convert the signals delivered by the graphic card of the subject PC into video signals, so that a VISCA recorder (video system control architecture; SONY CVD-1000) could be used to videotape the sequences of the stimuli on a Hi-8 video cassette.

The program used to synchronize eye movements with the videotaped stimuli was VIPER (Clermont, 1995). With the help of VIPER, researchers can observe a subject's eye movements which are shown as two cursors (one for the left eye and one for the right eye) moving across the stimuli. Although it is impossible to evaluate the observation of eye movements via videos in a quantitative way, the visualization of eye movements on a videotape provides the best way of obtaining qualitative data that give an impression of the dynamics of attention.

7.4.4 Procedure

Subjects were randomly assigned to the four experimental conditions. They were asked to learn the materials to such a degree that they felt they could successfully solve the cube puzzle. The results of their learning were tested immediately after they finished learning. While they were studying the materials, their eye movements were measured and recorded on a videotape.

The course of an experimental session (maximally 60 minutes) can be roughly divided into four phases:

- *Phase 1*: Subjects were given all the parts of the cube and 10 minutes to solve the puzzle on their own. The purpose of this phase was to let the subjects become acquainted with the puzzle and to get an idea of the problem they had to solve. Those who were unable to solve the puzzle were allowed to proceed to Phase 2.

- *Phase 2*: This was the information processing phase. During this phase, subjects had to learn how to solve the cube puzzle on a computer. While the subjects were studying the instructions, their eye movements were recorded. Basically, there was no time limit for this phase. Subjects could browse the instructional sections at their own pace; they were allowed to view each section for as long and as frequently as they liked. As soon as the subjects felt that they had learned how to solve the problem sufficiently well, they had to notify the experimenter to end this phase. The overall time a subject required for learning was recorded as his or her learning time.

- *Phase* 3: This was the test phase. During this phase, subjects were given all the cube parts again. They had to assemble them as quickly and correctly as possible in line with the instructions they had been given in Phase 2. The time for assembly was limited to 15 minutes. Altogether, there were nine operations to be performed to assemble the cube. Each correctly performed operation scored one point. The maximum score was 9 points. Aside from the assembly score, the time a subject took to assemble the cube was also recorded.

- *Phase* 4: During this phase, some secondary data were collected. First, subjects' spatial ability was tested by using the Cube Folding Test, which took about 10 minutes. Secondly, a questionnaire was administered to the subjects.

In the questionnaire, demographical details such as age, sex, and subject of study were ascertained. Then, the following questions were to be answered:

1. How would you rate the comprehensibility of the texts you have read or heard? (Very easy/ Easy/ Medium/ Difficult/ Very difficult)
2. How would you rate the comprehensibility of the pictures or the videos you have seen? (Very easy/ Easy/ Medium/ Difficult/ Very difficult)
3. What has been crucial to you in understanding the solution? (The texts/ The graphics/ Both)
4. What has supported you the most in memorizing the solution? (The texts/ The graphics/ Both)
5. How much experience did you already have with similar 3D-puzzles? (Very much/ Much/ Not so much/ None)

7.5 Results

7.5.1 Learning efficiency

In order to assess the learning outcomes, I would like to first of all define suitable criteria for the assessment. Since there was no time limit for studying the materials in this experiment, subjects' learning performance could not be judged only based on their assembly score. What is more appropriate is an achievement index that assesses the subjects' learning efficiency. In so doing, I have taken the following criteria into consideration:

-   The score for assembling the cube (C_score)
-   The time used for learning the instructions (LT)
-   The time taken for assembling the cube (AT)

Due to the negative correlation between AT and the C_score ($r = -0.779$; $p < 0.001$), it is more appropriate to measure learning efficiency as follows:

Learning efficiency (LE) = C_score/(LT+AT)

That is, learning efficiency was therefore high if the C_score was high while both learning time (LT) and assembly time (AT) were short.

It should be pointed out that the maximum time for assembly was 15 minutes. However, some subjects aborted the assembly task because they could not remember any further information from the materials to complete the cube. In this case, their C_score was registered as they aborted the task, and their AT was automatically

registered as 15 minutes because their performance would not have improved even if they had had more time for doing the task.

In addition to the experimental conditions, two further factors should be taken into account as well, namely, subjects' *spatial ability and the amount of experience* gained with similar 3D-puzzle games. These factors were entered as covariates in an analysis of covariance for examining the effect of MPs on learning efficiency. According to the ANCOVA, subjects' spatial ability and prior experience were equally distributed among the four experimental conditions.

Figure 8 shows the learning efficiency as a function of the experimental condition. Simple factor effects revealed that text mode had a significant effect on learning efficiency ($F(1; 42) = 11.006$; $p < 0.005$): LE was higher for spoken-text conditions (mean LE = 0.009) than for written-text conditions (mean LE = 0.007). In contrast, the effects of picture mode and the interaction between text mode and picture mode were not significant. However, this pattern of results was in line with a tendency towards an interaction between text mode and picture mode ($F(1; 42) = 2.672$; $p = 0.11$), which indicates that the effect of text mode on learning efficiency was more or less dependent upon the accompanying picture mode. Therefore, the superiority of spoken texts was applicable only to moving pictures but not to static pictures.

Pairwise comparisons using t-tests revealed that the mean learning efficiency of Condition (spoken/moving) was significantly higher than that of Condition (written/static) ($t(22) = -2.365$; $p < 0.05$), Condition (written/moving) ($t(22) = -2.773$; $p < 0.05$), and Condition (spoken/static) ($t(22) = -2.271$; $p < 0.05$). The differences between Condition (written/static) and Condition (written/moving), Condition (written/static) and Condition (spoken/static), and between Condition (written/moving) and Condition (spoken/static) were not significant.
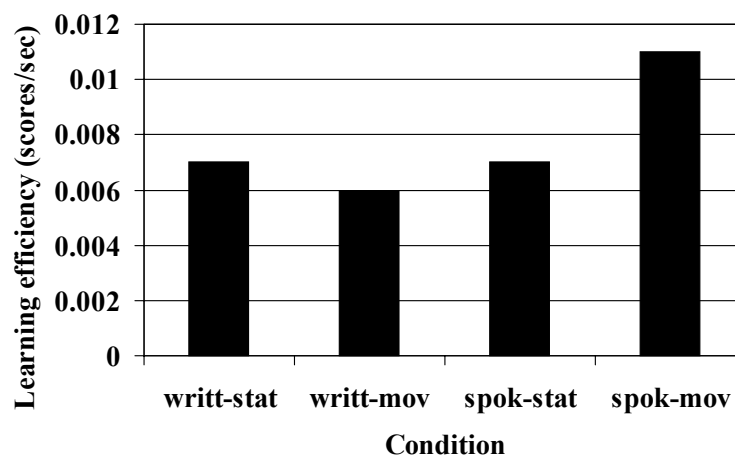

Figure 8: Learning efficiency as a function of experimental condition

7.5.2 Learning time

While the learning efficiency gives information about how efficiently subjects have learned, the learning time can show us how much time subjects have spent in processing the learning materials to achieve that performance. Figure 9 shows the mean learning time for the four experimental conditions. Condition (written/moving) had the longest learning time on average (mean = 675.263 sec) whereas Condition (spoken/moving) had the shortest such time (mean = 505.904 sec). The mean learning time was independent of text mode and picture mode. However, the interaction between text mode and picture mode was significant ($F_{(1; 44)} = 4.780$; $p < 0.05$). Pairwise comparisons using t-tests showed that the learning time in relation to Condition (spoken/moving) was significantly shorter than that for Condition (written/moving) ($t(22) = 2.104$; $p < 0.05$) and Condition (spoken/static) ($t(22) = 2.724$; $p < 0.05$).
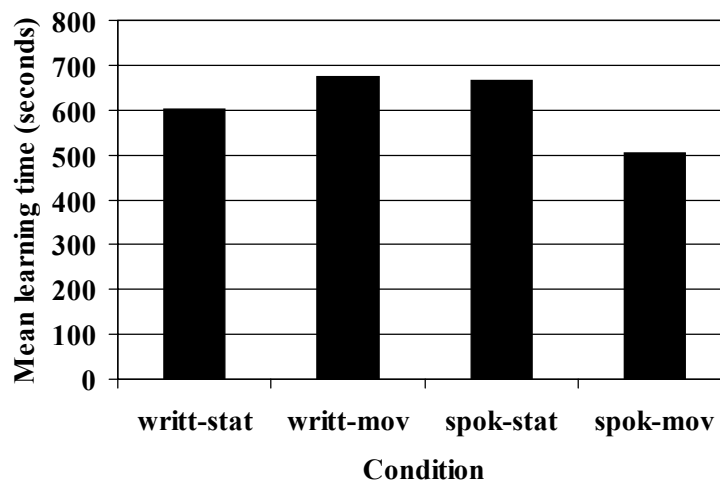
Figure 9: Mean learning time as a function of experimental condition

In order to outline the learning processes, I analyzed two dependent variables: the section time and the section frequency. The reason why I analyzed these two variables was that the total learning time of a subject was not able to reveal the policy related to how the subject devoted time to study a particular instructional section, or how frequently he/she needed to repeat (rehearse) each section, in order to memorize it well. To investigate the extent to which the information processing was exactly influenced by the presentation conditions or by the quantity or quality of the information to be learned, analyses of the section time and the section frequency were deemed to be essential.

7.5.3 Mean section time

This variable is concerned with the time subjects on average spent learning an instructional section. Factors affecting this variable were the presentation conditions, the number of propositions, the number of objects, and the number of visits to that section. Both between-conditions and within-condition analyses were carried out. Between-conditions analyses made it possible to relate any differences between presentation conditions to specific structural details concerning the multimedia information, whereas within-condition analyses were suited to revealing specific details regarding the course of information processing.

7.5.3.1 Between conditions analyses

First, the mean section time for the four experimental text-picture conditions was compared. Figure 10 shows the mean section time as a function of experimental condition:
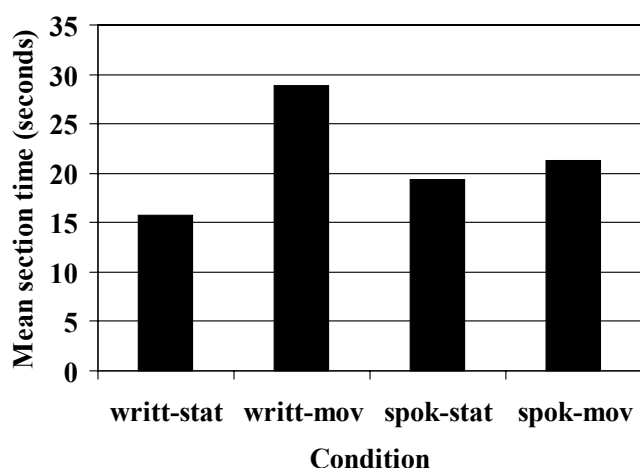
Figure 10: Mean section time by condition

Both text mode ($F(1; 1328) = 5.297$; $p < 0.05$) and picture mode ($F(1; 1328) = 73.110$; $p < 0.001$) as well as the interaction between them ($F(1; 1328) = 40.826$; $p < 0.001$) had a significant effect on the mean section time. When written texts or moving pictures were presented, the mean section time was longer. Post-hoc comparison using Bonferroni tests showed that the mean section time of Condition (written/moving) was significantly longer than that of the other conditions, whereas the mean section time of Condition (written/static) was the shortest among the four experimental conditions.

Figure 11 depicts the mean section time as a function of section. By comparing the mean section time between the four conditions separately for each section, significant differences could be found in sections S0, S1, S2, S3, S7, S8 and S9. This means that the presentation mode mostly took effect in the first and the last sections of the learning material.
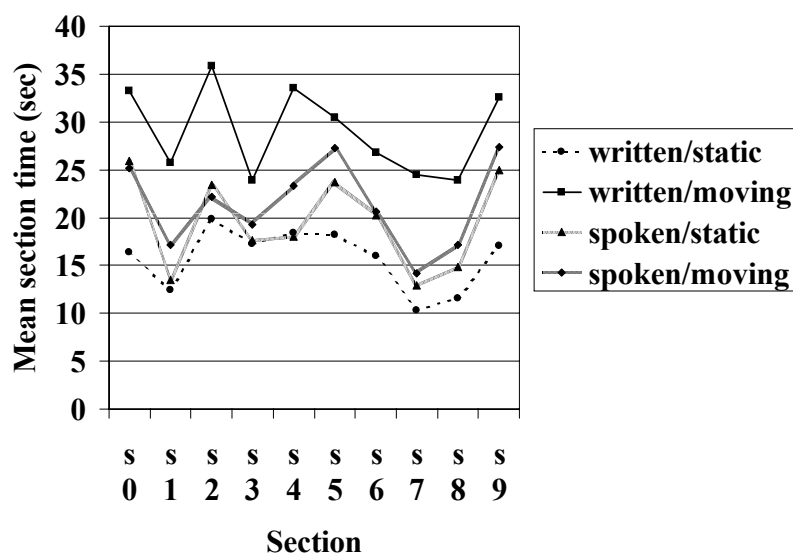
Figure 11: Mean section time as a function of experimental condition, by section

7.5.3.2 Within-condition analyses

The analysis of the mean section time within each condition indicated that there were some significant differences in the mean section time between sections for Condition (written/static) ($F(9; 440) = 2.201$; $p < 0.05$), Condition (spoken/static) ($F(9; 335) = 3.968$; $p < 0.001$), and Condition (spoken/moving) ($F(9; 250) = 3.203$; $p < 0.005$). The Bonferroni tests revealed that the differences for Condition (spoken/static) were between S0 and S1, S0 and S7, and S5 and S7, whereas those for Condition (spoken/moving) were between S5 and S7, and S7 and S9.

7.5.3.3 Effect of the number of propositions

The criteria used for counting the number of propositions were developed by Kintch (1974), which can be regarded as a valid measure of the semantic content of a text. Word categories regarded as constituting a proposition were verbs, adjectives,

adverbs, conjunctions, prepositions, and quantifiers. Table 1 shows the number of propositions in each section.

S0	S1	S2	S3	S4	S5	S6	S7	S8	S9
11	5	15	14	15	14	16	12	11	11

Table 1: Number of propositions of the ten instructional sections

Figure 12 shows the mean section time as a function of the number of propositions. The number of propositions had a significant effect on the mean section time ($F_{(5; 1326)} = 7.233$; $p < 0.001$). Due to the large variance between the values, it seemed appropriate to divide them into two conditions (few propositions (5, 11, or 12) vs. many propositions (14, 15, or 16)) to find out the general characteristics of the relationship that holds between the number of propositions and the mean section time.
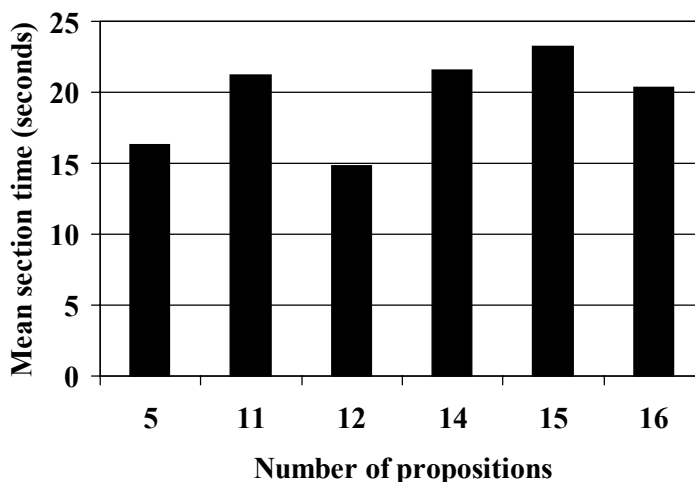
Figure 12: Mean section time as a function of the number of propositions

A t-test showed that subjects took significantly less time to learn the sections that contained few propositions (mean = 18.827 sec) than those containing many propositions (mean = 21.962 sec) ($t_{(1330)} = -3.526$; $p < 0.001$). Altogether, the data indicated that, the higher the number of propositions, the higher the mean section time. Pairwise comparisons by means of Bonferroni tests revealed that significant differences in the mean section time existed between propositions 5 and 14, 5 and 15, 11 and 12, 12 and 14, 12 and 15, 12 and 16.

7.5.3.4 The effect of the number of objects

The number of objects shown in the pictures can be considered to be an indicator of the complexity of the pictorial information. The method used for counting the objects in the diverse sections was as follows. Each part of the puzzle was counted as an object. If a part of the puzzle was fitted into another part or other parts, then the aggregated parts were regarded as one object. The number of objects was obtained by counting the objects that were to be seen in the pictures or video clips. The procedure for counting objects in the static pictures, however, differed slightly from that in relation to the video clips. In the static-picture version, there were two pictures in each section. The number of objects in a section was calculated by adding the number of objects in those two pictures. In the moving-picture version, the subjects saw how the parts of the puzzle were assembled in action. In this case, the number of objects in a section was obtained by counting the objects initially shown in the video clip. Tables 2 and 3, respectively, show the number of objects in the static and the moving pictures of each section:

S0	S1	S2	S3	S4	S5	S6	S7	S8	S9
13	6	6	5	3	3	3	3	5	3

Table 2: The number of objects in each section of the static-picture version

S0	S1	S2	S3	S4	S5	S6	S7	S8	S9
12	4	4	3	2	2	2	2	3	2

Table 3: The number of objects in each section of the moving-picture version

Figure 13 gives the mean section time as a function of the number of objects. The effect of the number of objects on the mean section time was significant ($F_{(6; 1325)} = 15.337$; $p < 0.001$). It needs to be borne in mind that the number of objects was dependent on picture mode: objects 5, 6, and 13 occurred only in the static pictures, whereas objects 2, 4, and 12 occurred only in the moving pictures. Therefore, the effect of the number of objects should be analyzed separately for static and for moving pictures.
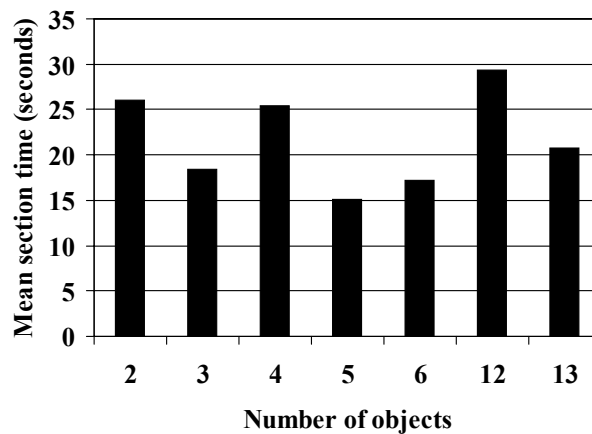
Figure 13: Mean section time as a function of the number of objects

By taking picture mode into account, the data showed that, for static pictures, there was a strong tendency (F(3; 791) = 2.397; p = 0.067) towards a positive relationship between the number of objects and the mean section time. For moving pictures, the number of objects had a significant effect on the mean section time (F(3; 533) = 3.121; p < 0.05). The mean section time increased as the number of objects increased. Pairwise comparisons by means of Bonferroni tests indicated that significant differences existed between objects 2 and 3, 2 and 5, 2 and 6, 3 and 4, 3 and 12, 4 and 5, 4 and 6, 5 and 12, 6 and 12.

7.5.3.5 The effect of the number of visits to the same section

Finally, the number of visits to the same section, i.e. the frequency with which a subject viewed the same section, had a significant effect on the mean section time (F(11; 1320) = 74.827; p < 0.001). The data are given in Figure 14. The mean section time decreased as the number of visits increased. It appears that subjects spent more time processing information during their first encounter with the learning materials. In their successive visits to the same section, subjects possibly concentrated only on the information they wanted to rehearse in order to memorize the materials.
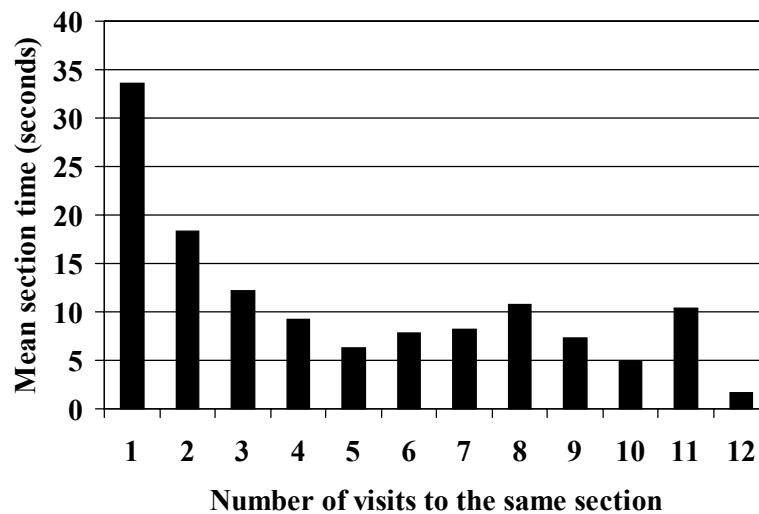
Figure 14: Mean section time as a function of the number of
visits to the same section

7.5.4 Mean section frequency

This variable is a measure of how often a section was viewed by the subjects on average. Figure 15 shows the mean section frequency as a function of experimental condition by section. The results of the ANOVA indicated that the mean section frequency was independent of the experimental condition. The simple factor effect of picture mode had a significant effect on the mean section frequency ($F(1; 36) = 78.733$; $p < 0.001$), in that all of the sections were viewed more frequently when *static pictures* were shown. Text mode had no effect on the mean section frequency.

It should be pointed out that the difference in the mean section frequency between static and moving pictures was most pronounced for Sections 4 and 5. This could be due to the fact that the texts of Sections 4 and 5 described a particular step in the assembly procedure, which could not be depicted very well by static pictures. In S5, the puzzle parts that were fitted together in S4 had to be first of all turned 90° to the left. After that, the rest of the puzzle parts could be assembled. Since the perspective for depicting the cube changed from S4 to S5, subjects who could not easily follow the change shown in the static pictures conceivably had to view S4 and S5 more often to understand or integrate the information.
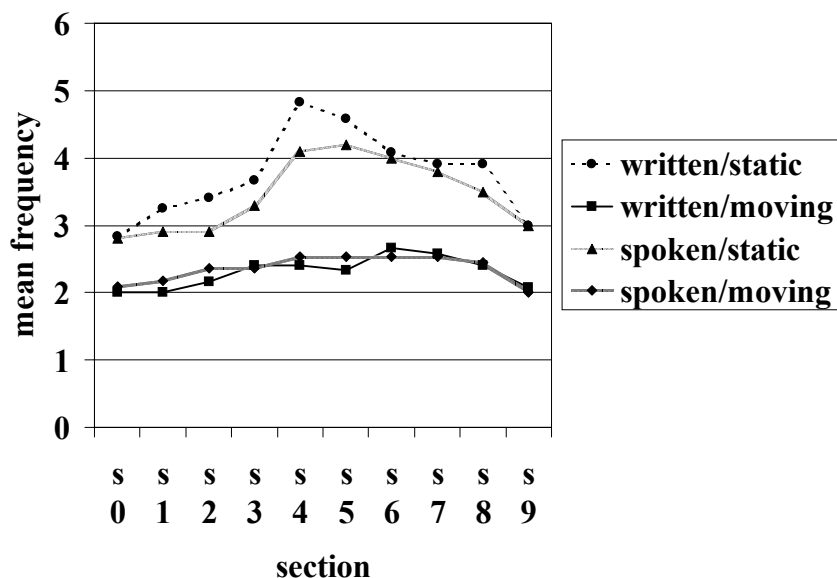
Figure 15: Mean section frequency as a function of experimental
condition, separately by section

The number of propositions had no effect on the mean section frequency. In
contrast, the number of objects did have such an effect ($F_{(6; 33)} = 6.609$; $p < 0.001$).
Figure 16 depicts the mean section frequency as a function of the number of objects.
Again, the presentation mode of pictures has to be considered because it affects the
interpretation of the data. The results showed that for both static ($t(18) = 3.626$; $p <
0.005$) and moving pictures ($t(18) = 3.395$, $p < 0.005$), the mean section frequency
decreased as the number of objects increased.

Figure 16: Mean section frequency as a function of the number of objects

7.5.5 Analyses of eye-movement data

7.5.5.1 Number of fixations

Figure 18 shows the mean number of fixations as a function of experimental condition. The results of the ANOVAs showed that text mode ($F_{(1; 1328)} = 31.301$; $p < 0.001$), picture mode ($F_{(1; 1328)} = 14.036$; $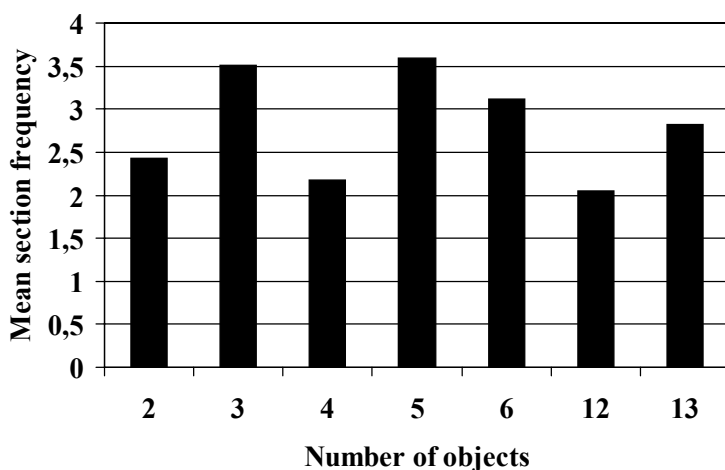p < 0.001$), and the interaction between text mode and picture mode ($F_{(1; 1328)} = 48.190$; $p < 0.001$) had significant effects on the mean number of fixations. The mean number of fixations was larger when written text rather than spoken text was presented, and when moving rather than static pictures were presented. Pairwise comparisons using Bonferroni tests showed that the mean number of fixations for Condition (written/moving) was substantially larger than that for Conditions (written/static), (spoken/static), and (spoken/moving).

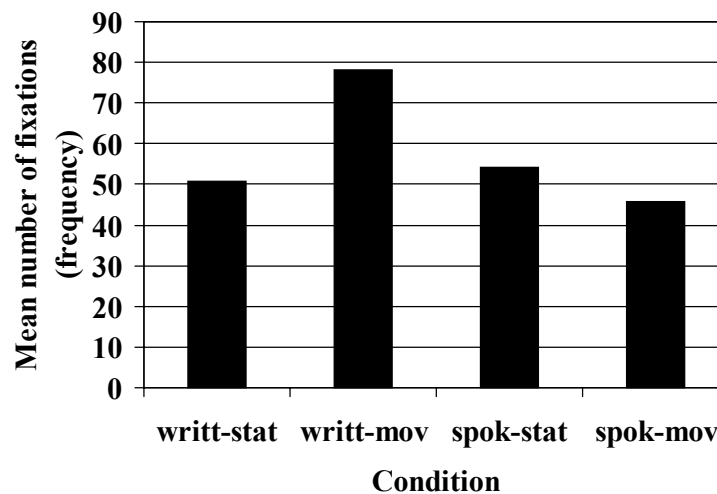Figure 18: Mean number of fixations as a function of experimental condition

7.5.5.1.1 Between conditions analyses

Figure 19 depicts the mean number of fixations for each instructional section for each of the four experimental conditions. When the mean numbers of fixations between conditions were compared section by section, significant differences between conditions were found in S0, S1, S2, S4, S5, S7 and S8.
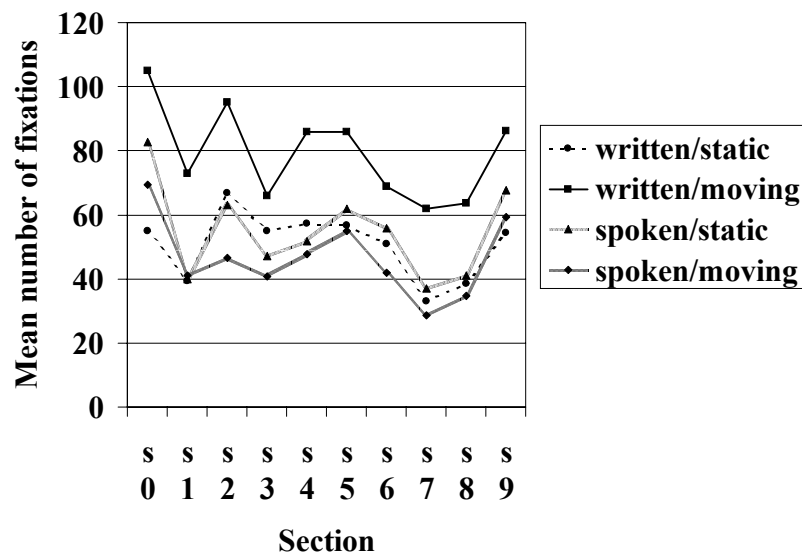
Figure 19: Mean number of fixations as a function of experimental
condition, by section

7.5.5.1.2 Within-condition analyses

In the within-condition analyses, significant differences in the number of fixations between sections were found for all experimental conditions: Condition (written/static) ($F_{(9; 440)} = 2.028$; $p < 0.05$); Condition (written/moving) ($F_{(9; 267)} = 1.982$; $p < 0.05$); Condition (spoken/static) ($F_{(9; 335)} = 3.813$; $p < 0.001$); and Condition (spoken/moving) ($F_{(9; 250)} = 4.243$; $p < 0.001$). These differences were more likely to result from the number of objects rather than from the number of propositions, because some differences existed between the sections that had the same number of propositions.

7.5.5.1.3 The effect of the number of propositions

The number of propositions exerted a significant influence on the mean number of fixations ($F_{(5; 1326)} = 6.696$; $p < 0.001$) (see Figure 20). Due to the large variance, it again seemed sensible to compare sections with few propositions with sections with many propositions in order to assess the general nature of the relationship. A t-test ($t_{(1330)} = -2.448$; $p < 0.05$) indicated that, on the whole, the mean number of fixations was positively related to the number of propositions.
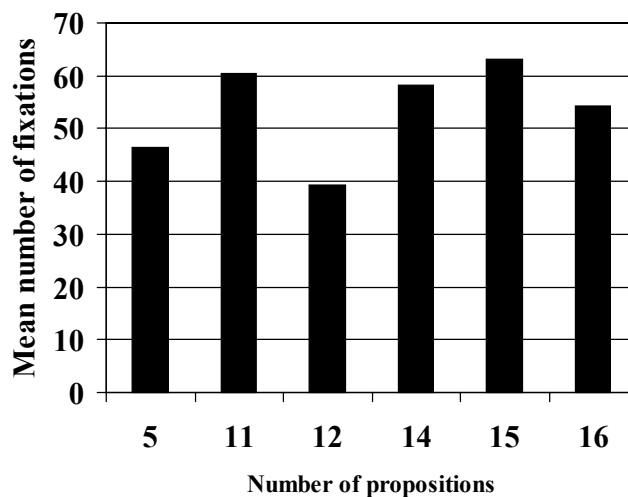
Figure 20: Mean number of fixations as a function of the
number of propositions

7.5.5.1.4 The effect of the number of objects

The data are given in Figure 21. The mean number of fixations also depended on the number of objects ($F_{(6; 1325)}$ = 7.598; $p < 0.001$). Specifically, pairwise comparisons by means of Bonferroni tests showed that significant differences in the mean number of fixations existed between sections with 2 and 5, 2 and 12, 3 and 12, 4 and 5, 5 and 12, 5 and 13, and 6 and 12 objects, respectively. Again, in regard to picture mode, a series of t-tests (static pictures: $t(793)$ = -1.761; $p = 0.07$; moving pictures: $t(535)$ = -2.725; $p < 0.01$) indicated that the general effect of the number of objects on the mean number of fixations was positive: the larger the number of objects, the larger the number of fixations.
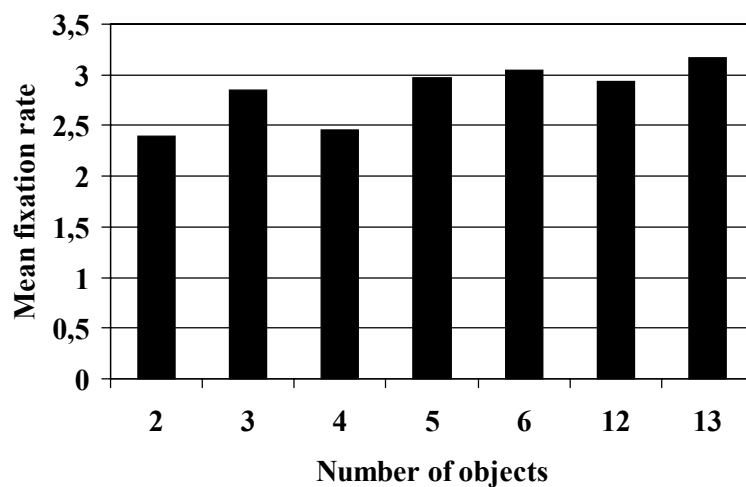
Figure 21: Mean number of fixations as a function of the number of objects

7.5.5.1.5 The effect of the number of visits to the same section

Figure 22 shows the data. The number of visits to the same section also had a significant effect on the mean number of fixations ($F_{(11; 1320)} = 69.265$; $p < 0.001$).
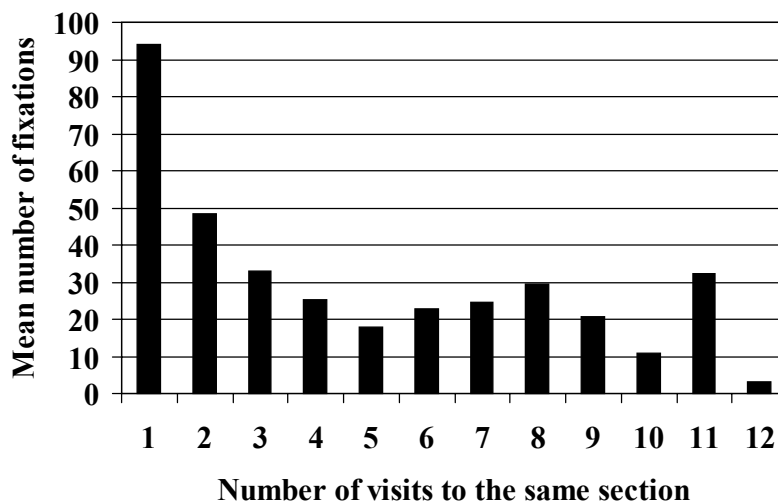
Figure 22: Mean number of fixations as a function of the
number of visits to the same section

The mean number of fixations decreased with an increase in the number of visits to the same section. According to Bonferroni tests, the mean number of fixations during the first and the second visits was significantly higher than that during the third visit or later.

7.5.5.2 Fixation duration in the text and in the picture region

The fixation duration in the text region was independent of the experimental condition. Figure 23 shows the mean fixation duration in the picture region as a function of experimental condition. The effect of text mode was not significant, whereas that of picture mode was significant ($F_{(1; 44)} = 32.610$; $p < 0.001$). The mean fixation duration in the picture region was significantly longer when moving pictures were presented. There was a strong tendency towards a text mode × picture mode interaction ($F_{(1; 44)} = 3.517$; $p = 0.067$). Pairwise comparisons by means of t-tests showed that the fixation duration in the region for moving pictures was significantly longer when text was presented auditorily ($t_{(22)} = 2.588$; $p < 0.05$).
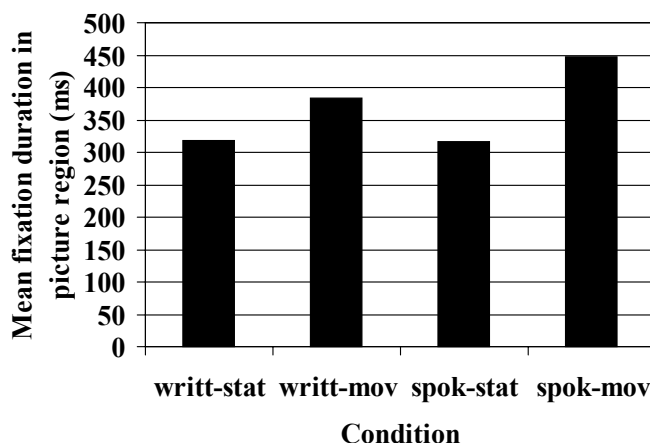
Figure 23: Mean fixation duration in the picture region as
a function of experimental condition

7.5.5.3 Fixation rate

The fixation rate was defined as the number of fixations per second. The data are shown in Figure 24. According to the ANOVAs, text mode ($F(1; 1328) = 181.924$; $p < 0.001$) as well as picture mode ($F(1; 1328) = 184.484$; $p < 0.001$) had a significant effect on the mean fixation rate. The interaction between text mode and picture mode was not significant. That is, the mean fixation rate was independent of the experimental condition.
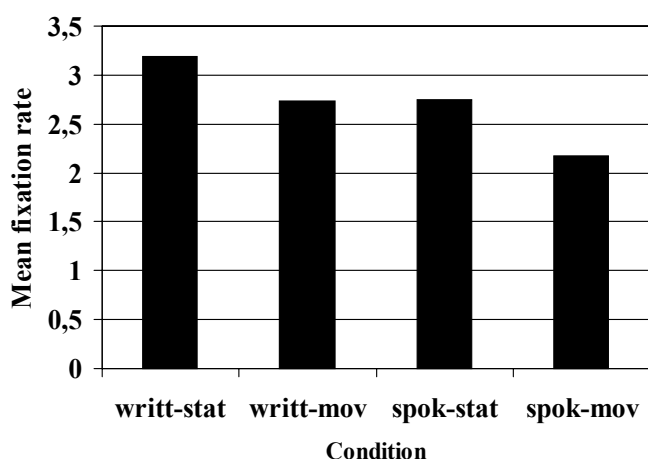
Figure 24: Mean fixation rate as a function of experimental condition

The mean fixation rate was higher when static rather than moving pictures were displayed, and when written rather than spoken texts were presented. Pairwise comparisons by means of Bonferroni tests indicated that the mean fixation rate of Condition (written/static) was significantly higher than those of the other conditions,

whereas the mean fixation rate of Condition (spoken/moving) was significantly lower than those of the other conditions.

7.5.5.3.1 Within-condition analyses

In the within-condition analyses significant differences between sections were found only in Condition (written/moving) ($F(9; 267) = 3.739$; $p < 0.001$) and Condition (spoken/moving) ($F(9; 250) = 2.368$; $p < 0.05$). For Condition (written/moving), the Bonferroni tests showed that the mean fixation rate in S0 was significantly higher than that in S2, S4, S5, S6, S7, or S8. Likewise, for Condition (spoken/moving), the mean fixation rate in S0 was substantially higher than that in S3, S4, or S5. Figure 25 presents the data.
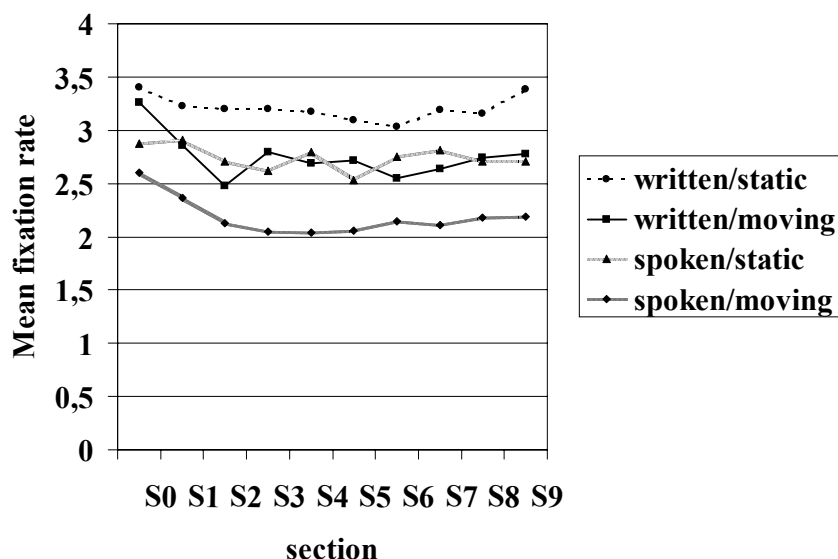
Figure 25: Mean fixation rate as a function of conditions, by section

7.5.5.3.2 The effect of the number of propositions

Figure 26 shows the mean fixation rate as a function of the number of propositions. The number of propositions had a significant effect on the mean fixation rate ($F(5; 1326) = 2.764$; $p < 0.05$). By comparing sections with few (5, 11, 12) propositions with sections with many (14, 15, 16) propositions, the results indicated that the mean fixation rate was inversely related to the number of propositions in general ($t(1330) = 3.205$; $p < 0.005$).
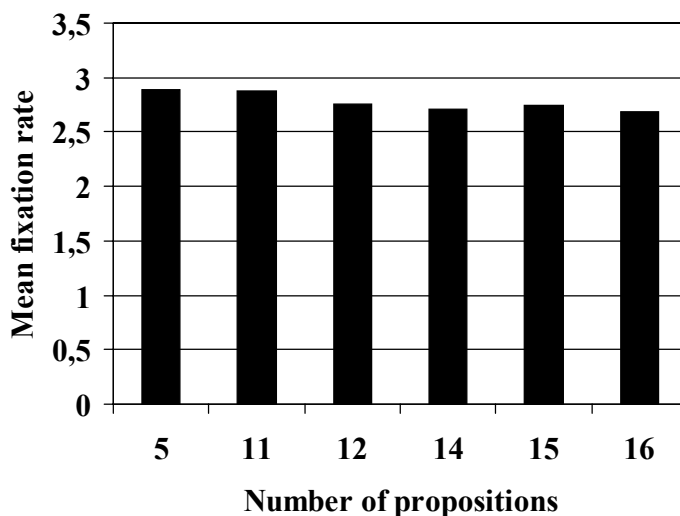
Figure 26: Mean fixation rate as a function of the number of propositions

7.5.5.3.3 The effect of the number of objects

Figure 27 shows the mean fixation rate as a function of the number of objects. The number of objects had a significant effect on the mean fixation rate ($F_{(6; 1325)} = 26.197$; $p < 0.001$). Again, picture mode was considered in the analysis of the general effect of the number of objects. The results showed that the mean fixation rate was positively related to the number of objects in general. For static pictures, there was a tendency towards the existence of a positive relationship between the mean fixation rate and the number of objects ($t(793) = -1.81$; $p = 0.07$). For moving pictures, the mean fixation rate increased as the number of objects increased ($t(535) = -3.134$; $p < 0.005$).
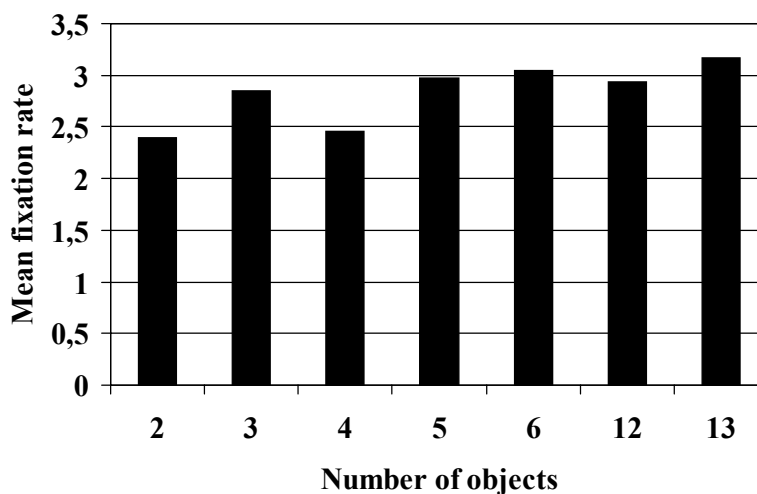
Figure 27: Mean fixation rate as a function of the number of objects

7.5.5.3.4 The effect of the number of visits to the same section

The number of visits to the same section did not exert a significant influence on the mean fixation rate. Within-condition analyses, however, revealed an influence of the number of visits to the same section for Conditions (written/static) and (spoken/static) but not for Conditions (written/moving) and (spoken/moving). Figure 28 shows the mean fixation rate as a function of the number of visits to the same section as a function of experimental condition. The Bonferroni tests showed that, for Condition (written/static), the fixation rate during the first visit was higher than that during the second and the third visit. For Condition (spoken/static), the fixation rate during the first visit was higher than that during the fourth and the fifth visit; in addition, the fixation rate during the second and the third visit was higher than that during the fifth visit.
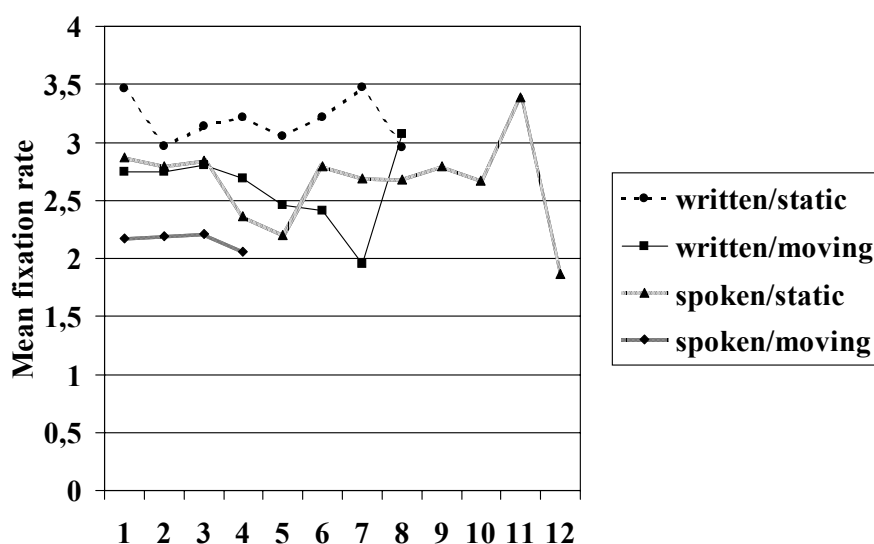
Figure 28: Mean fixation rate as a function of experimental condition,
by the number of visits to the same section

7.5.5.4 Number of gaze changes between the text and picture regions

It goes without saying that the analysis of gaze changes between the text and picture areas of the instructional display can be performed only for those conditions that include written texts, i.e. Condition (written/static) and Condition (written/moving). A t-test did not show any significant difference between Condition (written/static) and Condition (written/moving) in terms of the number of gaze changes between the text and picture regions.

7.5.5.5 The percentage of time spent in the picture versus the text region

The data are given in Figure 29. Again, this comparison could only involve Conditions (written/static) and (written/moving). First, it must be pointed out that the percentage of time subjects spent viewing the text and the pictures did not add up to 100% because subjects also spent some time looking at the other components on the instructional display, such as the control bar and the forward and backward buttons. According to the t-tests, there was no significant difference between the two conditions in terms of the percentage of time spent in the picture region, whereas the difference in terms of the percentage of time spent in the text region was significant ($t(22) = -2.504$; $p < 0.05$): Subjects in Condition (written/static) spent a significantly larger proportion of their time reading the text than did subjects in Condition (written/moving).
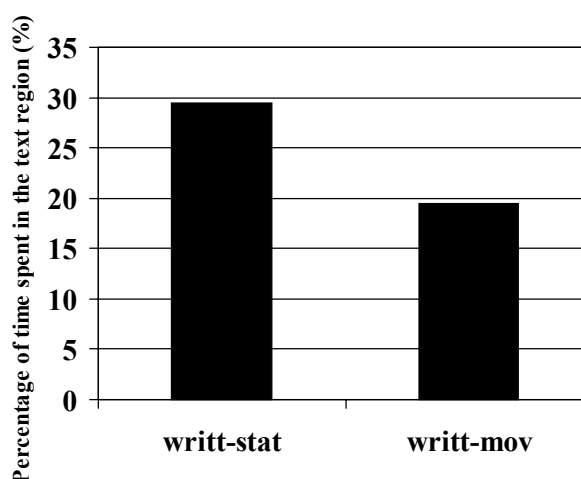
Figure 29: Percentage of time spent in the text region as a function of experimental conditions involving written texts

7.5.6 Questionnaire results

In addition to measuring subjects' learning efficiency and eye-movement behavior, I also inquired into subjects' rating of the comprehensibility of texts and pictures. The rating was conducted on the basis of a five-point scale: 1 for very difficult, 2 for difficult, 3 for medium, 4 for easy, and 5 for very easy. The results showed that 47.917% of the subjects rated text comprehensibility as easy (median = 4), while 60.417% of the subjects rated picture comprehensibility as easy (median = 4) as well. After considering the presentation modes for both texts and pictures, the results indicated that most of the subjects in written-text conditions rated text

comprehensibility between medium and easy (median = 3.5; 41.667% for 'medium'; 45.833% for 'easy'), whereas most of the subjects in spoken-text conditions rated its comprehensibility as easy (50%; median = 4). In contrast to text comprehensibility, picture comprehensibility was rated as easy (median = 4), both in the static- and the moving-picture conditions.

As to the question of whether the texts, the pictures or both of them were the crucial elements for understanding, the results showed that 58.333% of the subjects regarded pictures as essential elements for understanding the instructions, and 87.5% of the subjects regarded pictures as crucial elements for memorizing the instructions. Furthermore, 58.333% of the subjects in spoken-text conditions considered both texts and pictures as crucial to understanding, whereas 79.167% of the subjects in written-text conditions considered pictures as essential to understanding. 75% of the subjects in moving-picture conditions regarded pictures as the crucial element for understanding, and 54.167% of the subjects in static-picture conditions regarded both texts and pictures as crucial. The medium that subjects chose to support the memorization of instructions was rather consistent. Most of the subjects chose pictures as the primary memory aid regardless of the text and picture presentation modes adopted in the instructions.

7.6 Discussion

The main objective of this experiment was to investigate the influence of MPs on learning efficiency in a multimedia-based learning scenario. In addition, subjects' eye-movement behavior during learning was also observed to facilitate a more detailed analysis of information-processing strategies. In the following sections, I will start with a discussion of the connection between MPs and learning efficiency. Later, I shall discuss the information-processing strategies and the eye-movement behavior.

7.6.1 MPs and learning efficiency

It was hypothesized that moving pictures are superior to static pictures in presenting information about sequences of actions (procedural tasks), movements of objects, or changes in a state of affairs over time, and that, if the modality effect is true, presenting information in different modalities should facilitate the efficiency of information processing. Consequently, instructions comprising moving pictures and spoken texts should lead to the highest learning efficiency. The results indeed

showed that subjects in Condition [spoken/moving] achieved the highest learning efficiency. However, this result did not directly confirm the hypotheses mentioned above because it should be interpreted in line with a tendency toward a text mode × picture mode interaction.

In terms of the learning efficiency, the result was not consistent with the prediction that moving pictures are superior to static pictures because picture mode did not exert an effect on learning efficiency. The reasons for this result could be: first, there was no time limit for learning, whereby the effect of moving pictures was possibly weakened; second, the procedures for assembling the puzzle were sliced into several sections, and most of the sections did not comprise complicated actions except Sections 4 and 5. As a consequence, static pictures were, for the most part, as effective as moving pictures in demonstrating the assembly procedures.

Nevertheless, when the mean section frequency was considered, the advantages of moving pictures were revealed to be as follows: subjects in moving-picture conditions repeated the instructional sections less often than did subjects who received static pictures. It appears that information conveyed by moving pictures was easier to remember. Besides, a local effect was found in Sections 4 and 5 where subjects in static-picture conditions substantially repeated these two sections more frequently because more complicated assembly procedures and rotations of the puzzle parts were involved. Subjects processing static pictures had to mentally simulate the operations and rotate some puzzle parts in order to understand the instructions properly. In other words, complicated inferences had to be drawn, which led to higher section frequency for the two sections. Moving pictures, in contrast, could easily demonstrate the operations and the rotations of the puzzle parts as a continuous action. As a result, subjects could easily read off the actions they had to perform in relation to the puzzle parts from moving pictures, which was thus more efficient than inferring the actions from static pictures that only depicted the puzzle states before and after the operation.

Furthermore, the hypothesis that the dual-modality presentation leads to higher information-processing efficiency was only partly confirmed in this experiment. The superiority of the dual-modality presentation over the single-modality one was only restricted to moving-picture conditions. When moving pictures and written texts were shown, subjects failed to process both sources of information simultaneously. The video recordings of subjects' eye movements indicated that they were actually not able to read the text and watch the video at the same time, and they could hardly switch their gaze positions between the text and picture regions while the video clip was playing. Thus, the verbal and pictorial information was processed successively, which led to low learning efficiency.

Overall, the results with regard to the learning efficiency were consistent with the hypothesis that subjects should learn most efficiently if the learning materials are presented in the form of spoken texts and moving pictures rather than written texts and moving pictures. However, this result, in my opinion, is not necessarily attributed to the split-attention effect because the fact that moving pictures with written texts led to lower learning efficiency could also be explained by the strong structural interference induced by the given MP, which simply impeded the processes of text-picture integration.

Furthermore, it is interesting that subjects who received static pictures did not benefit from the dual-modality presentation at all. The eye-movement data revealed that subjects who had to process written text with static pictures integrated text and picture information by means of relatively rapid eye movements between text and pictures, which led to the same learning efficiency as achieved by subjects who had to process spoken text with static pictures. This could result from the fact that the pictorial information in this case was relatively easy to process. Thus, it did not impose an overload on the visuo-spatial sketchpad, so that working memory (or CE) could easily integrate the verbal and the pictorial information. Alternatively, the findings could also come about if the pictorial information was very difficult to process. Based on the study conducted by Sweller et al. (1997), if the information conveyed in a diagram is visually demanding, there are not enough resources available in working memory to integrate the information from pictures and spoken texts. As a consequence, the dual-modality presentation is not superior to the single-modality one. Thus to find out under which circumstances learners can really benefit from a dual-modality presentation, further research is required. A follow-up experiment is designed to address this issue.

7.6.2 Information-processing strategies

In an attempt to reconstruct subjects' information-processing strategies, I examined the effects of the following factors on the mean learning time, the mean section time, and the mean section frequency: 1) the four experimental conditions, including the analyses of simple factor effects—text mode and picture mode; 2) the number of propositions in the texts; 3) the number of objects in the pictures (separately for static vs. moving pictures); and 4) the number of visits to the same section.

The learning time in relation to Condition (spoken/moving) was the shortest among the four experimental conditions, whereas the learning efficiency for this condition was the highest. This result indicates that spoken texts combined with moving pictures, in comparison to the other MP-conditions, could convey the

information to be learned most efficiently. Since there were no significant differences in the mean learning time and in the mean learning efficiency for the other conditions, the analyses of the mean section time and mean section frequency may yield more detailed information about the differences in the information-processing strategies and the amount of cognitive effort underlying those strategies.

The mean section time was affected by different MPs: For instance, the mean section time was longer when written texts or moving pictures were presented. Subjects in Condition (written/moving) took more time to process the information presented in a section than did subjects in the other conditions. Since subjects in Condition (written/moving) could not watch the video and read the text simultaneously, a long section time did not come as a surprise. However, the mean section time for Condition (written/static) was the shortest among the conditions, though subjects could not read the text and view the pictures at the same time either. When comparing the mean fixation rate and the mean section frequency of the four conditions, it was found that subjects in Condition (written/static) had the highest mean fixation rate and the highest mean section frequency, whereas subjects in Condition (written/moving) had a lower mean fixation rate and almost the lowest mean section frequency. This indicates that subjects in Condition (written/static), who were by no means constrained in terms of the time they allocated to the learning materials, required many brief fixations to process the information presented in a section, including frequent gaze changes between the text and the static pictures; however, they had to return to the sections more frequently in order to memorize the information.

It should be noted that the mean section time for subjects in Conditions (written/moving), (spoken/moving), and (spoken/static) was to a certain extent contingent upon the time required for playing the video clips or the speech track. In other words, there was a 'base time' which, for technical reasons, could not be shortened. This most probably resulted in a longer mean section time for Conditions (written/moving), (spoken/static), and (spoken/moving). It follows that the mean section time alone does not constitute a valid parameter for information processing, but must rather be considered in the light of other, additional variables such as the mean section frequency and some eye-movement variables. Consider the relationship between the mean section time and the mean section frequency. The mean section frequency was negatively correlated with the mean section time ($r = -0.583$; $p < 0.001$). That is, although the subjects in Conditions (written/moving), (spoken/static), and (spoken/moving) required more time to view a section, they did not repeat each section as often as the subjects in Condition (written/static). That is the reason why

the mean learning time in relation to Conditions (written/static), (written/moving), and (spoken/static) did not differ.

Apart from the experimental conditions, the number of propositions, the number of objects and the number of visits to the same section were also analyzed as factors affecting the mean section time and the mean section frequency. Both the mean section time and the mean number of fixations were positively related to the number of propositions and the number of objects because the more information that was to be processed in a section, the longer was the period of time that subjects needed to spend on that section, and subjects would require more fixations to scan the information. The mean section frequency, on the other hand, was independent of the number of propositions but inversely related to the number of objects. In line with this result, I have assumed that the mean section frequency was not determined by the quantity of the information to be processed but rather by the quality of that information. In terms of the number of objects, a reason for this could be that S0, which gave the introduction to the learning materials, comprised the largest number of objects. It was, however, repeated the least because it was not quite relevant for learning in this case.

As to the number of visits to the same section, this was inversely related to the mean section time and the mean number of fixations. Specifically, the first two visits required more time and more fixations when compared with the later ones. Conceivably, subjects spent more time during the first two visits because they were dealing with new information. However, during later visits, the information could be accommodated under an already established schema. In other words, the main objective of the later visits was to rehearse the learning materials. Based on the assumption that the rehearsal should require less cognitive effort than the establishment of a new cognitive schema required during the first two visits and that the cognitive effort should have gradually declined as the number of rehearsals increased, a decrease in the mean section time and the mean number of fixations could be expected as the number of visits to the same section increased.

7.6.3 Interpretation of eye-movement data

a) Number of fixations:

The mean number of fixations was larger when visual text or moving pictures were presented. This result is mainly due to Condition (written/moving) where subjects received written text with moving pictures. The mean number of fixations for Condition (written/moving) was particularly high because subjects usually read the text and watched the video clips successively and because more time was

needed to watch a video clip being played than to inspect static pictures. Furthermore, the number of fixations was positively related to the number of propositions as well as to the number of objects shown in the pictures. Both the number of propositions and the number of objects are parameters of the amount of information to be processed. Thus, the data are consistent with the hypothesis that the more information that is to be processed, the more fixations that are required. In addition, the number of fixations was inversely related to the number of visits to the same section, which indicates that, during repeated visits to a section, subjects probably concentrated only on viewing the relevant information in order to learn it by heart. Hence, they used less time as well as fewer fixations to view each instructional section.

b) Mean fixation duration in the picture region:

The mean fixation duration in relation to moving pictures was longer than that in relation to static pictures. This might indicate that more intensive information processing was involved when subjects perceived moving pictures. However, this phenomenon could also result from the fundamental differences in eye movements in exploring the static and the moving pictures in this experiment. Since subjects did not often have to move their eyes in order to follow the actions shown in the video clips, their fixation duration could be prolonged accordingly. In contrast, when subjects were viewing static pictures, they had to move their eyes quite often in order to pick up the information scattered in the pictures. Therefore, fixation duration in this case tended to be short.

Nevertheless, another possibility that cannot be ruled out is that the eye-tracker might not be able to measure very slight eye movements (smooth pursuit) precisely, and it may tend to merge the fixations in rapid succession. Thus the interpretation of long fixation duration in moving pictures is actually very problematic. What this comes down to is that, with moving pictures, fixation duration cannot be regarded as a valid indicator of keeping a moving object in focus.

Moreover, the mean fixation duration in relation to moving pictures was longer when text was presented auditorily rather than visually. The reason why the mean fixation duration was significantly shorter when written text was presented was that subjects often switched their gaze between the text and the video regions when the video was not playing. As a consequence, part of the fixations in the video region were very short when the gaze positions changed rapidly between the written text and the still video element, which in turn led to a

shorter mean fixation duration in relation to this condition. This was not the case when text was presented auditorily.

c) Fixation rate:

The mean fixation rate was higher when written texts or static pictures were presented. In the first case, subjects often switched their gaze positions between text and pictures in order to integrate the verbal and pictorial information. In the second case, the subjects' eyes were free to inspect the photographs. The speed of the subjects' eye movements was therefore higher because it was not constrained by any movement shown in the visual display.

The effects of the number of propositions as well as the number of objects on the fixation rate gave rise to interesting insights regarding the relationships between mental effort and eye-movements. The fixation rate was inversely related to the number of propositions, but positively related to the number of objects. That is, if a scene contained many objects, subjects tended to scan these objects using many brief fixations. In contrast, if a text contained many propositions, subjects tended to process it by means of a few long fixations. Conceivably, a part of the mental resources was devoted to setting up the propositional text base, and, therefore, the speed of eye movements was reduced.

d) Number of gaze changes between the text and the picture regions:

The number of gaze position changes between the text and picture regions did not vary between conditions. However, the video recording of the stimuli that were overlaid with the subjects' gaze trajectories revealed that subjects seldom shifted their gaze positions between the text and the video while the video was still playing. Typically, subjects viewed the video clip and read the text separately, but often switched their gaze positions between the text and the video when the video was not playing, i.e. when the video stopped in its last frame after it had been played. Although the last frame of the video only showed the final state of the puzzle parts assembled according to a certain procedure, subjects tended to use it as an aid to establish referential links between the text and the video information.

The eye-movement behavior of subjects in Condition (written/moving) was different from that observed by d'Ydewalle et al. (1987) in their studies in which subjects were asked to watch a movie. d'Ydewalle et al. found that, despite the fact that a partial loss of the information from the image would be caused by alternating between viewing the image and reading the subtitles, subjects automatically read the subtitles regardless of whether this was necessary for understanding the movie. Obviously, people use different strategies to integrate

verbal and pictorial information depending on the situation. When watching TV or seeing a movie, a partial loss of the pictorial information can easily be compensated for; it in no way impairs the understanding of the movie. In an experiment in which subjects are required to learn multimedia instructions with little allowance for redundancy, skipping any pictorial or verbal information can be a serious obstacle to comprehension and, therefore, may impair learning. As a consequence, subjects in this study did not alternate between watching the video and reading the text.

e) The percentage of time spent in the picture versus the text region:

There was a difference between Condition (written/static) and Condition (written/moving) in terms of the percentage of time spent in the text region: subjects in Condition (written/static) spent proportionally more time reading the text than did subjects in Condition (written/moving). This indicates that subjects in Condition (written/static) paid more attention to the textual information than did subjects in Condition (written/moving). Conceivably, subjects in Condition (written/moving) relied less on the information presented by the written texts during learning, because the integration of text and video information was difficult per se. The subjects therefore preferred studying the videos to the texts. In fact, the preferences of those subjects were confirmed by their answers to the questionnaire: all of the subjects used the videos as the crucial element in understanding the instructions, and almost all of the subjects used the videos to memorize the instructions.

7.6.4 Observations based on the questionnaire results

Subjects who had been given moving pictures as instructions regarded pictures as the crucial element for understanding, whereas subjects who had been given static pictures as instructions regarded both the texts and the pictures as crucial. Subjects who received spoken texts regarded both texts and pictures as crucial elements for understanding, whereas subjects who received written texts regarded pictures as the crucial element for understanding. The medium that most of the subjects used to memorize the instructions was pictures.

7.6.5 Further issues to be addressed

According to the results of Experiment 1, it is still unclear whether the split-attention effect and the modality effect are valid. To further investigate this issue, I attempt to

reexamine those effects from the perspective of Baddeley's working memory model. As I have argued in Chapter 5, the claims made in relation to the split-attention effect and the modality effect are based on an incorrect interpretation of the processes of information processing in working memory. Both dual coding theory and Baddeley's working memory model do not support such claims. Consider, for instance, Baddeley's working memory model. Presenting both verbal and pictorial information visually should not overburden the visuo-spatial sketchpad because verbal information should not be processed by the visuo-spatial sketchpad but rather by the phonological loop. On the contrary, it is assumed that the dual-modality presentation would be more likely to induce a split-attention effect in comparison with the single-modality presentation because the attention will have to split only when different sources of information are processed at the same time. With dual-modality presentation, working memory has to process verbal and pictorial information simultaneously. This is not the case with single-modality presentation. In the next chapter, I shall report on a follow-up experiment designed to examine the validity of the split-attention and dual modality effects.

8. EXPERIMENT 2

The objective of this experiment is to examine the role of working memory in learning in terms of the way in which the learning materials are presented. I attempt to find out: 1) under which circumstances the dual-modality (auditory plus visual) presentation (DMP) is indeed superior to the single-modality (only visual) presentation (SMP) when learning materials are presented in the form of diagrams (static pictures) combined with visual or auditory text; and 2) whether using animation as a guide for visual attention and/or as an aid for constructing mental models can successfully reduce the load imposed by the complex diagrams on the visuo-spatial sketchpad, and therefore release working memory (or CE, respectively) from being overloaded by DMP.

8.1 Hypotheses

According to the cognitive load theory, learning does not occur if the capacity of working memory is exceeded. The overload of working memory may either result from the complexity of the content of information or the inappropriate display of learning material. Based on that, if the complexity of the content of information is held at a non-demanding level while the presentation mode of information varies in complexity, the connection between the cognitive load engendered by the instructional display and learning efficiency can be better revealed.

By keeping the complexity of verbal information at the same (non-demanding) level, I assume that DMP can only be beneficial if the visual information is neither very low- nor very high-demanding. Since the visual (pictorial) and the auditory (verbal) information should be integrated by the central executive unit (CE) in working memory before they are forwarded to long-term memory, the processes of integrating information can be impaired if CE is overloaded. Consequently, if the pictorial information is highly complicated, DMP, of course, is no better than SMP given that both presentations would overburden working memory anyway. However, another possibility that should not be ruled out is that DMP is more likely to overload working memory than SMP because the former actually imposes a heavier load on CE by inducing it to process visual and auditory information simultaneously. Secondly, if the diagram is very easy, DMP is not necessarily superior to SMP, because the total visual demand in relation to both conditions is low. For the SMP

condition, I assume that subjects can use rapid eye movements to coordinate textual and pictorial information, and thereby, perform as well as the subjects in the DMP condition.

Moreover, I expect that using animation can be advantageous to learning 1) if the information to be learned is involved in sequences of action (e.g., movements of objects) or alteration of a state of affairs, etc., 2) if the visual display of the learning material is visually highly-demanding, and 3) if the animation is presented with auditory text together. In the first case, animation should help learners to build mental models required for understanding the information. In the second case, animation should be able to guide learners' visual attention, and as a consequence, can lessen the cognitive load on the visuo-spatial sketchpad. Finally, in the third case, animation combined with written text should produce strong structural interference (one cannot view an animation and read a text at the same time) and diminished learning performance, whereas this is not the case when animation is synchronized with auditory text.

As to the eye-movement behavior, I predict that picture complexity or the number of objects, respectively, will influence the speed of learners' eye movements. The more objects that are depicted in a picture, the faster the eye movements should be because this is a more efficient way of scanning the scattered objects that convey the relevant pictorial information required for the comprehension of the instructions. Accordingly, the large number of fixations, the short fixation duration, the high fixation rate, and long saccades are associated with rapid eye movements when subjects inspect a static picture with a large number of objects. When animation is presented, I assume that visual distractors in this case should not distract learners' visual attention very much, since learners are more likely to follow the motion and concentrate on the relevant information than pay attention to any irrelevant objects in the picture. Thus, the number of fixations, the fixation rate, and saccade length are not necessarily related to the number of objects.

Text mode, on the other hand, should affect the speed of eye movements as well. I assume that the fixation rate should be lower if auditory text is presented because learners have to process the verbal and pictorial information simultaneously. The attentional resource will be split during information processing, and, therefore, eye movements should slow down. This can be revealed by longer fixation durations, a lower fixation rate, and shorter saccade length compared to those experimental conditions in which text is presented visually.

The percentage of time that learners spend viewing a picture should be positively related to the number of objects, since the more information that is to be processed in the picture, the more learning time is required. Moreover, the number

of gazes changing between the text and the picture regions should be influenced by the picture complexity as well as by the number of propositions. As picture complexity intensifies, the number of gazes switching between the text and picture regions should increase because the integration of the textual and pictorial information is hindered by complex pictorial information. More changing of gazes can make the establishment of referential connections between textual and pictorial information easier. Similarly, the number of gaze changes should be positively related to the number of propositions, since the more information that is conveyed by the text, the less likely it is that learners will remember all of the textual information at one time. Hence, the integration of text and picture information can be carried out more easily if subjects frequently switch gazes between the text and the pictures. When animation is presented, it is difficult to integrate textual and pictorial information by switching gazes because learners will always lose some information in animation when they are reading the text. Therefore, there should be less gaze changes between the text and picture regions when animation is displayed.

8.2 Variables and Design

8.2.1 Independent variables

The effects of picture complexity and text mode, especially the interaction of these two factors on learning were investigated. A 4 (*picture complexity*) by 2 (*text mode*) factorial design was applied. The factor picture complexity was varied in three levels – *simple, medium, and complex* – for the static pictures, and in an additional level – *animation*. Certainly, it is very difficult to define the criteria for judging picture complexity because a multitude of factors (both qualitative and quantitative) may exert an influence on picture complexity. The criteria I employed in this experiment were the number of objects shown in the picture and the visual complexity of the objects per se. In the simple diagrams, only simple objects that were absolutely necessary for explaining the information to be learned were shown in the picture. In the medium diagrams, a larger number of simple objects were used for depicting the same information. Additionally, five to six irrelevant simple objects were added to the picture to intensify the pictorial complexity. Their function should be regarded as that of visual distractors. In the complex diagrams, not only visually more complicated objects but also more visual distractors were employed. Finally, in the additional level, I used animation in the complex diagrams to examine whether animation could successfully guide learners visual attention, and thus, reduce the

load on the visuo-spatial sketchpad as well as enhance learning results. Another factor text mode had two levels: *visual* and *auditory.* Together with the four picture types, eight experimental conditions were created:

1. **Condition (simp-visu):**   simple diagrams with visual texts
2. **Condition (simp-audi):**   simple diagrams with auditory texts
3. **Condition (med-visu):**   diagrams with medium complexity combined with visual texts
4. **Condition (med-audi):**   diagrams with medium complexity combined with auditory texts
5. **Condition (comp-visu):**   complex diagrams with visual texts
6. **Condition (comp-audi):**   complex diagrams with auditory texts
7. **Condition (ani-visu):**   animation with visual texts
8. **Condition (ani-audi):**   animation with auditory texts

In addition to picture complexity and text mode, the effects of *the number of propositions* in the texts, and *the number of objects* (including two subsets: *the number of relevant objects* and *the number of visual distractors*) in the pictures were also examined.

It should be noted that subjects' prior knowledge related to the information to be learned in the instruction is likely to affect their learning behavior and learning efficiency. Since the main task of the subjects was to learn the movement rules of seven Chinese-chess pieces, those who had expertise in playing European chess could certainly learn the rules more quickly and easily, independent of the display of the instructions. As a consequence, those who had expertise in playing chess or games involving similar rules were excluded from participation in this experiment.

8.2.2 Dependent variables

8.2.2.1 Learning variables

The dependent variable for assaying learning outcomes was *the error rate*. Besides, *the learning time*, *the task time*, the time subjects spent on studying each instructional section (*the mean section time*), and the frequency of each instructional section that was viewed by subjects (*the mean section frequency*) were investigated in relation to the learning processes.

8.2.2.2 Eye-movement variables

The following eye-movement variables were analyzed:
- The number of fixations
- The fixation duration
- The fixation rate
- The saccade length
- The number of gaze changes between the text and the picture regions
- The percentage of time spent in the picture region and in the text region

8.3 Method

8.3.1 Subjects

The participants (N = 40) were students at the University of Bielefeld between the ages of 18 and 38. All subjects were native German speakers and had no or only a little experience of playing chess or games with similar rules to chess. They were paid for their participation in this experiment.

8.3.2 Materials

The learning material was concerned with the movement rules of Chinese-chess pieces. Since there were seven different pieces in Chinese chess, all of the instructions were divided into 7 instructional sections. Every section was edited in HTML and presented by means of a program called V-Designer on the computer screen (of the subject PC) with a high color, 16-bit resolution of 1024×768 pixels.

The V-Designer program developed by Thomas Clermont (2001) is a control program for the eye-tracker system. It provides a user-friendly visual programming environment, which enables inexperienced programmers to create their own control program for the implementation of eye-tracking experiments. Moreover, V-Designer is Microsoft-Windows based, so that standard Windows hardware and software interfaces are accessible. Multimedia presentations of experimental stimuli such as video or sound are supported as well. Most important of all, V-Designer provides a solution to the typical problem of timing in Windows environments by implementing an independent timing function, which gives a highly accurate account of run-time behavior (Koesling, Clermont, and Ritter, 2001).

For the first six experimental conditions, each section of the learning material contained a diagram (static picture), a text presented either visually or auditorily, and the links connecting to other sections as well as a link to end the learning program. The diagram was located in the upper part of the screen whereas the links were embedded at the bottom. The visually-presented text was located in between. Unlike the first experiment, there was no control-bar for subjects to pause or stop playing the auditory text. Yet, subjects could simply press any key on the keyboard to replay the auditory text.

Examples of the learning material can be found in Figure 1 to Figure 5. All the diagrams were 800×600 pixels in size. On the left part of the diagram, there was a drawing (200×200 pixels in size) of a Chinese-chess piece with its name beneath it. In the simple diagrams, only one possible move (based on the movement rules) for every single Chinese-chess piece was visualized. A blue dot symbolized the starting position of the piece whereas the end-position was marked by a black circle. The blue arrows indicated the direction as well as the length of a possible movement. The links to other instructional sections were put in a table, conveyed by the small drawings (50×50 pixels in size) of the chess pieces with their names above them.
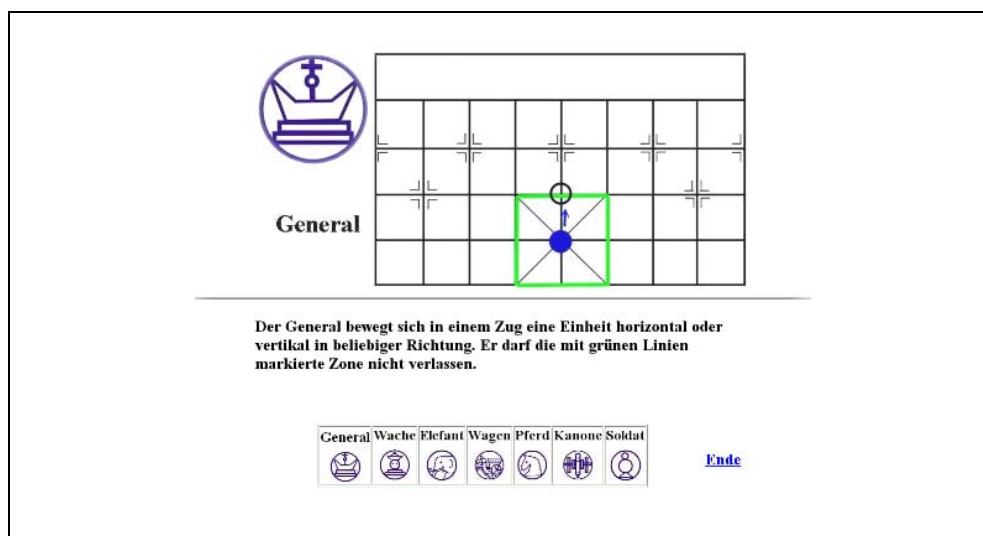
Figure 1: An example of an instructional section of Condition (simp-visu). Every section contains a simple diagram and a visual text.

In the diagrams with medium complexity, all the possible moving directions from a starting point are visualized. Besides, there were 5 or 6 visual distractors depicted by gray dots in every diagram. The gray dots represent some other pieces on the chess board, and they are additionally used to distract subjects' attention, and therefore, enhance picture complexity.

Figure 2: An example of an instructional section of Condition (med-visu).

For the complicated diagrams used in complex diagrams, small drawings of the chess pieces instead of the dots were employed. Not only were all possible movements visualized but also more visual distractors were presented in the diagrams.
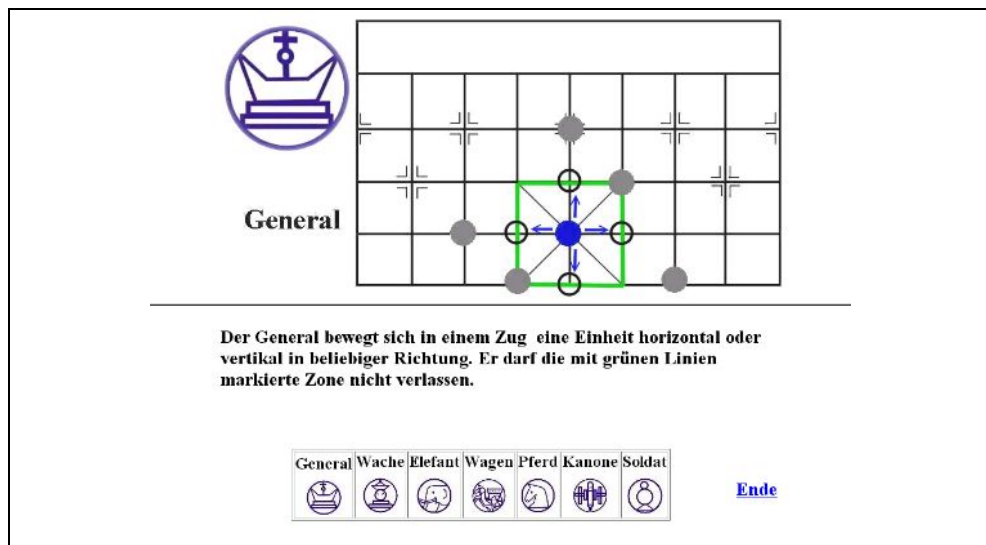

Figure 3: An example of an instructional section of Condition (comp-visu).
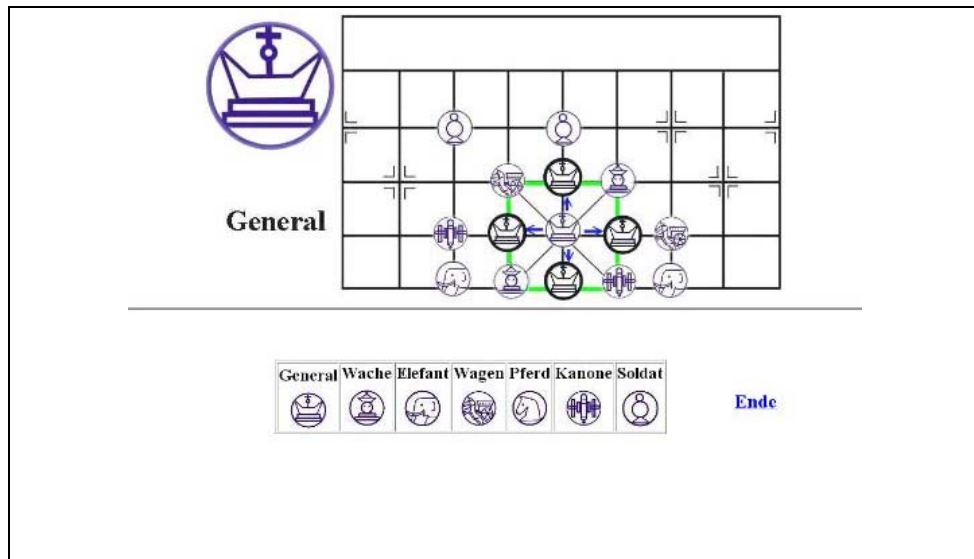
Figure 4: An example of an instructional section of Condition (comp-audi). In this condition, subjects were given auditory text.

For the last two experimental conditions, I employed animation for displaying the movement rules. Again, animation was presented with either visual or auditory text. The same number of visual distractors shown in Condition (comp-visu) and Condition (comp-audi) were fitted in the background of the display of animation as well. Figure 5 shows an example of an instructional section of Condition (ani-visu).
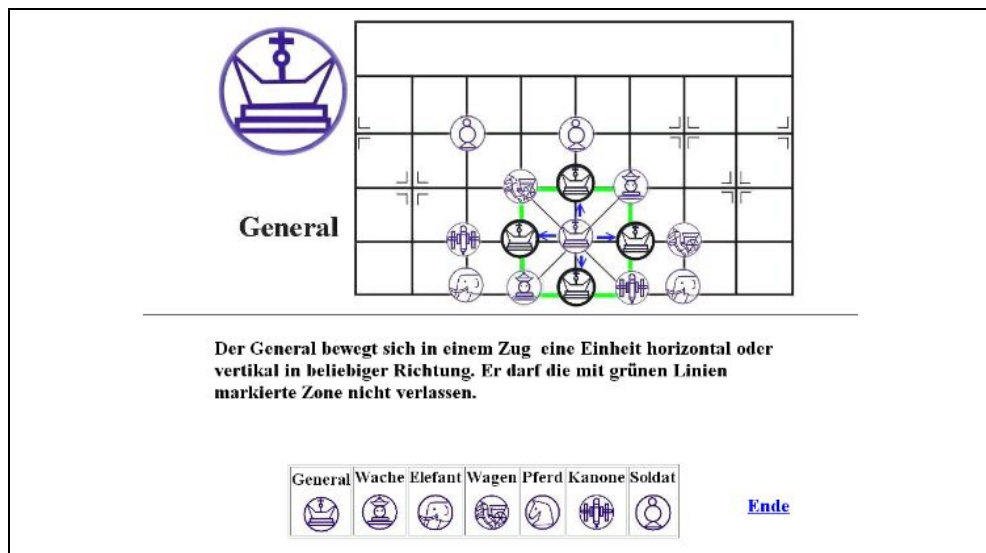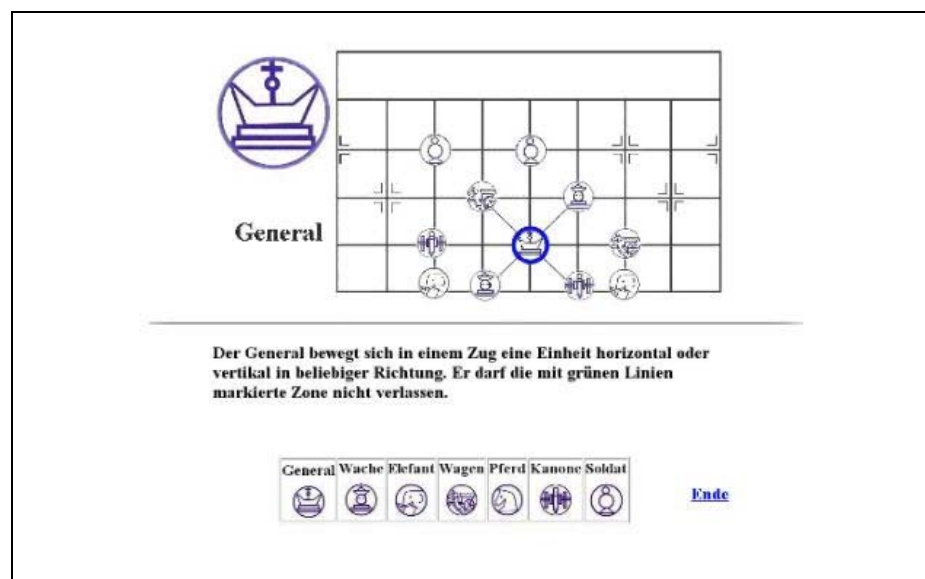

Figure 5: An example of an instructional section of Condition (ani-visu).

The chess piece for which the rules were described by the instruction was marked with a blue circle every time in its starting position shortly before it moved to a possible position on the chess board. The animation showed how the piece moved to all the possible positions that it could potentially reach in one move from its current position. In some instructional sections, the positions to which a chess piece was not allowed to move were signaled by inserting red crosses in the referential positions.

8.3.3 Apparatus

The same SMI EyeLink eye-tracking system as in Experiment 1 was used to measure subjects' eye movements. There was, however, a change in the hardware of the subjects' PC which contained an AMD Athlon 600 MHz processor and a new graphics card (ELSA ERAZOR III). With the aid of the new graphics card, the stimuli viewed by the subjects during the experiment were delivered directly in the form of video signals to the VISCA recorder, so that the sequences of stimuli viewed by each subject could be videotaped. The synchronization of eye movements and the video was again carried out by the Viper program.

8.3.4 Procedures

On arrival, subjects were asked about their experiences of playing chess or games similar to chess, so that the number of subjects with different amounts of experience (chess experts were excluded) could be balanced between conditions. Subsequently, basic information about Chinese chess as well as their respective tasks were introduced to subjects. After calibrating the eye-tracker, another instruction concerning how to use the learning environment was shown on the screen of the subjects' PC. Subjects were asked to learn the movement rules of Chinese-chess pieces on the computer with concentration. There was no time limit for learning, but subjects were asked to learn the information as quickly as they could. When the subjects finished learning, they were given a test to assess their learning results and another test to roughly estimate their prior knowledge of chess. In addition, they had to fill out a questionnaire which comprised more or less the same questions as did the one used in the first experiment.

There were eight questions in the test for assessing subjects' learning results. Each question tested the movement rules of a Chinese-chess piece. For example, the question for testing the rules regarding the elephant is shown in Figure 6. Subjects were asked to mark the piece or pieces which was/were threatened by the blue

elephant. There were two questions about the piece "soldier" because the rules were different depending on which zone the soldier was in.

The test for estimating prior knowledge of chess consisted of six questions. Subjects had to write down the definition of the following six specialist terms regarding chess: blitz tournament, en passant, ELO, castling, gambit, and fictitious attack. The score a subject achieved in this test was supposed to be positively correlated to his or her prior knowledge about chess.

8.4 Results

8.4.1 Assessment of learning results

First, I would like to discuss the method for assessing the subjects' learning results. In Experiment 1, I employed "learning efficiency" as the measurement of subjects' learning performance. The learning efficiency was calculated by:

$$\textbf{Learning efficiency = C\_score/(LT+AT)}$$

that is, the correct scores subjects achieved in assembling the cube puzzle divided by the time subjects spent learning the instructions on the computer and the time taken to assemble the puzzle. In the current experiment, I use the error rate to assess the learning results instead of learning efficiency because the error rate was shown to be independent of any reaction time components: neither the correlation between the learning time and the error rate ($r = -0.095$, $n = 40$, $p = 0.56$) nor that between the task time and the error rate ($r = 0.241$, $n = 40$, $p = 0.134$) was found to be significant.
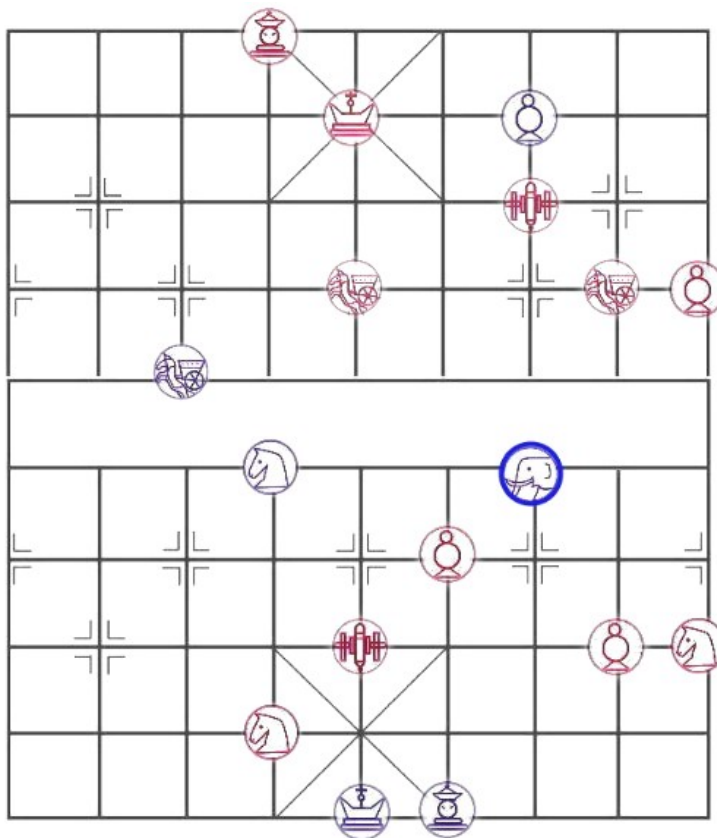
Figure 6 : A sample question used in the test for assessing learning
results. All three rules about the elephant were tested in a question.

8.4.1.1 Error rate

Figure 7 depicts the mean error rate of the eight experimental conditions. The audio-visual conditions were presented in the gray color, while the visual-only conditions were presented in black. The results of the ANOVA showed that both picture complexity and text mode had no significant effect on the error rate. However, the interaction between picture complexity and text mode was significant ($F(3; 32) = 5.948$; $p < 0.005$). Moreover, subjects' experience of chess or similar games exerted no effect on the error rate.

      Post hoc pairwise comparison using Fisher's Least-Significant-Difference tests (LSD tests) showed that the differences between Condition (comp-visu) and Condition (comp-audi), Condition (comp-visu) and Condition (ani-visu), Condition (comp-audi) and Condition (ani-audi) as well as Condition (ani-visu) and Condition (ani-audi) were significant.

Figure 7: Mean error rate as a function of experimental condition.

These results only partially correspond to the hypotheses that the DMP conditions should not outperform the SMP conditions if the pictorial information is visually very easy or very highly-demanding. There was no significant difference in the error rate between Condition (simp-visu) and Condition (simp-audi), but the differ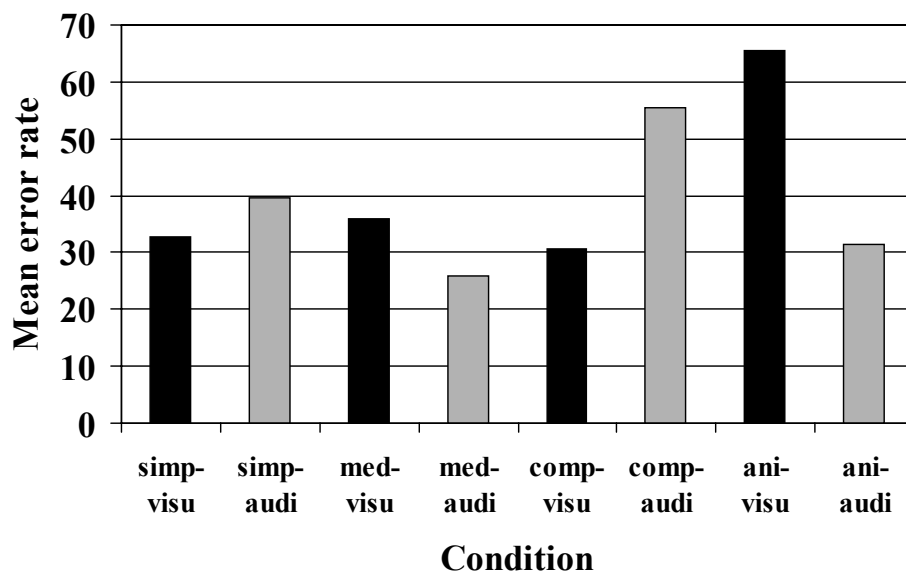ence between Condition (comp-visu) and Condition (comp-audi) was significant according to the LSD test. When the mean error rates for Condition (comp-visu) and Condition (comp-audi) were compared, the mean error rate for DMP (Condition (comp-audi): mean = 55.5%) was significantly higher than that for SMP (Condition (comp-visu): mean = 30.7%). The explanation for this result could possibly be that there was interference between the visual and auditory information processing. Since the total amount of information coming from the visual and auditory channels exceeded the capacity of the CE, subjects failed to process visual and auditory information simultaneously. Consequently, the integration of visual and auditory information could have been damaged.

According to the hypotheses, the DMP should be superior to the SMP if the demand for pictorial information is at a medium level. However, the difference in the error rate between Condition (med-visu) and Condition (med-audi) was not significant, though the error rate in relation to Condition (med-audi) was slightly lower than that for Condition (med-visu). The error rate for Condition (ani-visu) was the highest among all the experimental conditions and was substantially higher than that for Condition (ani-audi), which was consistent with the results derived in Experiment 1.

8.4.1.2 Learning time

Figure 8 depicts the learning time and task time for each of the eight experimental conditions. Picture complexity exerted a significant effect on the learning time ($F(3; 672) = 2.881$; $p < 0.05$), whereas text mode did not. The mean learning time for materials with medium picture complexity was significantly shorter than for those with simple pictures. There was a tendency towards an interaction between picture complexity and text mode ($F(3; 672) = 2.271$; $p = 0.079$). In addition, the prior experience with chess or similar games tended to have a slight influence on the learning time as well ($F(2; 37) = 2.809$; $p = 0.07$). Subjects without prior knowledge of chess or similar games tended to take more time learning than subjects with a small or medium amount of prior knowledge. Furthermore, the correlation between learning time and task time was significant ($r = 0.43$; $n = 40$; $p < 0.01$). Learning time was positively correlated with task time.
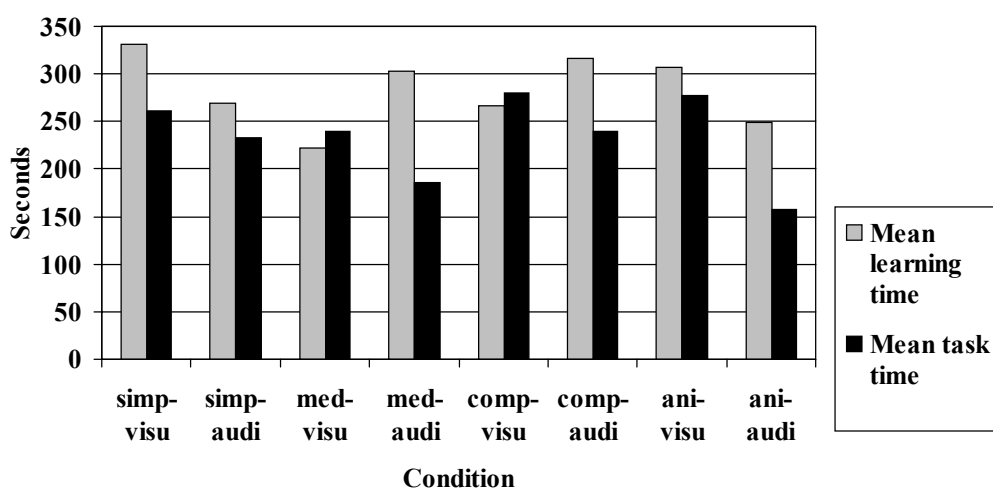
Figure 8: Mean learning time and mean task time as a function
of the experimental condition

The correlation between the learning time and the error rate was not significant. Moreover, it is inevitable that there is always a certain amount of base time required for playing a sound or video file. Thus, the learning time in the case of the audio-visual as well the as animation conditions should be potentially longer than that in the case of the visual-only conditions where static pictures are shown. Nevertheless, the Bonferroni tests pointed out that there was no significant difference in the learning time between the experimental conditions. Since subjects were allowed to view the instructions as long as they wanted, the effect of experimental conditions on the learning time might not have been transparent. As a consequence,

the effect of experimental conditions on the learning time should be better uncovered by the time that subjects spent in learning each single instructional section (section time).

8.4.2 Mean section time

Figure 9 conveys the mean section time as a function of the experimental condition. According to the ANOVA, there was no simple effect of either picture complexity or text mode, but the interaction between these two factors was significant ($F(3; 672) = 5.596$; $p < 0.005$). The results of Bonferroni tests indicated that the mean section time for Condition (med-visu) was significantly shorter than that for the other conditions.
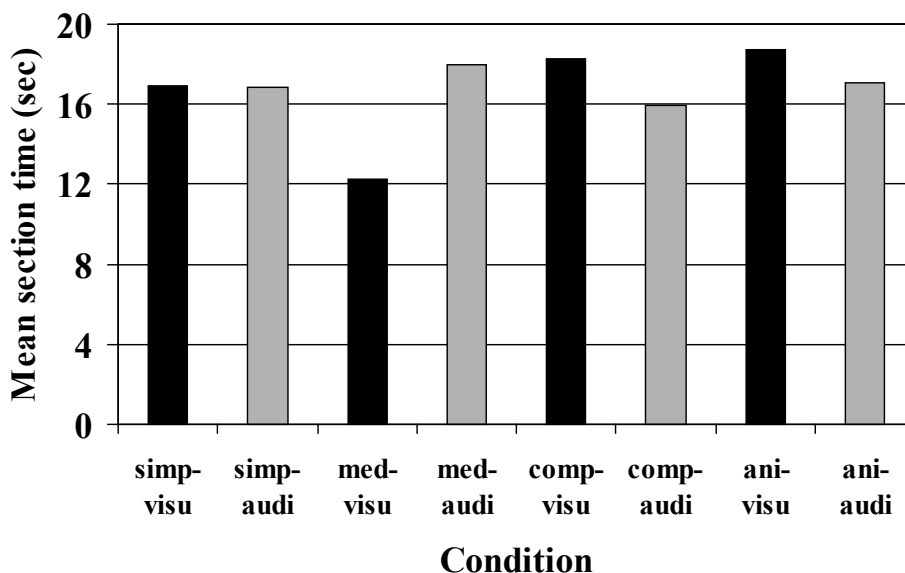
Figure 9: Mean section time as a function of the experimental condition

The mean section time calculated in accordance with instructional sections is shown in Figure 10 and Figure 11. In the between-group analysis, the difference in the mean section time was not significantly affected by the experimental conditions.
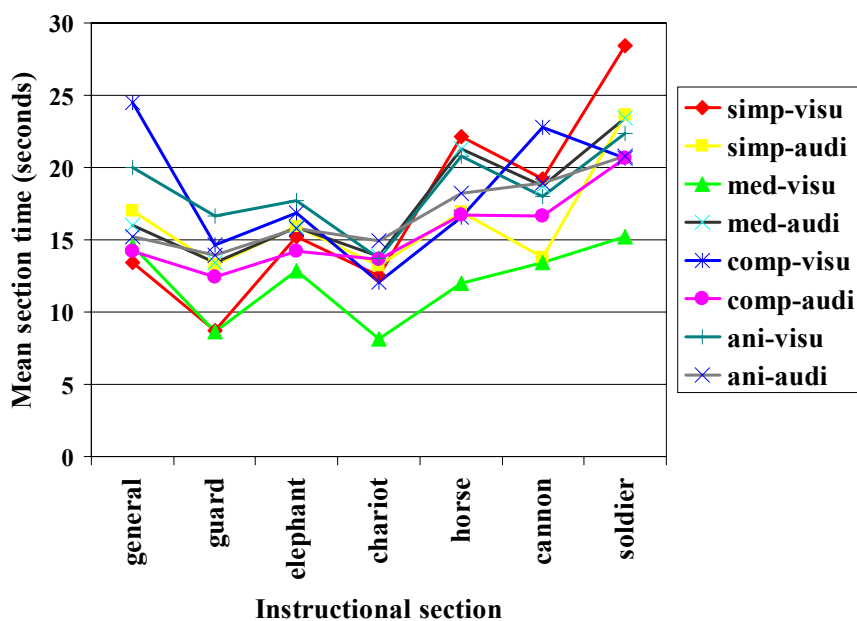
Figure 10: Mean section time as a function of experimental condition, by instructional sections
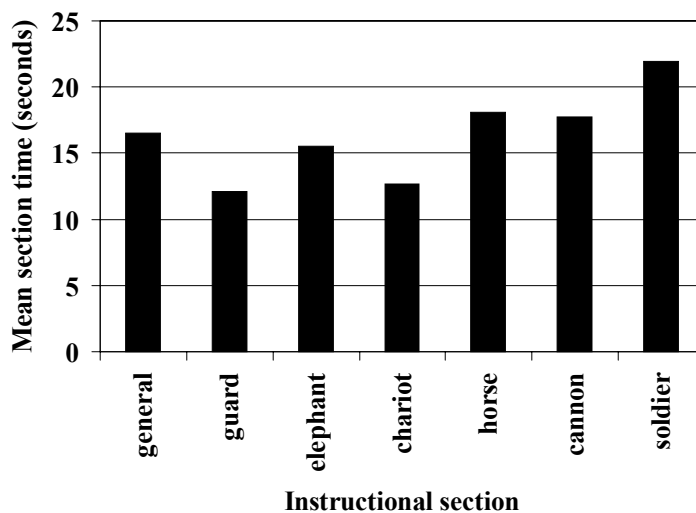
Figure 11: Mean section time as a function of instructional section

There were significant differences in the learning time between the instructional sections ($F_{(6; 673)} = 10.689$; $p < 0.001$). According to the Bonferroni tests, the mean section time of the Section 'soldier' was significantly longer than that of the other sections, whereas the mean section times of the Sections 'guard' and 'chariot' were significantly shorter than those in relation to the Sections 'horse', 'cannon', and 'soldier'. The cause of the differences in the mean section time could be the number of propositions, which will be analyzed subsequently.

8. Experiment 2

In addition to picture complexity and text mode, other factors such as the number of propositions in the texts and the number of objects in the pictures may also influence the section time. Since the number of objects comprises the number of visual distractors as well as the number of relevant objects, and each of them might have a different influence, it is better to examine the effect of the two different object types separately. I, therefore, first give an overview of how the quantity of those factors was distributed in the instructional sections. Tables 1 to 4 depict the number of propositions, objects, visual distractors and relevant objects in each of the corresponding sections.

Chinese-chess pieces	general	guard	elephant	chariot	horse	cannon	soldier
Number of propositions	11	9	11	9	14	11	13

Table 1: The number of propositions in the corresponding instructional section

	Conditions			
	(simp-visu/-audi)	(med-visu/-audi)	(comp-visu/-audi)	(ani-visu/-audi)
Chinese-chess pieces	number of objects	number of objects	number of objects	number of objects
general	2	10	15	11
guard	2	10	15	11
elephant	4	10	16	12
chariot	2	10	15	11
horse	5	15	20	12
cannon	4	11	16	13
soldier	3	15	16	9

Table 2: The number of objects in the respective instructional section

	Conditions			
	(simp-visu/-audi)	(med-visu/-audi)	(comp-visu/-audi)	(ani-visu/-audi)
general	0	5	10	10
guard	0	5	10	10
elephant	0	5	10	10
chariot	0	5	10	10
horse	0	5	10	10
cannon	0	5	10	10
soldier	0	5	8	8

Table 3: The number of visual distractors in the relational instructional section

	Conditions			
	(simp-visu/-audi)	(med-visu/-audi)	(comp-visu/-audi)	(ani-visu/-audi)
Chinese-chess pieces	number of relevant objects	number of relevant objects	number of relevant objects	number of relevant objects
general	2	5	5	1
guard	2	5	5	1
elephant	4	6	6	2
chariot	2	5	5	1
horse	5	10	10	2
cannon	4	6	6	3
soldier	3	10	8	1

Table 4: The number of relevant objects in the respective instructional section

Again, I adopted the same definitions of propositions developed by Kintsch (1974) as the criteria to count the number of propositions. Word categories regarded as predicates are verbs, adverbs, adjectives, conjunctions, prepositions, and quantifiers. The number of objects includes figures that refer to the positions of the relevant and the non-relevant (visual distractors) chess pieces on the chess board. The relevant objects are the objects that carry relevant meaning for elucidating the movement rules. On the contrary, the visual distractors represent some other chess

pieces which are irrelevant to the movement rules. In the diagrams that are of 'simple' and 'medium' complexity, for instance, the figures regarded as objects consist of blue and gray dots as well as black circles. The blue arrows and the red crosses are included in the number of objects.

8.4.2.1 The effect of the number of propositions

The data are given in Figure 12. According to the ANOVA, the effect of the number of propositions on the mean section time was significant ($F_{(3; 676)} = 20.465$; $p < 0.001$). The results indicated that the mean section time was positively related to the number of propositions.
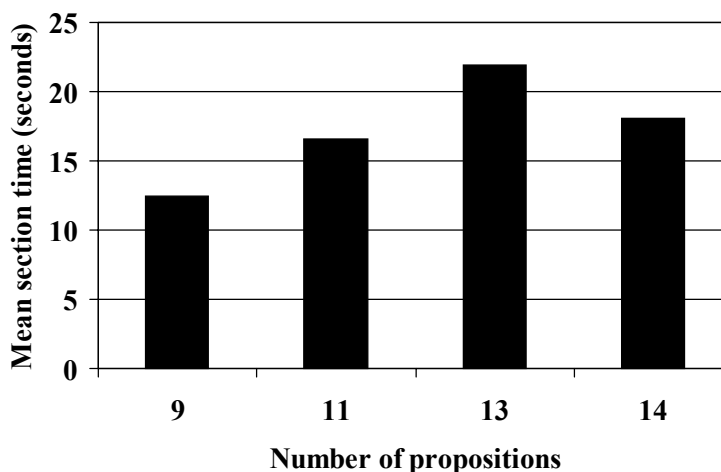
Figure 12: Mean section time as a function of the number of propositions

On the other hand, the effect of the number of objects, the number of visual distractors, and the number of relevant objects on the mean section time was also examined. It should be mentioned that I used 'quantity' instead of the absolute number of visual distractors as well as of relevant objects because the amount of them was varied regularly between the levels of picture complexity. Accordingly, the absolute number in this case is not important but it is only the variation in the amount that is of interest. For the sake of convenience, I distinguished three levels of the quantity of visual distractors and the quantity of relevant objects. For the quantity of visual distractors, the three levels were 'no' (0), 'medium' (5), and 'high' (10). As to the quantity of the relevant objects, the 'low' quantity corresponded to 1 to 3 objects, 'medium' contained 4 to 5, while 'high' consisted of 6 to 10 objects.

Moreover, it should be noted that the picture mode (static vs. animated) will influence the effects of the number of objects, the quantity of visual distractors as well as the quantity of relevant objects. The number of objects in animation, for

example, affects section time in a different way than it does when static pictures are used. Since subjects' visual attention is attracted or guided by the motion, the effect of the number of objects or the visual distractors, respectively, on the mean section time is probably much weaker than that in static pictures. Thus, to avoid contamination of the data analysis, it is better to distinguish the effect of static-picture conditions from that of the animation conditions. In this experiment, I have mainly evaluated the effects of these factors in relation to the static-picture conditions. This principle is also employed with regard to the ongoing evaluations of the effects concerning those factors. In addition, if there is an interaction between the effects of text mode and picture complexity, there is also probably an interaction between text mode and the number of objects to be expected, because picture complexity is partly defined by the number of objects. This should also be taken into account when analyzing the effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects.

8.4.2.2 The effect of the number of objects

Figure 13 depicts the mean section time as a function of the number of objects. The results showed that the effect of the number of objects on the mean section time for static-picture conditions was significant ($F(8; 516) = 5.852$; $p < 0.001$): the mean section time was positively related to the number of objects. Due to the large variance between the values, the relationship between the mean section time and the number of objects was not linear. According to the Bonferroni tests, significant differences were found to exist between the 2/3, 2/16, 3/4, 3/10, 3/11, 3/15, 10/16 objects. In order to roughly estimate the general effect of the number of objects on the mean section time, the number of objects was divided into two conditions at the median (=10). A t-test ($t(523) = -1.747$, $p = 0.081$) showed that the mean section time of the sections with few objects (< = 10) (mean = 15.508 sec) tended to be shorter than that of the sections with many objects (> 10) (mean = 17.124 sec).
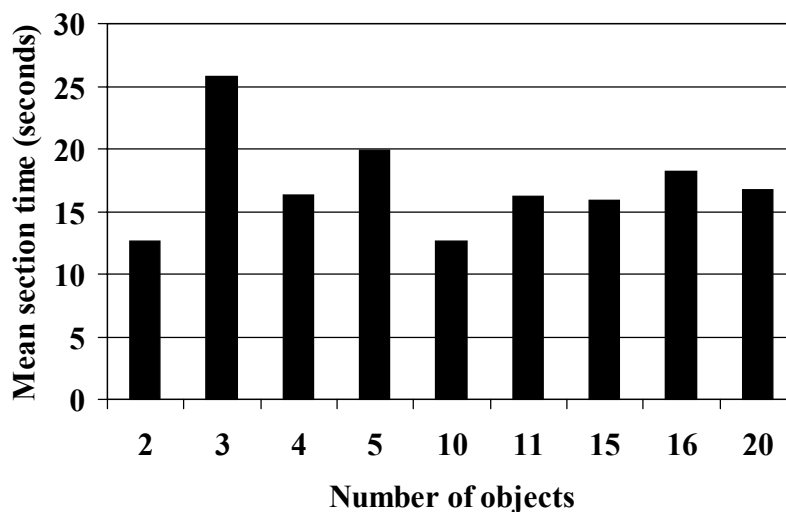
Figure 13: Mean section time as a function of the number of objects
in the static-picture conditions

The quantity of the visual distractors had no effect on mean section time, whereas the effect of the quantity of relevant objects was significant ($F(2; 522) = 3.697$; $p < 0.5$). The data relating to the quantity of the relevant objects are shown in Figure 14. When the text mode was considered, the effect of the quantity of relevant objects was only significant when auditory text ($F(2; 260) = 9.142$; $p < 0.001$) was presented. The Bonferroni tests pointed out that the difference between 'medium' and 'high' was significant.
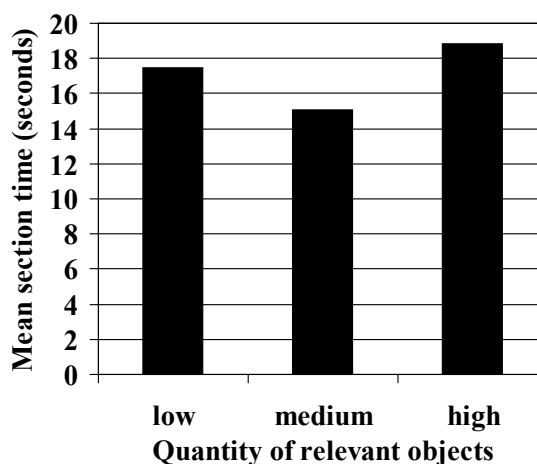
Figure 14: Mean section time as a function of the quantity of
relevant objects, when text was presented auditorily

8.4.2.3 The effect of the number of visits to the same section

The factor 'the number of visits to the same section' was also investigated in connection with the mean section time. Figure 15 depicts the mean section time as a function of the number of visits to the same section. It stands to reason that mean section time was inversely related to the number of visits to the same section.
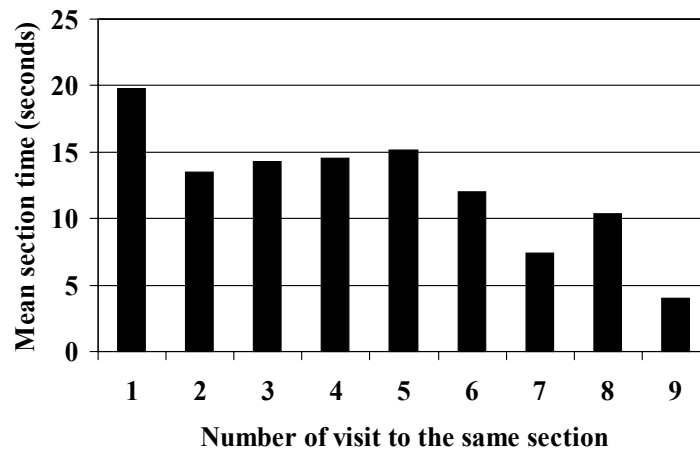
Figure 15: Mean section time as a function of the number of visits to the same section

8.4.3 Mean section frequency

Since subjects were free to view a section as often as they wanted, it is also important to examine whether factors like picture complexity, text mode, the number of propositions, the number of objects, and the quantity of visual distractors determine how often a subject views a certain instructional section.

Figures 16 and 17 show the mean section frequency as a function of experimental condition. The results indicated that picture complexity had a significant effect on the mean section frequency ($F(3; 672) = 16.839$; $p < 0.001$). Mean section frequency was significantly lower when animation was presented. Among the conditions where static pictures were displayed, the differences in the mean section frequency were not significant. In addition, there was a strong tendency that indicated that the mean section frequency was lower when auditory text was presented ($F(1; 672) = 3.680$; $p = 0.055$). The interaction between picture complexity and text mode was also significant ($F(3; 672) = 68.550$; $p < 0.001$). According to the Bonferroni tests, the mean section frequency of Condition (simp-visu) and Condition (comp-audi) were significantly higher than that of the other conditions, whereas the

frequency of Condition (comp-visu) and Condition (ani-audi) were substantially lower.
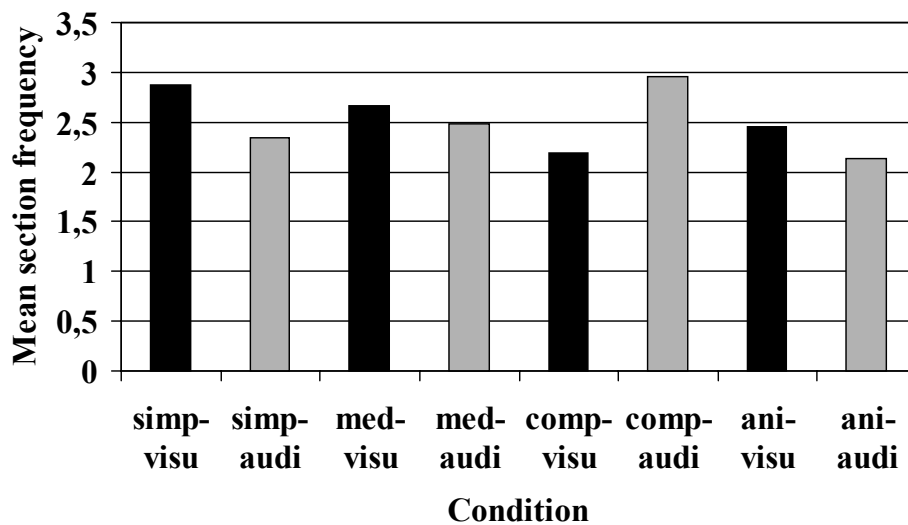
Figure 16: Mean section frequency as a function of experimental condition
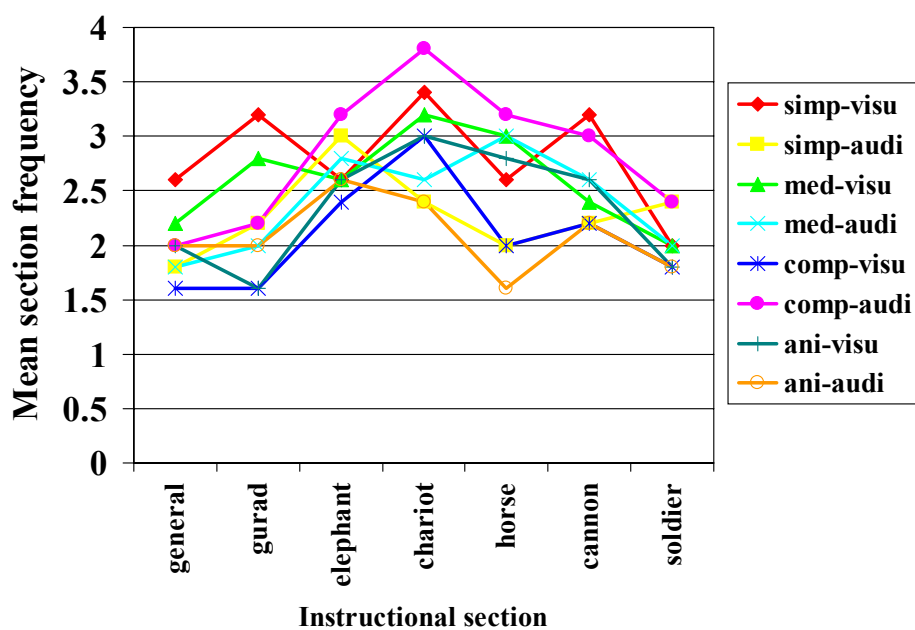
Figure 17: Mean section frequency as a function of experimental condition, by instructional sections

8.4.3.1 The effect of the number of propositions

The data are given in Figure 18. The mean section frequency was significantly influenced by the number of propositions ($F(3; 521) = 35.732$; $p < 0.001$). The data indicated that mean section frequency declined as the number of propositions increased. According to the Bonferroni tests, significant differences lay between the 9/11, 9/13, 11/13, 11/14, and 13/14 propositions.
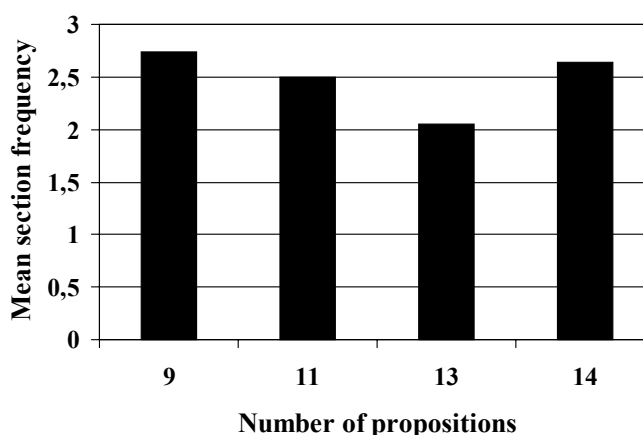
Figure 18: Mean section frequency as a function of the number of propositions

8.4.3.2 The effect of the number of objects

The data are depicted in Figure 19. The mean section frequency was significantly influenced by the number of objects ($F(8; 516) = 4.077$; $p < 0.001$) but not by the quantity of visual distractors or the quantity of relevant objects. The Bonferroni tests showed that significant differences existed between the 2/3, 3 /4, 3/15, 3/20, and 4/5 objects. Again, due to the large variance in the values, the relationship between the mean section frequency and the number of objects was not linear. I therefore divided the number of objects into three conditions: low (= 2, 3, or 4 objects), medium (= 5, 10, or 11 objects), and high (= 15, 16, or 20 objects). The Bonferroni tests showed that the significant difference was only between 'low' (mean = 2.677) and 'medium' (mean = 2.444).
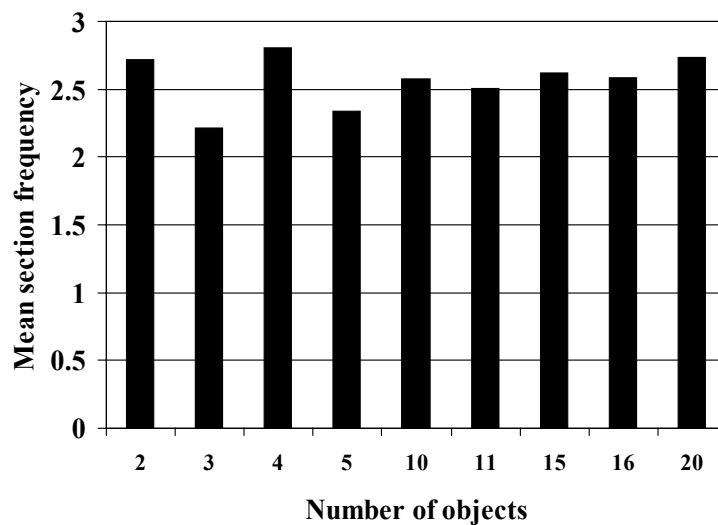
Figure 19: Mean section frequency as a function of the number
of objects in the static-picture conditions

8.4.4 Analyses of eye-movement data

To investigate the dynamics of subjects' visual attention, i.e. the viewing strategies for integrating the pictorial and textual information, quantitative as well as qualitative analyses of eye-movement data were carried out. In the first part of this section, I will first of all focus on the quantitative evaluations. Afterwards, the qualitative analyses of some subjects' viewing patterns will be described.

8.4.4.1 Quantitative analyses of eye-movement data

The evaluations are mainly focused on how subjects' fixation duration, the number of fixations, the fixation rate, and saccade length vary along with the factors – picture complexity, text mode, the number of propositions, the number of objects, the quantity of visual distractors, and the quantity of relevant objects. The analyses of every dependent variable consist of two parts. In the first part, I investigate the effects of picture complexity and text mode on the eye-movement variables in a more general way, so that the evaluation of the effects is concerned with the eye movements in regard to all components of the instructional sections that comprise text, pictures, and control buttons. In the second part, more detailed analyses are implemented so that the effects of all of the factors mentioned above are examined separately for the picture as well as for the text region (if text was presented visually) on the display of the instructional sections. It should be noted that the effects relating

to the number of objects are analyzed separately for static pictures and for animation, in order to avoid contamination.

8.4.4.1.1 Mean fixation duration

Figure 20 depicts the mean fixation duration as a function of picture complexity. The effect of picture complexity ($F(3; 672) = 23.446$; $p < 0.001$) on the mean fixation duration was significant. The Bonferroni tests indicated that the mean fixation duration of picture complexity for 'simple' as well as 'animation' were significantly higher than those for 'medium' and 'complex'. The difference between 'simple' and 'animation' was not significant.
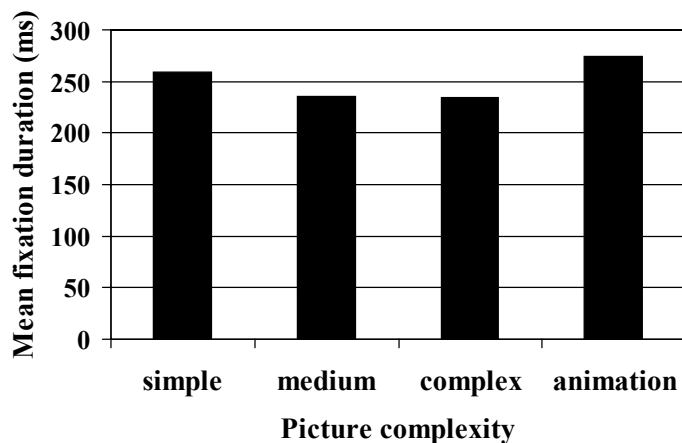
Figure 20: Mean fixation duration as a function of picture complexity

Text mode ($F(1; 672) = 124.215$; $p < 0.001$) and the interaction between picture complexity and text mode ($F(3; 672) = 13.688$; $p < 0.001$) also had a significant effect on the mean fixation duration. The mean fixation duration was significantly longer when auditory text was presented (mean = 272.042 ms) than when visual text was shown (mean = 228.744 ms). The effect of the interaction between picture complexity and text mode as depicted in Figure 21 indicated that animation plus auditory text produced considerably longer fixation duration than did other experimental conditions. When auditory text was presented, the mean fixation duration was much longer than when text was presented visually, with the result that the mean fixation durations of audio-visual conditions were always longer than those of visual-only conditions. The Bonferroni tests pointed out that the mean fixation durations of Condition (ani-audi) and Condition (simp-audi) were significantly longer than those of the other conditions.
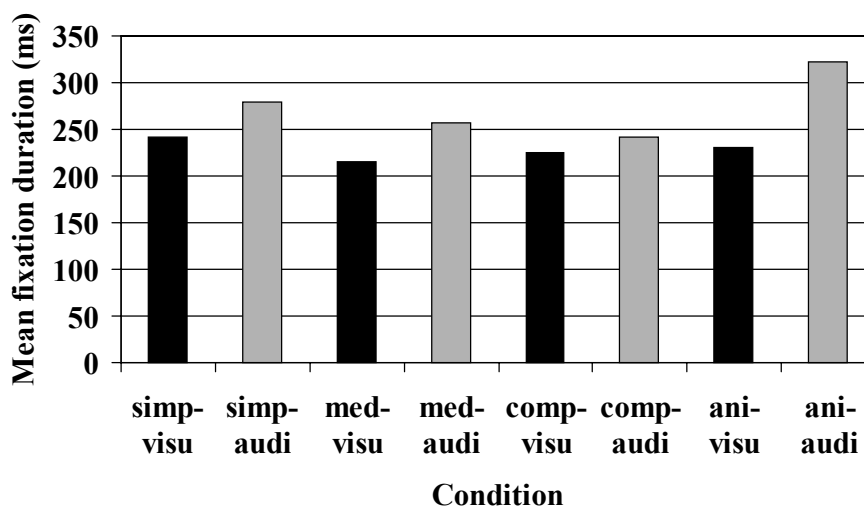
Figure 21: Mean fixation duration as a function of experimental condition

Mean fixation duration in the picture region

Mean fixation duration in the picture region was significantly influenced by picture complexity ($F(3; 672) = 39.181$; $p < 0.001$) and text mode ($F(1; 672) = 107.561$; $p < 0.001$) as well as by the interaction between these two factors ($F(3; 672) = 8.294$; $p < 0.001$). The effect of picture complexity clearly showed that mean fixation duration in the picture region was drastically longer when pictures were animated. As to the effect of text mode, the mean fixation duration in the picture region was significantly longer (mean = 280.887 ms) when auditory text rather than visual text was presented (mean = 233.599 ms). Due to the effect of the interaction between picture complexity and text mode, when static pictures and visual text were shown, there was no significant difference in the mean fixation duration in the picture region. Yet, when static pictures were combined with auditory text, significant differences could be observed. Figure 22 shows the mean fixation duration in the picture region when text was presented visually. According to the Bonferroni tests, 'animation' was significantly higher than the other conditions, while there were no significant differences between 'simple', 'medium', and 'complex'. Figure 23 depicts the mean fixation duration in the picture region when text was presented auditorily. The Bonferroni tests showed that 'simple' was significantly higher than 'medium' and 'complex'. 'Animation' was significantly higher than the other conditions.
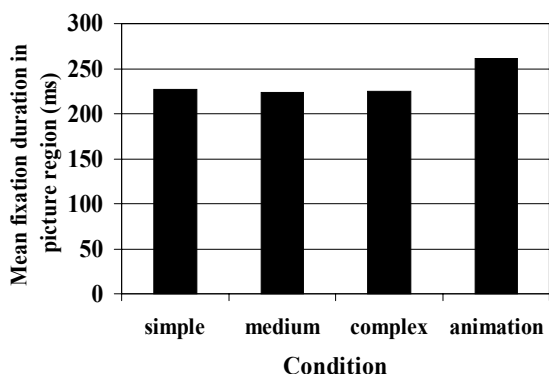
Figure 22: Mean fixation duration in the picture region when text was presented visually
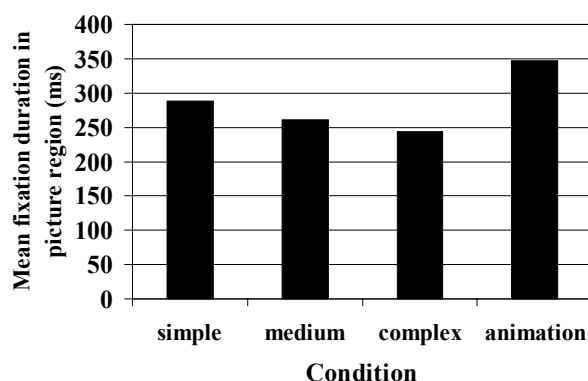
Figure 23: Mean fixation duration in picture region when text was presented auditorily

The effects of the number of propositions, objects and the quantity of visual distractors as well as relevant objects on the mean fixation duration in the picture region were also examined:

The effect of the number of propositions

The mean fixation duration in the picture region was independent of the number of propositions.

The effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects

Since picture complexity was partially defined by the number of objects, and there was an interaction between the effects of text mode and picture complexity on the mean fixation duration in the picture region, an interaction between the effects of text mode and the number of objects and between that of text mode and the quantity of visual distractors or between that of text mode and the quantity of relevant objects could be expected. The results showed that when text was presented visually, the number of objects, the quantity of visual distractors and the quantity of relevant objects had no effect, whereas the mean fixation duration in the picture region was significantly influenced by these factors when text was presented auditorily (number of objects: $F(8; 254) = 3.565$; $p < 0.005$; quantity of visual distractors: $F(2; 260) = 12.288$; $p < 0.001$; quantity of relevant objects: $F(2; 260) = 11.426$; $p < 0.001$). As the results indicated, the mean fixation duration in the picture region was inversely

related to the number of objects (see Figure 24), the quantity of visual distractors (see Figure 25), and the quantity of relevant objects (see Figure 26).

Figure 24: Mean fixation duration in the picture region as a function of the number of objects in the conditions where static pictures and auditory text were presented

Comparing the mean values of the number of objects <= 10 (mean = 277.175 ms) with those > 10 (mean = 249.015 ms), a t-test (t(523) = 2.597, p < 0.05) showed that the mean fixation duration in the picture region declined substantially as the number of objects increased.
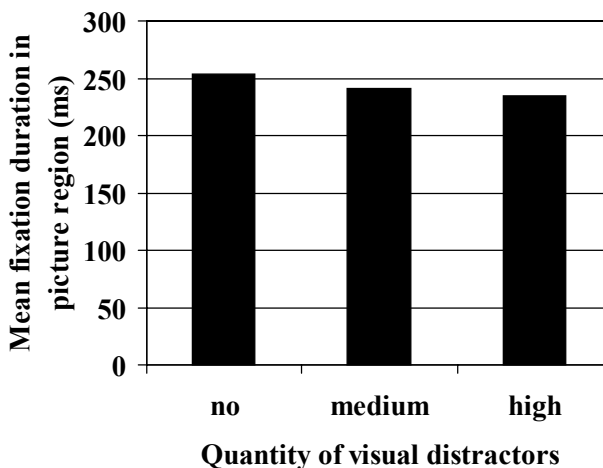
Figure 25: Mean fixation duration in the picture region as a function of the quantity of visual distractors. The Bonferroni tests showed that mean fixation duration was significantly longer when no visual distractors were displayed in the pictures.
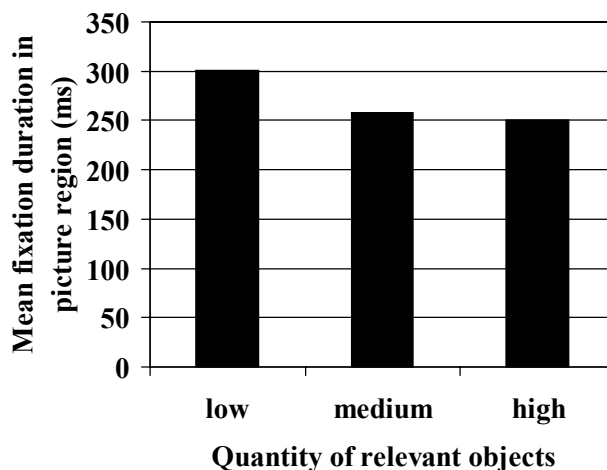
Figure 26: Mean fixation duration in the picture region as a function of the quantity of relevant objects. According to the Bonferroni tests, the mean fixation duration for the 'low' quantity was significantly longer than for the 'medium' and 'high' quantity.

The effect of the number of objects on the mean fixation duration in animation was also significant (F(3; 69) = 2.963; p < 0.05) when text was presented auditorily. Mean fixation duration decreased as the number of objects in animation increased (see Figure 27).
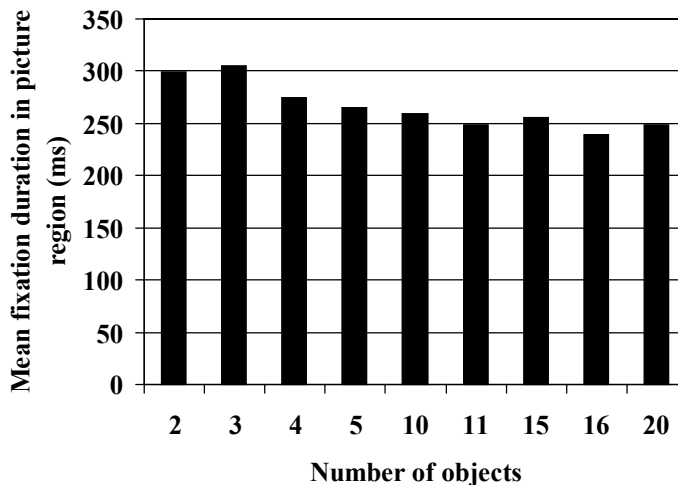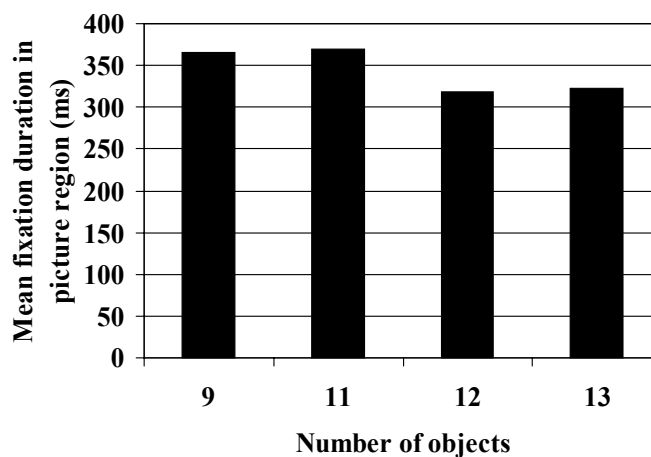

Figure 27: Mean fixation duration in the picture region as a function of the number of objects when animation and auditory text were presented.

In keeping with the results conveyed above, it appears that, irrespective of what kind of function the objects in the pictures had (relevant or not relevant), the

160

mean fixation duration in the picture region always declined as the quantity of the objects increased.

Mean fixation duration in the text region

The mean fixation duration in the text region was independent of the experimental conditions. Besides, the number of propositions had no effect on the mean fixation duration in the text region.

8.4.4.1.2 Mean number of fixations

The values of the mean number of fixations for each of the eight experimental conditions are depicted in Figure 28. The ANOVA showed that the mean number of fixations was not significantly influenced by picture complexity but by text mode ($F(1; 672) = 4.895$; $p < 0.05$) and by the interaction between picture complexity and text mode ($F(3; 672) = 5.349$; $p < 0.005$). The mean number of fixations was significantly larger when text was presented visually. The results indicated that the mean number of fixations in the visual-only conditions increased substantially when picture complexity was strongly enhanced, whereas there was no significant difference in the mean number of fixations in the audio-visual conditions. According to the Bonferroni tests, significant differences were found to exist between Condition (med-visu) and Condition (comp-visu), Condition (med-visu) and Condition (ani-visu), as well as Condition (comp-visu) and Condition (ani-audi).
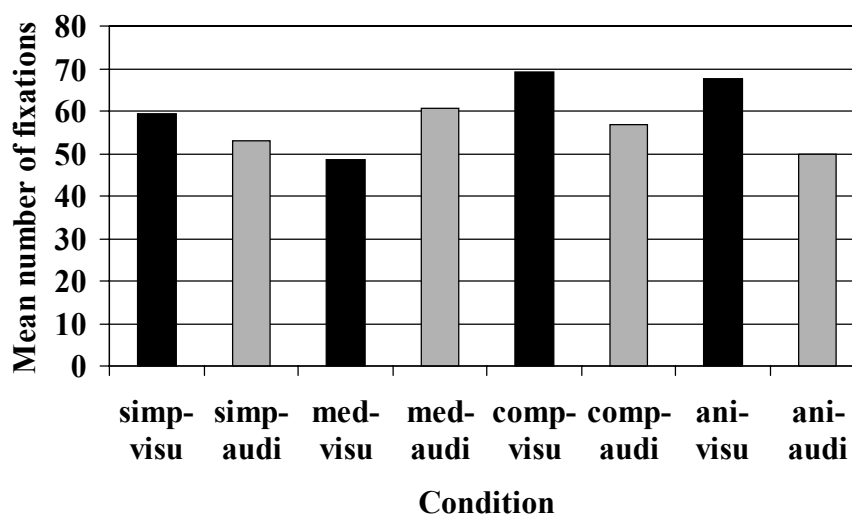
Figure 28: Mean number of fixations as a function of experimental condition

161

Mean number of fixations in the picture region

Figure 29 depicts the mean number of fixations in the picture region for each of the eight experimental conditions. The mean number of fixations in the picture region was significantly influenced by picture complexity ($F(3; 672) = 9.128$; $p < 0.001$), and text mode ($F(1; 672) = 81.633$; $p < 0.001$) as well as the interaction between the two factors ($F(3; 672) = 4.613$; $p < 0.005$). The audio-visual conditions had significantly more fixations in the picture region than did the visual-only conditions when static pictures were presented. The difference was most pronounced with materials of medium complexity. The Bonferroni tests showed that the mean number of fixations in the picture region with 'simple' picture complexity was significantly smaller than that with 'medium' or 'complex' picture complexity, and when there was 'animation'. More precisely, Condition (simp-visu) was significantly smaller than the other conditions except for Condition (med-visu). Besides, significant differences also existed between Condition (simp-audi) and Condition (med-audi), Condition (med-visu) and Condition (med-audi), Condition (comp-visu) and Condition (comp-audi). There was no significant difference in the mean number of fixations in the picture region between audio-visual and visual-only conditions when animation was presented.
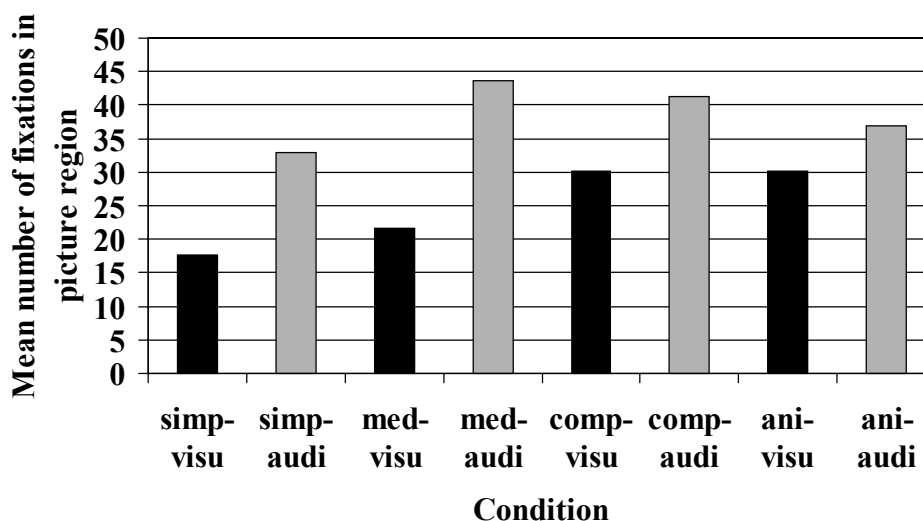
Figure 29: Mean number of fixations in the picture region as a function of experimental condition.

The effect of the number of propositions

Figure 30 shows the mean number of fixations in the picture region as a function of the number of propositions. The effect of the number of propositions was significant

$(F(3; 676) = 11.562; p < 0.001)$. As the results indicated, the mean number of fixations in the picture region increased as the number of propositions increased. This is actually a natural outcome when the text was presented auditorily because, the longer the text, the longer subjects would inspect the pictures, and therefore, the larger the number of fixations in the picture region. Interestingly, the number of propositions was still positively related to the mean number of fixations in the picture region, even though the text was presented visually $(F(3; 340) = 3.125; p < 0.05)$. The data were given in Figure 31.
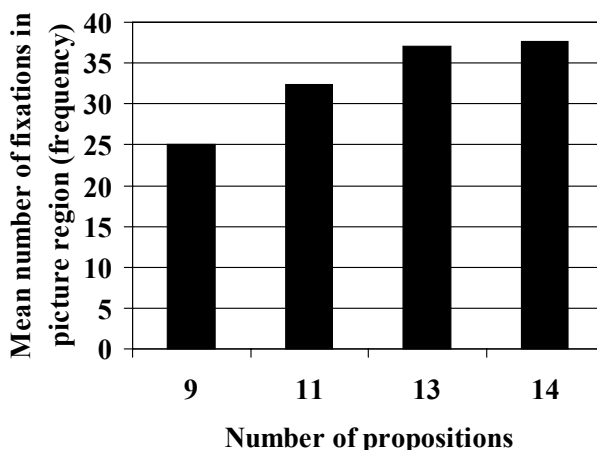
Figure 30: Mean number of fixations in the picture region as a function of the number of propositions
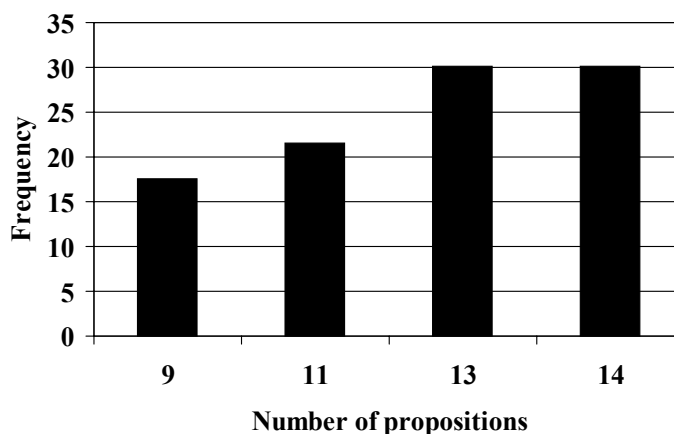
Figure 31: Mean number of fixations in the picture region as a function of the number of propositions when text was presented visually

The effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects

Figure 32 shows the effect of the number of objects on the mean number of fixations in the picture region ($F_{(8; 516)} = 5.369$; $p < 0.001$). The data clearly show that the mean number of fixations in the picture region increased as the number of objects increased.
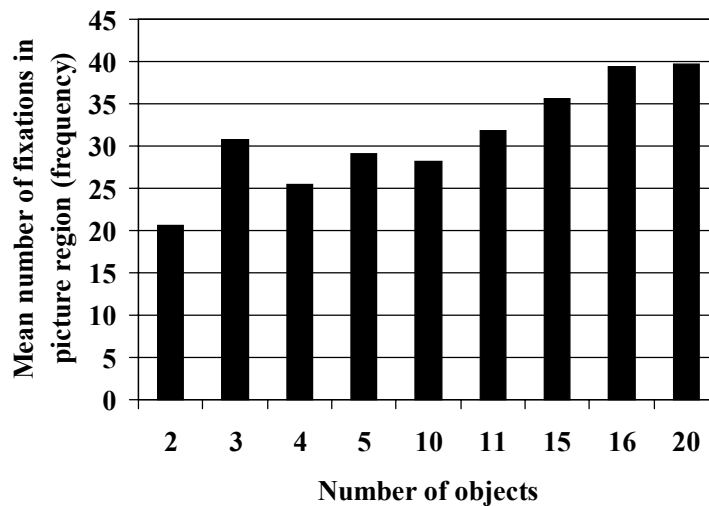
Figure 32: Mean number of fixations in the picture region as a function of the number of objects

The quantity of visual distractors ($F_{(2; 522)} = 13.264$; $p < 0.001$) (see Figure 33) and the quantity of relevant objects ($F_{(2; 522)} = 18.601$; $p < 0.001$) (see Figure 34) also had an effect on the mean number of fixations in the picture region. The mean number of fixations in the picture region was a positive function of the quantity of visual distractors and the quantity of relevant objects.
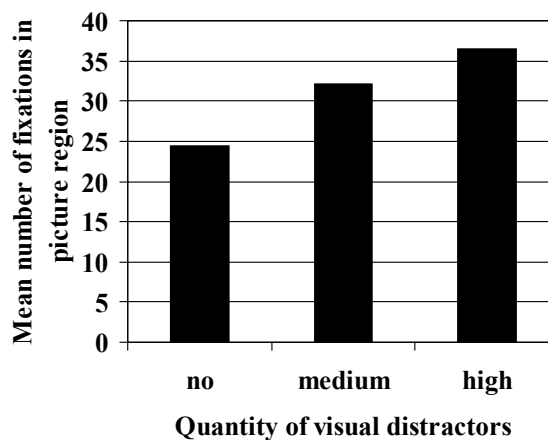
Figure 33: Mean number of fixations in the picture region as a positive function of the quantity of visual distractors. The Bonferroni

tests showed that the differences between 'no' and 'medium' as well as 'no' and 'high' were significant.
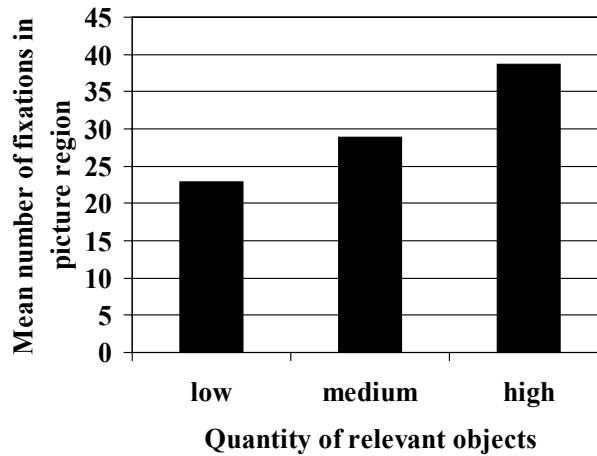
Figure 34: Mean number of fixations in the picture region as a function of the quantity of relevant objects. The Bonferroni tests showed that the quantity level 'high' was significantly greater than the 'low' and 'medium' levels

The results concerned with the mean number of fixations in the picture region are related to those to do with the mean fixation duration in the picture region. It appears that the learning strategy that subjects employed was to examine the objects in the pictures with short but many fixations. The objects of different functions (relevant or non-relevant) did not influence the eye-movement behavior in different ways.

Mean number of fixations in the text region

The mean number of fixations in the text region was independent of the experimental conditions. However, the mean number of fixations in the text region was positively related to the number of propositions (F(3; 340) = 5.388, p < 0.005). The data are given in Figure 35. The Bonferroni tests showed that '9' was significantly smaller than '11' and '13'.
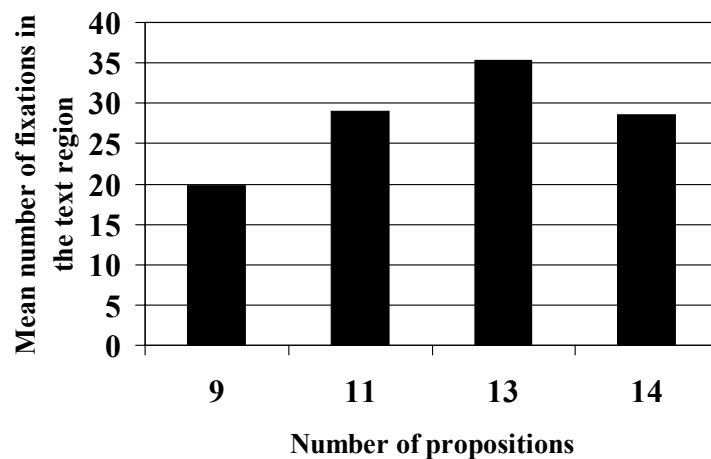
Figure 35: Mean number of fixations in the text region as a
function of the number of propositions

8.4.4.1.3 Mean fixation rate

The fixation rate was calculated as the number of fixations divided by the viewing
time. The reason for analyzing the fixation rate is that the fixation rate can give
insight into how fast a subject was viewing or scanning the instructions, which
provides information relating to the subject's strategies regarding information
processing and the eye-movement behavior triggered by different visual displays and
text modes.

Figure 36 shows the mean fixation rate as a function of experimental
condition. The results showed that the mean fixation rate was significantly
influenced by picture complexity ($F(3; 670) = 25.069$; $p < 0.001$), text mode ($F(1;
670) = 168.649$; $p < 0.001$), and the interaction between these two factors ($F(3; 670)
= 8.235$; $p < 0.001$). When pictures were static, the fixation rate for 'medium' picture
complexity was significantly higher than that for 'simple' complexity, whereas there
were no significant differences between 'simple' and 'complex' or 'medium' and
'complex'. Moreover, the mean fixation rate declined significantly when animation
was presented. When text was presented visually, the fixation rate (mean = 4.086
fixations/sec) was significantly higher than when auditory text was presented (mean
= 3.399 fixations/sec). According to the Bonferroni tests, the fixation rate of
Condition (med-visu) was significantly higher while those of Condition (ani-audi)
and Condition (simp-audi) were substantially lower when compared with those of the
other conditions. Additionally, significant differences were also found between
Condition (simp-visu) and Condition (med-audi), Condition (med-audi) and
Condition (comp-visu).
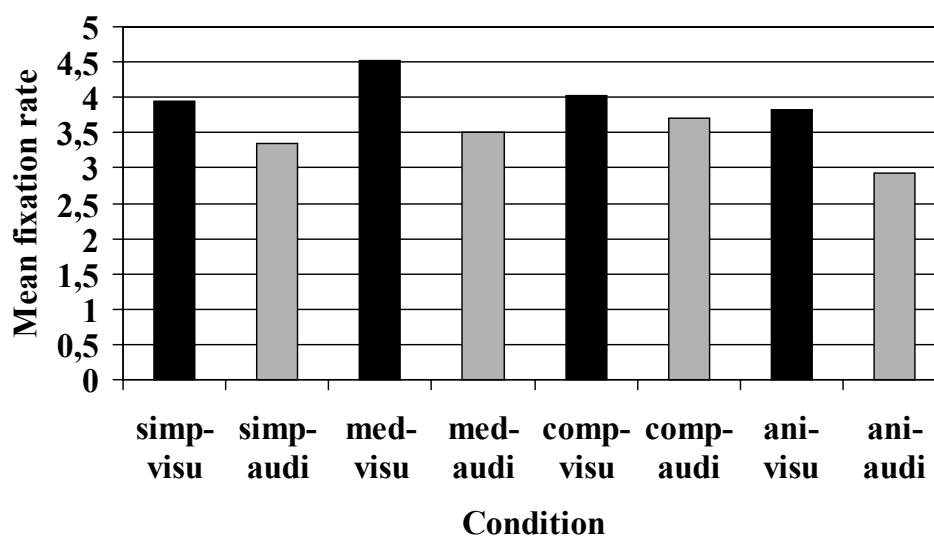
Figure 36: Mean fixation rate as a function of experimental condition. Again, audio-visual conditions are gray in color, and the visual-only conditions are black.

Mean fixation rate in the picture region

Figure 37 shows the mean fixation rate as a function of experimental condition. Picture complexity ($F(3; 668) = 28.584$; $p < 0.001$) as well as text mode ($F(1; 668) = 62.017$; $p < 0.001$) had a significant effect on the mean fixation rate in the picture region. The interaction between picture complexity and text mode was not significant but still pointed to a tendency ($F(3; 668) = 2,195$, $p = 0.087$). When text was presented auditorily, the mean fixation rate in the picture region (mean = 3.321 fixations/sec) was substantially lower than when the text was presented visually (mean = 3.666 fixations/sec). The Bonferroni tests pointed out that Condition (ani-audi) was significantly lower than other conditions. In addition, significant differences were also found between Condition (simp-visu) and Condition (simp-audi), Condition (comp-visu) and Condition (comp-audi), Condition (ani-visu) and Condition (ani-audi), but not between Condition (med-visu) and Condition (med-audi). On the other hand, the mean fixation rate in the picture region was significantly lower when animation was presented. Figure 38 shows the effect of picture complexity on the mean fixation rate in the picture region. The Bonferroni tests showed that 'animation' was significantly lower than the other conditions. Besides, 'complex' was significantly higher than 'simple'.
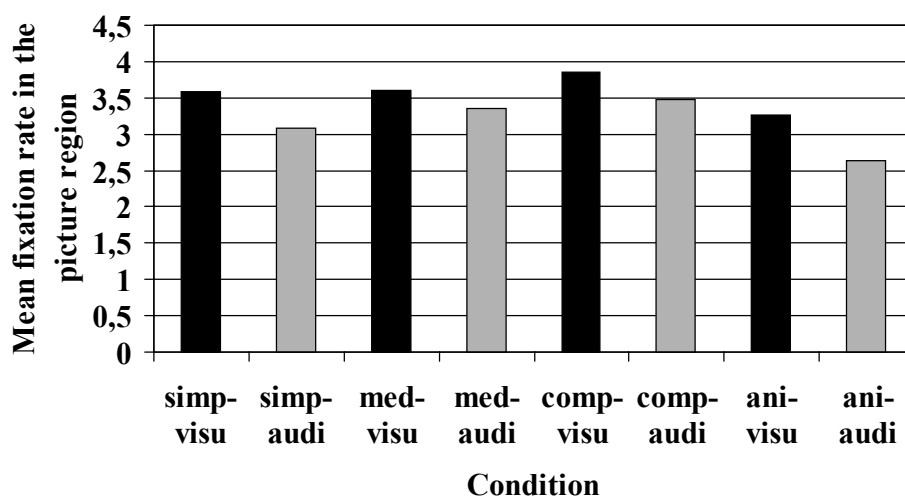
Figure 37: Mean fixation rate in the picture region of experimental condition
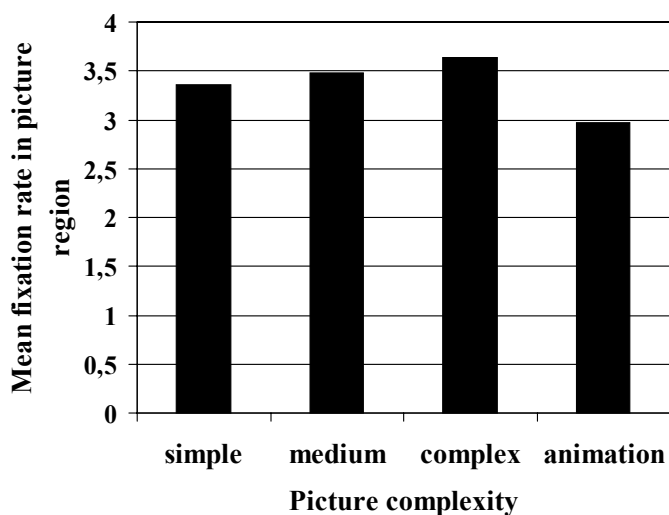
Figure 38: Mean fixation rate in the picture region as a function of picture complexity

The effect of the number of propositions

The number of propositions exerted no effect on the mean fixation rate in the picture region.

The effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects

The number of objects (F(8; 254) = 3.58; p < 0.005), the quantity of visual distractors (F(2; 260) = 11.356; p < 0.001) as well as the quantity of relevant objects (F(2; 260) = 9.885; p < 0.001) had significant effects on the mean fixation rate in the picture region when texts were presented auditorily. When texts were presented visually, the effects of those factors were not significant. Figure 39 shows the effect of the number of objects. The mean fixation rate in the picture region was positively related to the number of objects.
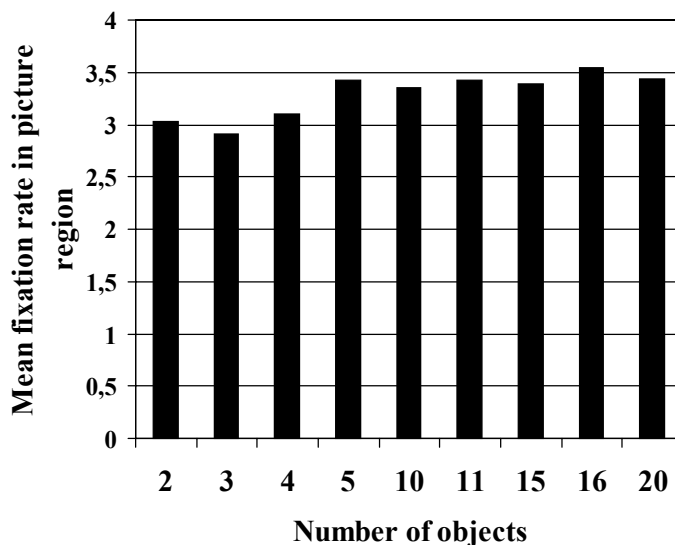
Figure 39: Mean fixation rate in the picture region as a function
of the number of objects

Figure 40 shows the effect of the quantity of visual distractors. The mean fixation rate in the picture region was positively related to the quantity of visual distractors. The Bonferroni tests indicated that the mean fixation rate was significantly lower when there were no visual distractors presented in the diagrams.
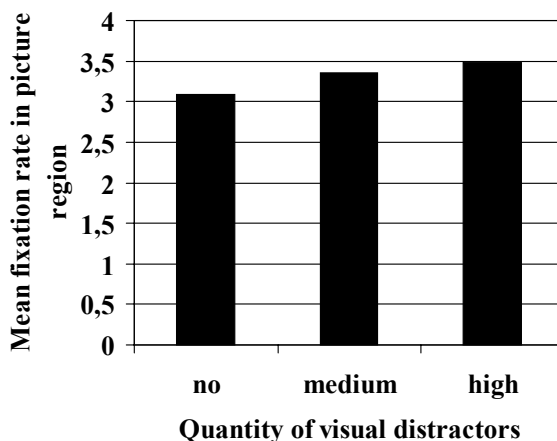
Figure 40: Mean fixation rate in the picture region as a function
of the quantity of visual distractors

Figure 41 shows the effect of the quantity of relevant objects. The mean fixation rate in the picture region was also positively related to the quantity of relevant objects. According to the Bonferroni tests, the mean fixation rate of 'low' was significantly lower than those of 'medium' and 'high'.
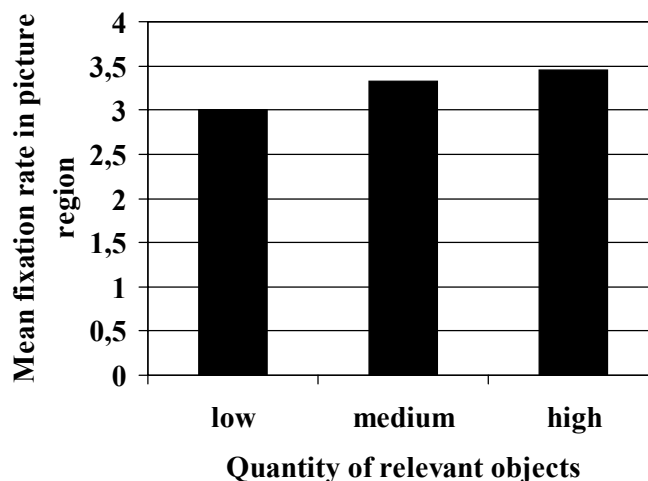
Figure 41: Mean fixation rate in the picture region as a function of the quantity of relevant objects

Mean fixation rate in the text region

The fixation rate in the text region was independent of the experimental conditions and the number of propositions.

8.4.4.1.4 Mean saccade length

Figure 42 depicts the mean saccade length of the eight experimental conditions. Picture complexity ($F(3; 672) = 5.115; p < 0.005$), text mode ($F(1; 672) = 151.354; p < 0.001$) as well as the interaction between these two factors ($F(3; 672) = 6.395; p < 0.001$) had significant effects on the mean saccade length. Mean saccade length was negatively related to picture complexity. However, the only significant difference lay between the 'simple' and 'medium' levels of picture complexity. There was no significant difference in saccade length between the 'medium', 'high' and 'animation' levels. When text was shown visually (mean = 139.015 pixels), mean saccade length was significantly greater than when text was presented auditorily

(mean = 107.101 pixels). Thus, the Bonferroni tests pointed out that the mean saccade length in relation to the audio-visual conditions were significantly shorter than those concerned with the visual-only conditions: the effect was most pronounced between the conditions for which simple diagrams were used. The mean saccade length of Condition (simp-visu) was significantly longer than that of the other conditions.
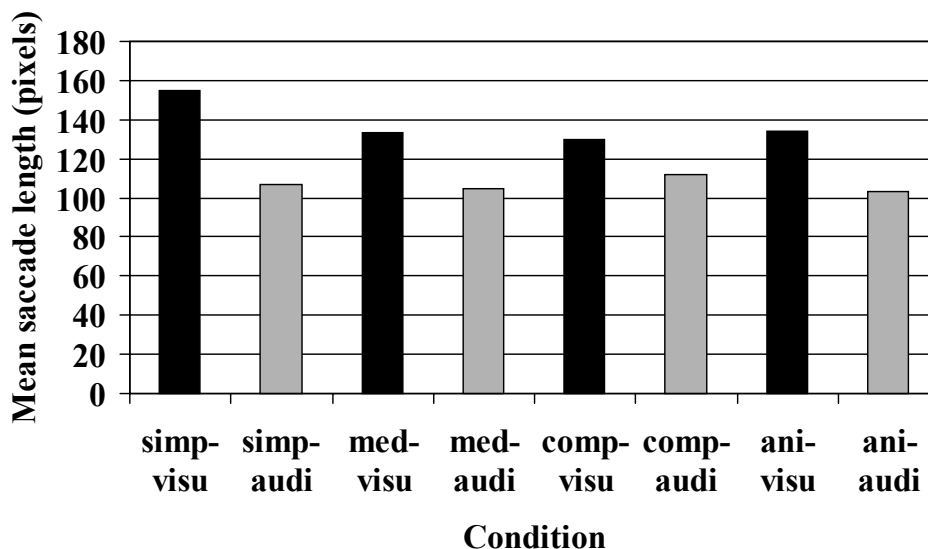
Figure 42: Mean saccade length as a function of experimental condition

Mean saccade length in the picture region

Figure 43 demonstrates the mean saccade length in the picture region of the eight experimental conditions. The mean saccade length in the picture region was not affected by picture complexity but by text mode ($F(1; 672) = 11.535; p < 0.005$) and by the interaction between picture complexity and text mode ($F(3; 672) = 2.802; p < 0.05$). Again, when text was presented auditorily (mean = 78.699 pixels), the mean saccade length in the picture region was significantly shorter than when visual text was presented (mean = 87.622 pixels). According to the Bonferroni tests, there was a significant difference between Condition (simp-visu) and Condition (ani-audi).
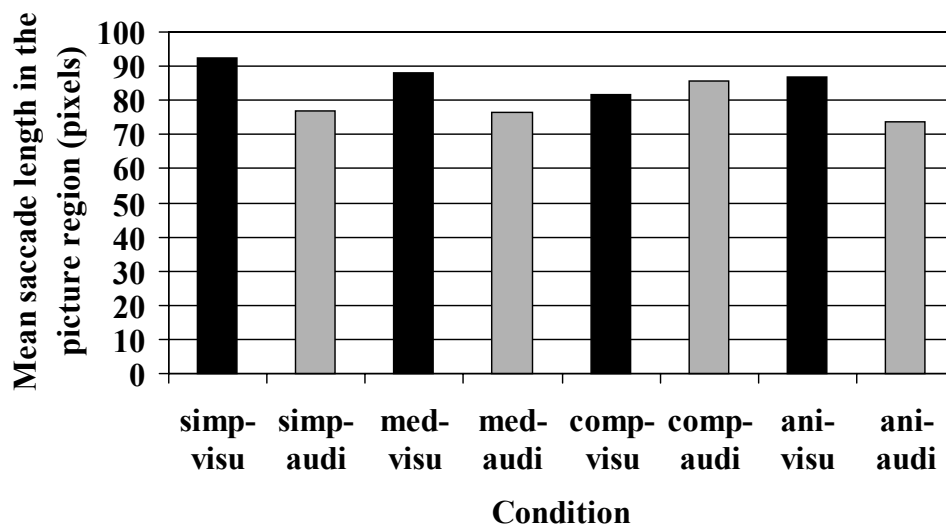
Figure 43 : Mean saccade length in the picture region as a function of experimental condition

The effect of the number of propositions

The number of propositions had a significant effect on the mean saccade length in the picture region. The mean saccade length in the picture region was inversely related to the number of propositions. However, this significant effect was restricted to when text was presented auditorily ($F(3; 332) = 4.320$; $p < 0.01$). When visual text was shown, there was no significant difference in the mean saccade length in the picture region. The difference in this effect was juxtaposed by Figures 44 and 45.
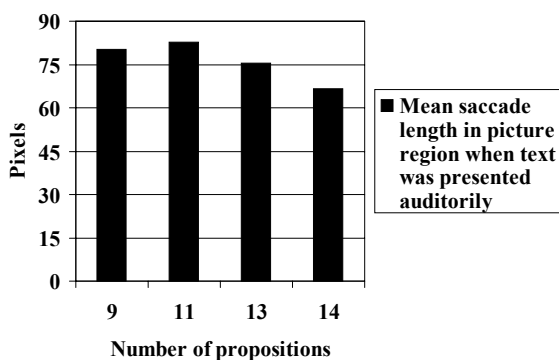
Figure 44: Mean saccade length was affected by the number of propositions
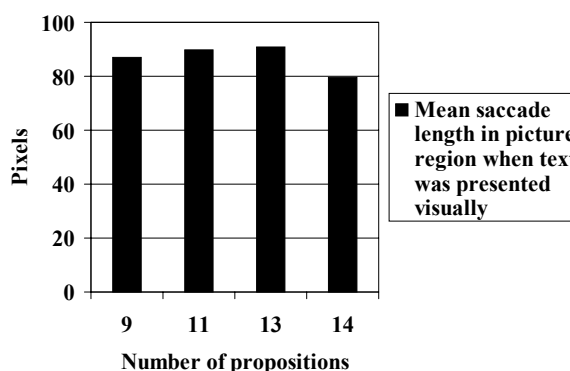
Figure 45: Mean saccade length was independent of the number of propositions

172

The effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects

The mean saccade length in the picture region was also significantly influenced by the number of objects when text was displayed visually (F(8; 253) = 2.757; p < 0.05) (see Figure 46). The Bonferroni tests indicated that the significant effect was exclusively due to the number of objects '4'. This corresponds to the two instructional sections (elephant and cannon) in Condition (simp-visu) and Condition (simp-audi) in which the objects in the diagrams were more sparsely arrayed than those in the diagrams of the other instructional sections. These arrangements possibly triggered longer saccades when subjects were viewing these two sections.
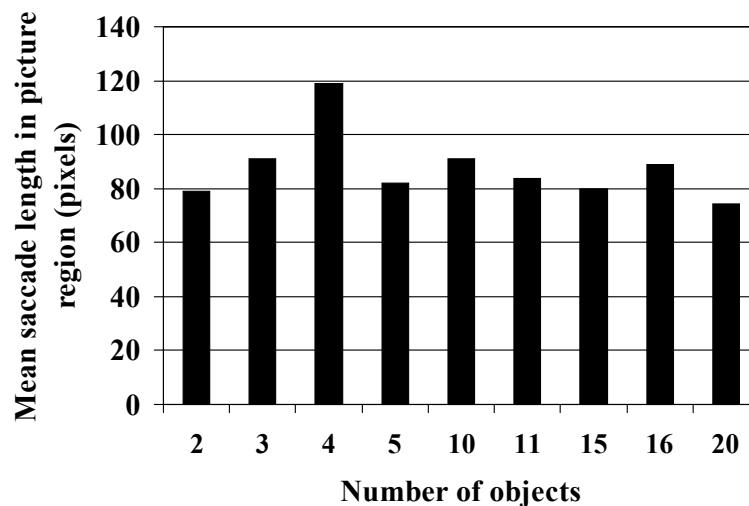
Figure 46: Mean saccade length in the picture region as a function of the number of objects when text was presented visually

When text was presented auditorily, the mean saccade length in the picture region was positively related to the number of objects (F(8; 254) = 5.946; p < 0.001) (see Figure 47). A t-test (t(261) = -2.498; p < 0.05) showed that the mean value for the sections with objects <= 10 (75.556 pixels) was significantly shorter than that for the sections with objects > 10 (84.204 pixels). Moreover, the effect of the number of objects on the mean saccade length in the picture region was not significant when animation was presented. Nevertheless, there was a tendency that indicated that the mean saccade length became larger as the number of objects increased (F(3; 151) = 2.329; p = 0.07).
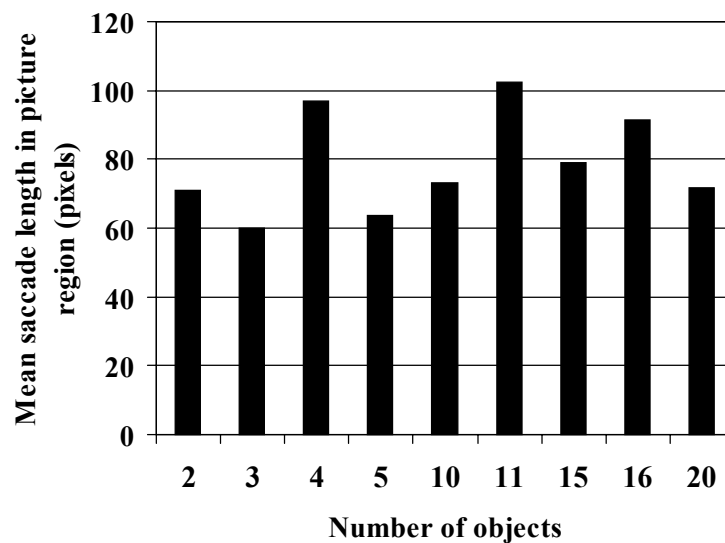
Figure 47: Mean saccade length in the picture region as a function of the number of objects when text was presented auditorily

Since the effect of the number of propositions on the mean saccade length in the picture region was opposite to the effect of the number of objects, an important question is which of the two factors had a stronger influence on subjects' eye movements. According to a multiple linear regression equation that captured the relationship between these variables, the following results were obtained:

$$\text{Mean saccade length} = 110.395 - 3.405 \times (\text{number of propositios}) + 0.707 \times (\text{number of objects})$$

The weight of the number of propositions (-3.405) was much larger than that in relation to the number of objects (0.707), which indicated that the number of propositions had a stronger effect on the mean saccade length in the picture region than did the number of objects.

The effect of the quantity of visual distractors was not significant, whereas the mean saccade length in the picture region was significantly affected by the quantity of relevant objects when text was presented auditorily ($F(2; 260) = 5.207$; $p < 0.01$) (see Figure 48). The mean saccade length in the picture region became larger as the quantity of relevant objects increased. According to the Bonferroni tests, 'low' was significantly lower than 'medium' and 'high'.
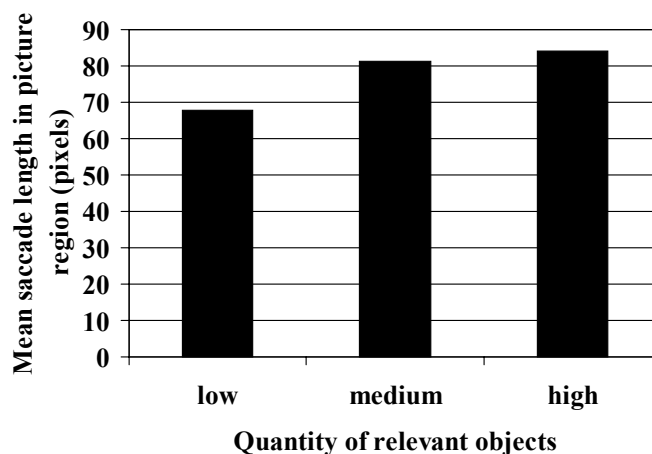
Figure 48: Mean saccade length in the picture region as a function of the quantity of relevant objects

Mean saccade length in the text region

The data are given in Figure 49. The mean saccade length in the text region was significantly influenced by picture complexity ($F(3; 340) = 7.410$; $p < 0.001$). The Bonferroni tests indicated that significant differences existed between Condition (simp-visu) and Condition (ani-visu) as well as Condition (comp-visu) and Condition (ani-visu). When pictures were static, there was no significant difference in the mean saccade length in the text region, whereas the mean saccade length in the text region became significantly longer when animation was presented.
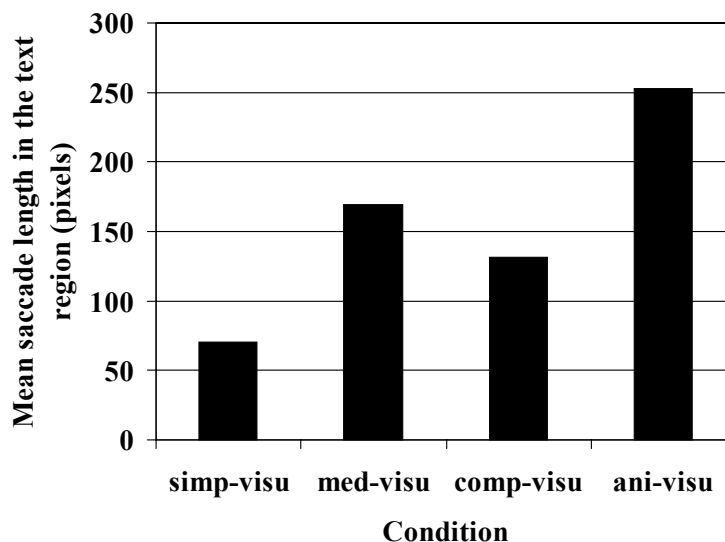
Figure 49: Mean saccade length in the text region of the corresponding experimental conditions

8.4.4.1.5 Number of gaze changes between the picture and the text region

This variable can yield information concerning how often subjects in the visual-only conditions switch their gazes between the picture and the text region, which can reveal their strategies for integrating the textual and pictorial information. The data are shown in Figure 50. The results showed that picture complexity had a significant effect on the mean number of gaze changes between picture and text ($F_{(3; 340)}$ = 4.221; $p < 0.01$). According to the Bonferroni tests, a significant difference was found to exist between 'medium' and 'animation'. The differences between 'simple', 'medium', and 'complex' were not significant. When animation was shown, subjects switched their gazes between text and picture significantly more often (mean = 9.598) than when static pictures were displayed (mean = 7.378) ($F_{(1; 342)}$ = 8.52; $p < 0.01$).
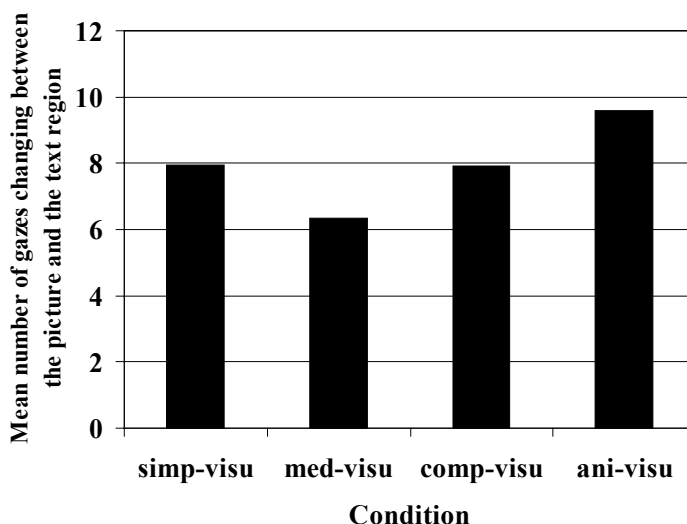
Figure 50: Mean number of gaze changes between the picture and the text region as a function of picture complexity

The effect of the number of propositions

The number of propositions had an effect on the mean number of gaze changes between the picture and the text region ($F_{(3; 340)}$ = 4.648; $p < 0.005$) (see Figure 51). The Bonferroni tests pointed out that the significant difference was between '9' and '11'.
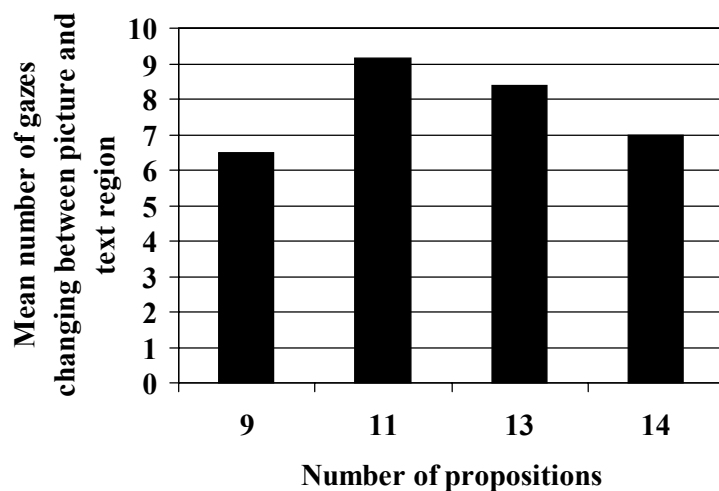
Figure 51: Mean number of gaze changes between the picture and the text region as a function of the number of propositions

Moreover, the number of objects, the quantity of visual distractors, and the quantity of relevant objects had no effect on the mean number of gaze changes between the picture and the text region.

8.4.4.1.6 Percentage of time spent in the picture region

This variable provided information concerning the percentage of time subjects spent inspecting the pictures when they were learning an instructional section. It was calculated by: (The sum of the fixation durations collected in the picture region) / (The section time of the given section)*100%. The results showed that when text was presented visually, the percentage of time spent in the picture region was significantly influenced by picture complexity ($F(3; 676) = 32.156$; $p < 0.001$) (see Figure 52).
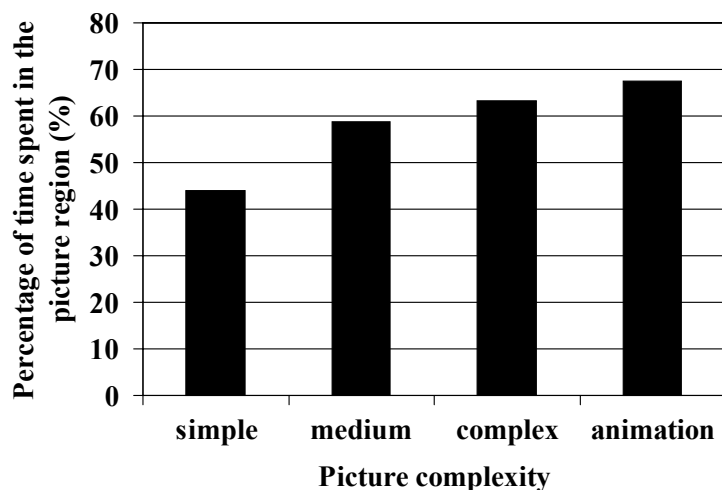
Figure 52: Percentage of time spent in the picture region as a function of picture complexity

Subjects that received simple static pictures with visual text spent significantly less time inspecting pictures than did subjects in other experimental conditions. According to the Bonferroni tests, 'simple' was significantly less than 'medium', 'complex', and 'animation'. Besides, 'animation' was significantly greater than 'medium'.

The effect of the number of propositions

The percentage of time spent in the picture region was independent of the number of propositions.

The effects of the number of objects, the quantity of visual distractors, and the quantity of relevant objects

The number of objects had a significant effect on the percentage of time spent in the picture region ($F_{(8; 516)} = 7.915$; $p < 0.001$) (see Figure 53). The data showed that the more objects there were in pictures, the more time subjects spent viewing these pictures.
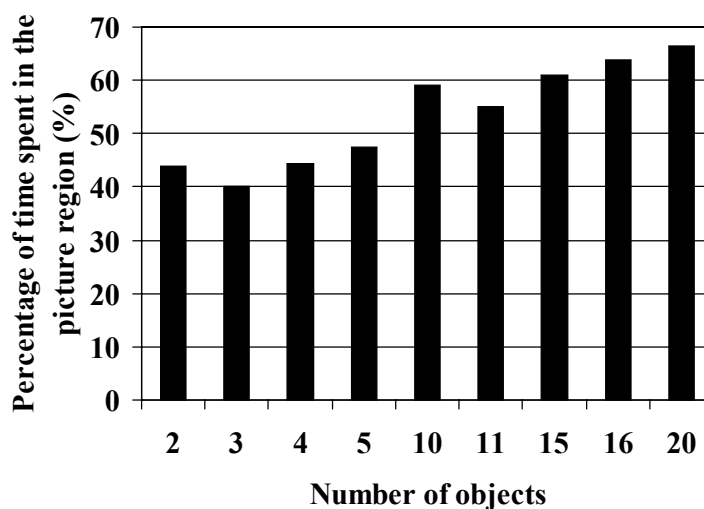
Figure 53: Percentage of time spent in the picture region as a
function of the number of objects.

Moreover, the effects of the quantity of visual distractors ($F_{(2; 522)}$ = 30.683; p <
0.001) and the quantity of relevant objects ($F_{(2; 522)}$ = 18.593; p < 0.001) were also
significant. Figure 54 shows that the percentage of time spent in the picture region
was positively related to the quantity of visual distractors. The Bonferroni tests
indicated that 'no' was significantly lower than 'medium' or 'high'. The difference
between 'medium' and 'high' was not significant. Overall, when the visual distractors
were absent, subjects spent substantially less time in the picture region than did
subjects that received visual distractors.
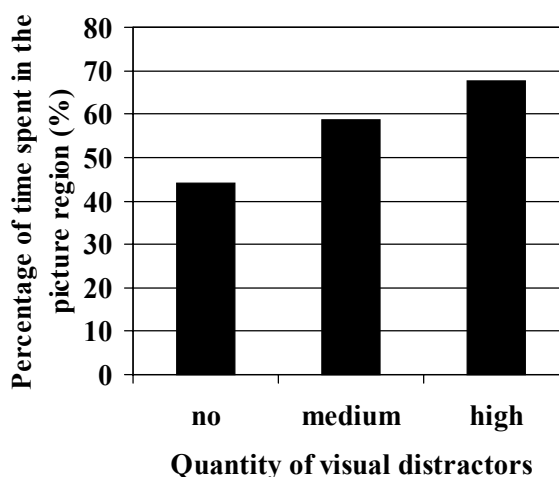
Figure 54: Percentage of time spent in the picture region as a
function of the quantity of visual distractors

Figure 55 depicts the effect of the quantity of relevant objects. The percentage of time spent in the picture region was positively related to the quantity of relevant objects. The Bonferroni tests showed that 'low' was significantly lower than 'medium' and 'high'. The difference between 'medium' and 'high' was also significant.
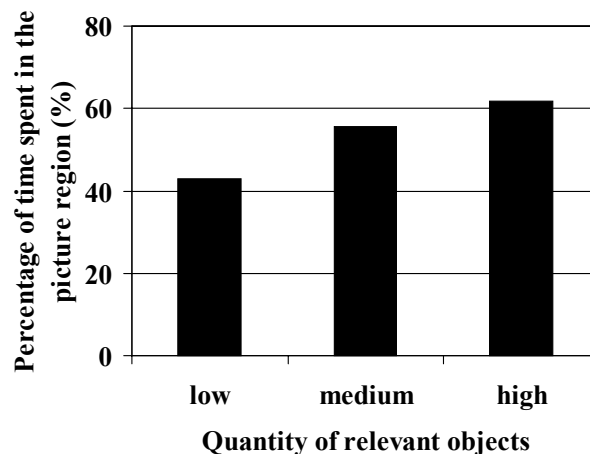
Figure 55: Percentage of time spent in the picture region as a function of the quantity of relevant objects

8.4.4.1.7 Percentage of time spent in the text region

This variable provides information regarding the percentage of time subjects spent reading the text when they were learning an instructional section. It was calculated by: (The sum of the fixation durations collected in the text region) / (The section time of the given section)*100%. The percentage of time that subjects spent in the text region was significantly affected by the experimental conditions ($F(3; 340) = 7.827$; $p < 0.001$). The data are depicted in Figure 56. Subjects that received diagrams with 'simple' or 'medium' complexity spent significantly more time reading text than did subjects that received complex diagrams or animation with visual text. The Bonferroni tests showed that the differences between Condition (simp-visu) and Condition (comp-visu), Condition (simp-visu) and Condition (ani-visu), Condition (med-visu) and Condition (comp-visu) as well as Condition (med-visu) and Condition (ani-visu) were significant. In addition, the number of propositions had no effect on the percentage of time spent in the text region.
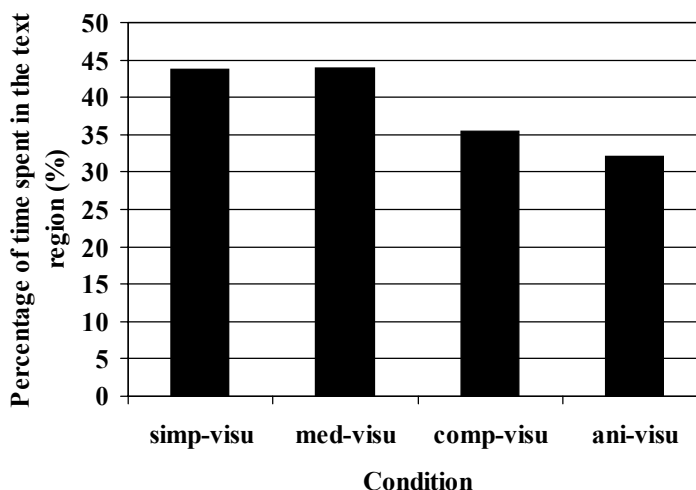
Figure 56: Percentage of time spent in the text region of the corresponding experimental conditions

8.4.4.1.8 Percentage of time spent in the picture versus in the text region

By comparing the percentage of time subjects spent inspecting the pictures with the percentage of time that they spent reading the text, we can find out whether subjects paid more attention to the pictures or to the written texts. Figure 57 depicts the data.
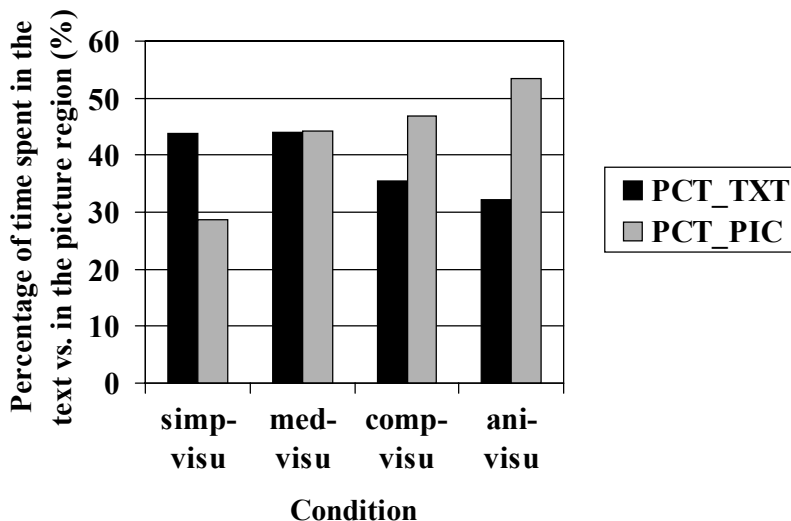
Figure 57: Percentage of time spent in the picture vs. in the text region. Percentage of time spent in the text region was marked in black whereas that spent in the picture region was marked in gray.

Pairwise comparisons using t-tests showed that subjects in Condition (simp-visu) (t(97) = -4.237; p < 0.001) spent substantially more time reading text than viewing the pictures while subjects in Condition (comp-visu) (t(72) = 2.837; p < 0.001) and

Condition (ani-visu) (t(81) = 5.477; p < 0.001) spent significantly more time viewing pictures or animation than reading text.

8.4.4.2 Qualitative analyses of eye-movement data

As to the qualitative analyses of eye-movement data, which were carried out as case studies, some subjects' eye-movement trajectories were analyzed to investigate how the subjects integrated the textual and pictorial information, and how their visual attention was spatially and temporally distributed. In the following subsections, I shall report on the eye-movement trajectories of three subjects: one subject was from Condition (comp-audi); another one was from Condition (med-audi), and the other one was from Condition (med-visu).

8.4.4.2.1 Case 1

The first example of a subject's eye-movement trajectories is shown in Figure 58. Each number represents a fixation position on the display of the given section in chronological order. The subject was for the first time inspecting the complex diagram and listening to the verbal instructions regarding the movement rules for the Chinese chess piece referred to as a 'cannon'. During this section there were all in all 56 fixations enumerated by their temporal order and plotted on the screen shot of the Section 'cannon'. The corresponding fixation durations are listed in Table 5. In addition, a time line is given to align auditory and oculomotor events.

As the data showed, during the first 1305 milliseconds, the subject was looking at the menu buttons. During that time three fixations were located at the label 'Kanone', while the subject was hearing the words 'die Kanone' (the cannon). The inspection of the diagram started at the fixation 8 (f. 8). From f. 20 to f. 28, the subject looked at the horizontal direction, and then examined the vertical direction from f. 29 to f. 40. While the auditory text was saying 'zum Schlagen' (to capture), a longer fixation duration (f. 33: 552 ms) was observed at the critical position on the diagram where a red flame signified that the cannon had captured the opponent's knight. Shortly after the second sentence ended, the subject successively considered the 'cannon' positions at the bottom as well as in the middle of the diagram, and the captured 'knight' of the opponent, which indicated that she was trying to understand what the second sentence meant.

Furthermore, the most frequently fixated piece in the diagram was the 'cannon' in its starting position because a large number of fixations were directed at this piece. This was one of the most crucial visual elements for understanding the

information illustrated by the diagram. Besides, longer fixation durations were especially observed in relation to the 'cannon' in the left end position (f. 22: 596 ms; f. 41: 676 ms) as well as in regard to the captured 'knight' of the opponent (f. 14: 404 ms; f. 33: 552 ms), and the starting position (f. 28: 408 ms). On the other hand, there were a small number of fixations located at some visual distractors that were close to the relevant chess pieces. The other visual distractors that were a bit far from the relevant pieces were simply ignored. That is, visual distractors did distract visual attention under the condition that they were spatially near the relevant visual elements.

To sum up, the eye-movement trajectories revealed the dynamics of the subject's visual attention while she was listening to the verbal instructions and trying to integrate them with the pictorial information presented in a visually demanding diagram. The relevant objects in the diagram received considerably more or longer fixations than the visual distractors. The subject's eye movements were affected by the auditory text, which could be observed from her inspection of the relevant objects that the auditory text referred to, shortly after she heard that text.

Figure 58: An example of a subject's gaze trajectories, while the subject was hearing the text and inspecting the complex diagram regarding the 'cannon'.

0   1305 (ms)   2804 (ms)   4165 (ms)   5228 (ms)   6137 (ms)   7405 (ms)

   (f. 1 - 7)   (f. 8 - 12)   (f. 13 -16)   (f. 17 - 20)   (f. 21 - 22)   (f. 23 - 28)
Die   Kanone bewegt sich in einem Zug beliebig viele Einheiten   horizontal   oder vertikal in beliebiger Richtung.
(The   cannon   moves   in one move   arbitrarily many units   horizontally   or vertically in any   direction.)

7405 (ms)   8209 (ms)   9060 (ms)   9487 (ms)   10260 (ms)   11501 (ms)

   (f. 29 - 31)   (f. 32 - 34)   (f. 35 - 36)   (f. 37 - 40)   (f. 41- 44)
(P a u s e   ) Zum Schlagen   muss   sie   über eine Figur   springen.
(p a u s e   ) (To capture,   it   must   jump over   a   piece.)

Nr. of fix.	1	2	3	4	5	6	7	8	9	10	11	12
FD (ms)	144	108	184	188	48	72	96	268	96	232	140	192
Nr. of fix.	13	14	15	16	17	18	19	20	21	22	23	24
FD (ms)	100	404	312	228	196	196	300	320	192	596	108	300
Nr. of fix.	25	26	27	28	29	30	31	32	33	34	35	36
FD (ms)	216	148	200	408	176	144	156	148	552	228	168	316
Nr. of fix.	37	38	39	40	41	42	43	44	45	46	47	48
FD (ms)	188	352	172	192	676	140	216	288	272	308	144	264
Nr. of fix.	49	50	51	52	53	54	55	56				
FD (ms)	112	176	272	256	152	240	212	268				

Table 5 shows the fixation durations of the fixations depicted in Figure 58

8.4.4.2.2 Case 2

The second example shows the eye-movement trajectories of another subject who was viewing the diagram with medium complexity and listening to verbal instructions regarding the movement rules for the 'soldier' for the first time. The eye-movement trajectories are depicted in Figure 59, and the corresponding fixation durations are shown in Table 6. Again, a time line is given to align auditory and oculomotor events.

Similarly, there was clear evidence of the influence of the auditory text on the viewing patterns of the subject. The subject cast a very short glance at the menu button 'Soldat', while she was hearing the words 'der Soldat'. She first looked at the starting position of the 'soldier' (the blue dot below on the left in the diagram), the black circle above it, and then shifted her fixation to the blue arrow (to the left of the blue dot), and looked back to the starting position and upwards to the black circle again, when she was hearing "eine Einheit vorwärts (one unit forwards)." At f. 25 (412 ms) and f. 29 (544 ms), she looked at the black circle above the starting position with long fixation durations, while the zone of the opponent was just mentioned by the text because the circle was located in the zone of the opponent. Finally, as the auditory text was saying that the 'soldier' may never move backwards, her gazes were directed toward the black circle with a red cross below the second starting position of the 'soldier' (the blue dot above to the right in the diagram), and moved between the second starting position and the black circles twice. Immediately after that, she switched her gazes to the blue arrow below the first starting position of the 'soldier', and looked downwards toward the black circle with a red cross. This shows clearly that the subject was trying to integrate the verbal and the pictorial information that referred to each other.

In this example, the majority of the fixations were located around the first starting position of the 'soldier', the blue arrows as well as the black circles above, below, and to the left of it. There were only two short fixations located close to the visual distractors, which indicated that the visual attention of this subject was not much diverted by the visual distractors. Most of the time, the subject's eye movements were related to the verbal instructions. As soon as the information delivered by the auditory text was processed, the subject's eyes immediately reacted to it, which provided clear evidence that the subject's eye movements were closely time locked to the auditory text. That is, the allocation of attention could well be observed based on the fixation positions in such a learning situation. In addition, the instantaneous reaction of the eyes to the auditory verbal input during a sentence also corresponds to the immediacy principle of language processing proposed by Just and Carpenter (1980, 1987).

However, it should be noted that the instant reaction of the eyes to the auditory text could be observed more frequently if the visual input was not very complicated. By comparing the eye-movement trajectories of case 1 with those of case 2, it was found that there was some delay in the reaction of eyes to the auditory text when the diagram was visually complicated. In this case, the subject usually reacted to the text shortly after she heard the text, but not immediately. This is a typical PRP (psychological refractory period) effect when a subject is performing a dual task. Due to the limited capacity of working memory, a trade-off in attentional resource must take place when two tasks have to be performed simultaneously. A longer delay in the reaction time of the second task can be observed if the first task is cognitively demanding.

8.4.4.2.3 Case 3

The third example shows the eye-movement trajectories of a subject who was reading the text and viewing the diagram with medium complexity. The eye-movement trajectories are shown in Figure 60, and the corresponding fixation durations are given in Table 7. The immediacy of language processing can also be observed in that the subject switched her eyes from the text to the diagram instantly at the end of a phrase, but not at the sentence boundary. The change in fixation positions was an indicator of the integration of the textual and pictorial information. This activity was revealed at two places. The first one was from f. 13 to f. 16. The subject read the first sentence until the word 'diagonal', and then instantaneously moved her eyes to the diagram where the starting position of 'guard' and the two blue arrows to the right of it were located. Her fixation jumped back to the word 'diagonal' at f. 16, and the reading continued. From f. 28 to f. 31, the eye-movement trajectories clearly showed that the subject looked at the green line (at f. 29) in the diagram immediately after she read the second sentence as far as the words 'mit grünen Linien markierte' (marked with green lines).

Most of the fixations in the picture region were aimed at the starting position of the 'guard' (the blue dot), the two blue arrows to the right of it, and the black circle above to the right. Longer fixation durations could also be observed in these locations. Since the instructions were all presented visually, the integration of text and picture information required switching fixations between text and picture.
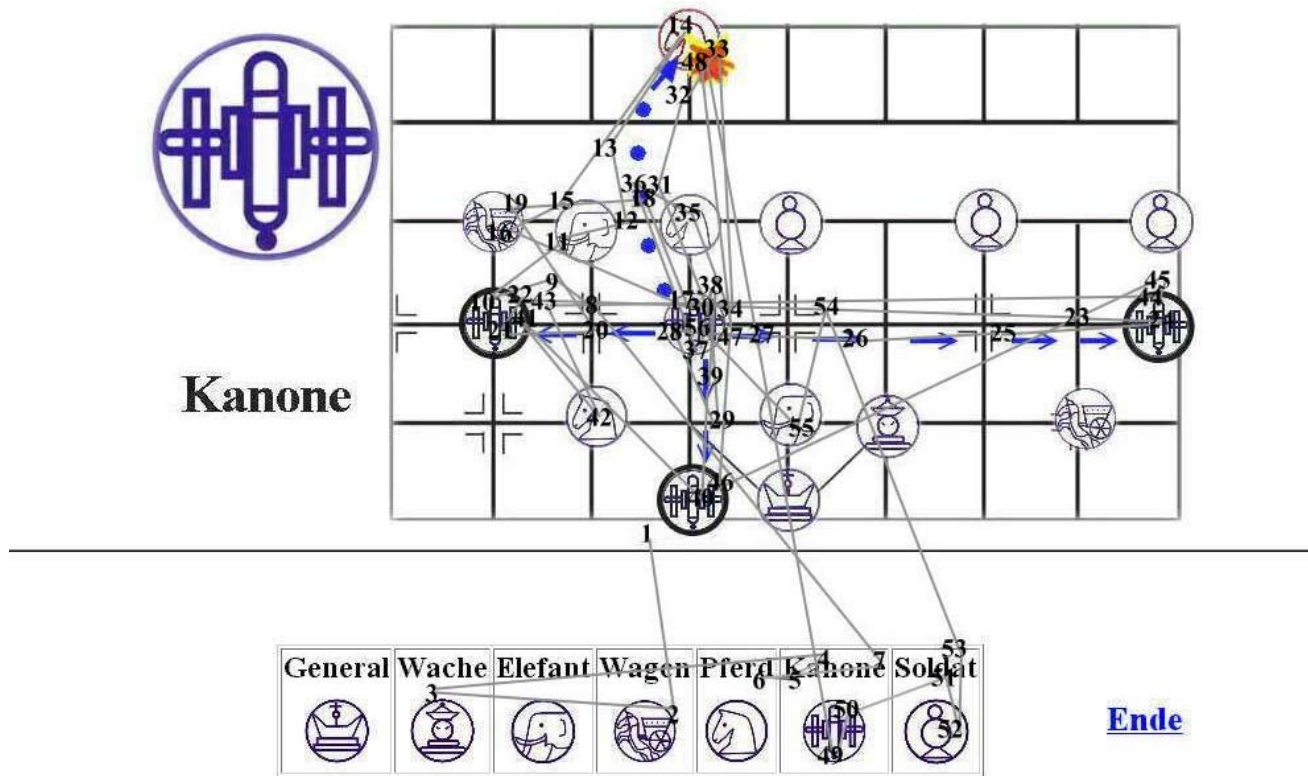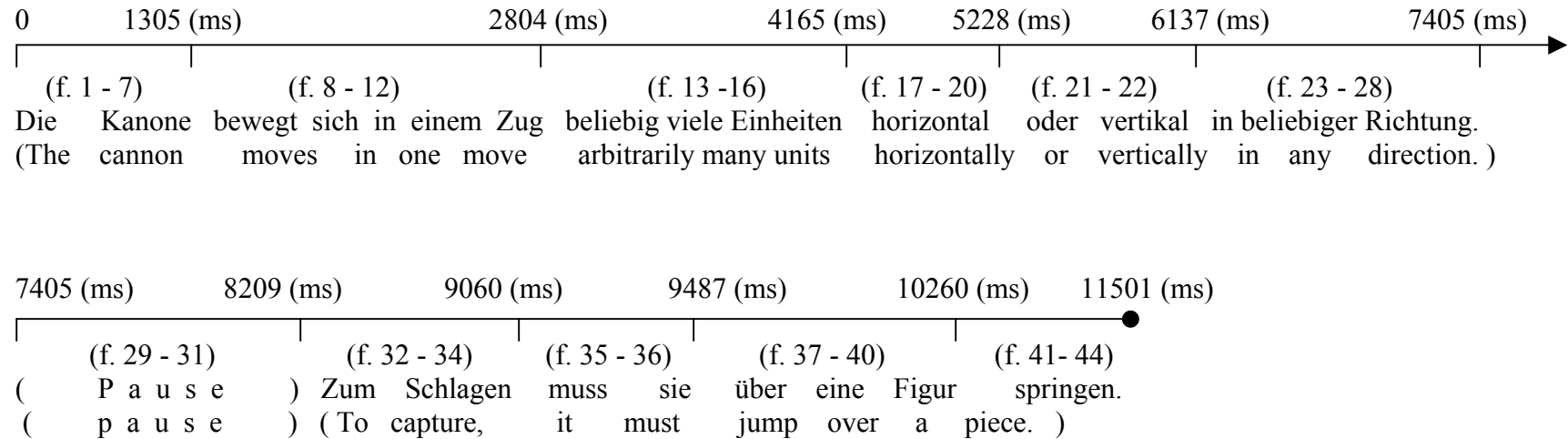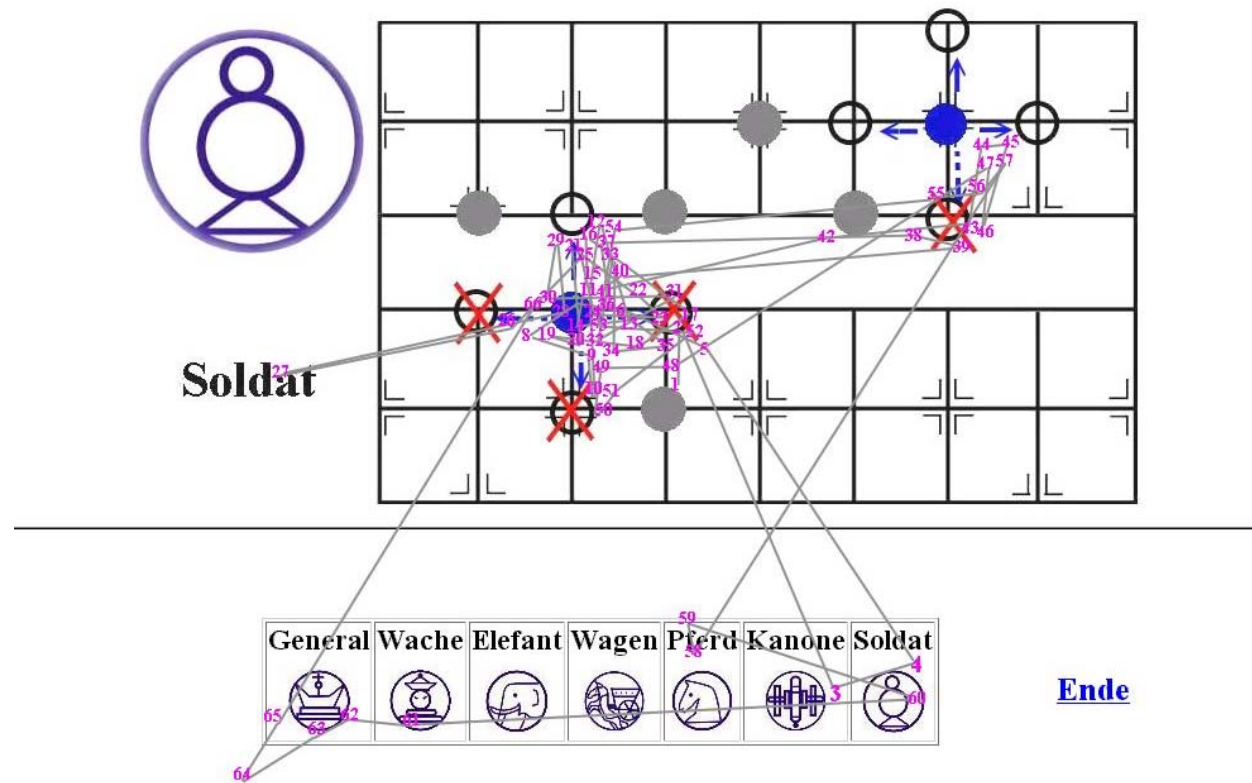
Figure 59: An example of a subject's gaze trajectories, while the subject was hearing the text and inspecting the diagram with medium complexity

0   1175 (ms)   2693 (ms)   4166 (ms)   4666 (ms)   5415 (ms)   6783 (ms)

(f. 1 - 6)   (f. 7 - 10)   (f. 11 - 15)   (f. 16 - 17)   (f. 18 - 19)   (f. 20 - 25)

Der   Soldat bewegt sich in einem Zug   eine Einheit vorwärts.   (P a u s e) Nachdem er   in die Zone des Gegners

(The soldier   moves in   one   move   one unit forwards.) (p a u s e) (After he has moved into his opponent's zone,

6783 (ms)   7799 (ms)   8685 (ms)   9368 (ms)   10051 (ms)   10624 (ms)   11037 (ms)   12164 (ms)

13751 (ms)

(f. 26 - 28)   (f. 29 - 30)   (f. 31 - 32)   (f. 33 - 35)   (f. 36 - 37)   (f. 38 - 39)   (f. 40 - 44)   (f. 45 - 50)

gegangen   ist, darf er sich auch eine   Einheit nach links   oder   rechts   bewegen, aber   nie   rückwärts.

he may also move one   unit   to the left or the right,   but   never   backwards.)

Nr. of fix.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
FD (ms)	292	132	136	180	224	268	372	196	336	340	204	304
Nr. of fix.	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
FD (ms)	260	168	232	232	328	264	208	276	260	160	212	212
Nr. of fix.	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
FD (ms)	412	136	116	332	544	208	256	260	308	240	256	176
Nr. of fix.	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
FD (ms)	420	192	200	204	220	148	236	380	148	288	344	204
Nr. of fix.	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
FD (ms)	728	312	96	184	492	176	192	292	96	80	80	124
Nr. of fix.	**61**	**62**	**63**	**64**	**65**	**66**	**67**					
FD (ms)	564	240	244	240	404	208	336					

Table 6 shows the fixation durations of the fixations depicted in Figure 59
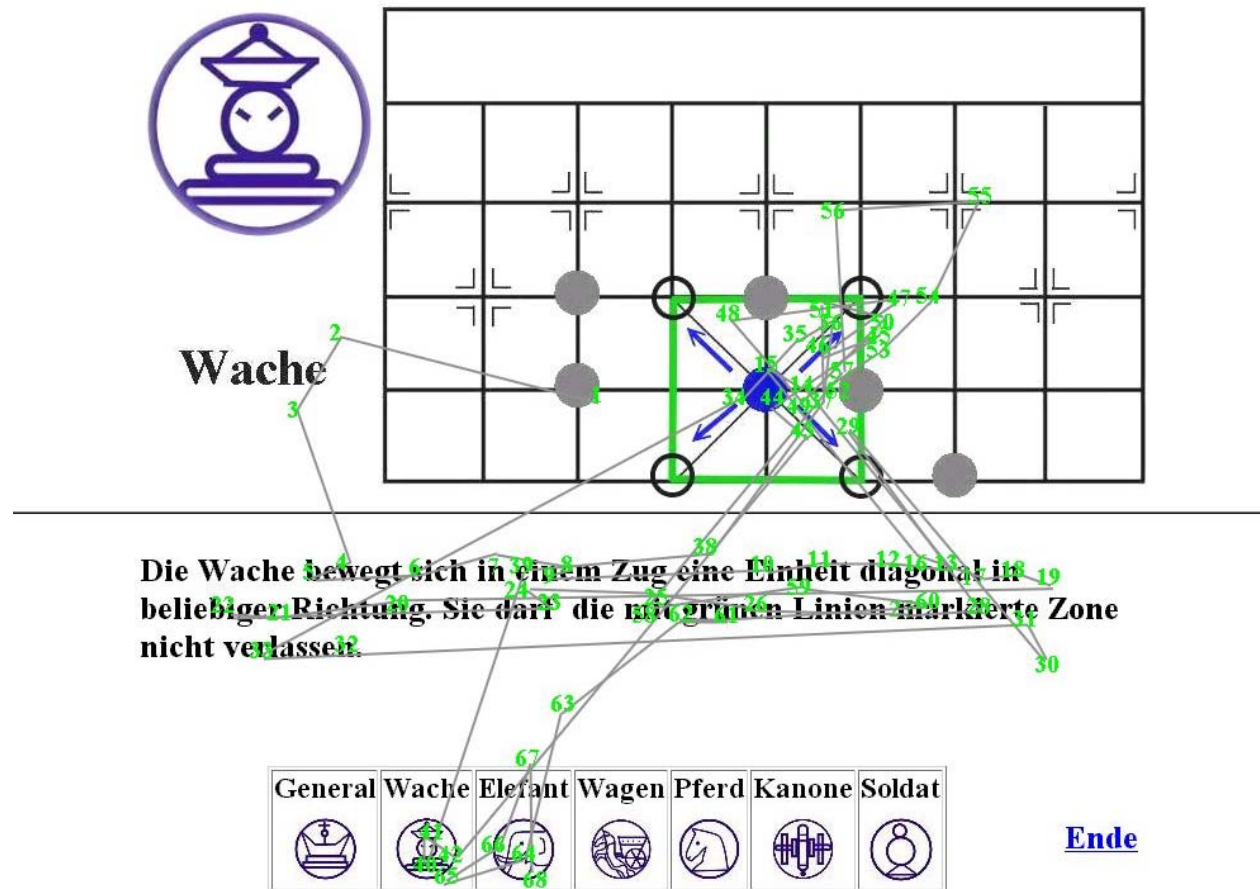
Figure 60: An example of a subject's gaze trajectories, while the subject was reading the text and inspecting the diagram with medium complexity

Die Wache bewegt sich in einem Zug eine Einheit diagonal in beliebiger Richtung.

(The guard   moves   in one move one unit diagonally in any   direction.)

 Sie darf die mit grünen Linien markierten Zone nicht verlassen.

(He may not leave the zone marked by the green lines.)

Nr. of fix.	1	2	3	4	5	6	7	8	9	10	11	12
FD (ms)	184	128	116	140	180	144	180	132	236	260	180	172
Nr. of fix.	13	14	15	16	17	18	19	20	21	22	23	24
FD (ms)	132	584	220	148	208	200	216	116	128	272	128	156
Nr. of fix.	25	26	27	28	29	30	31	32	33	34	35	36
FD (ms)	192	128	272	176	232	116	124	108	208	236	236	196
Nr. of fix.	37	38	39	40	41	42	43	44	45	46	47	48
FD (ms)	168	136	176	152	276	144	276	240	352	184	400	480
Nr. of fix.	49	50	51	52	53	54	55	56	57	58	59	60
FD (ms)	352	248	308	196	212	236	272	104	684	224	224	136
Nr. of fix.	61	62	63	64	65	66	67	68				
FD (ms)	160	364	124	104	224	152	536	336				

Table 7 shows the fixation durations of the fixations depicted in Figure 60

8.4.5 Questionnaire results

The same questionnaire as in the previous experiment was employed in this experiment. Subjects were asked to give their ratings of the comprehensibility of the texts and pictures (diagrams or animations) presented in the learning materials. The ratings were based on a five-level scale ranging from 1 (very difficult) to 5 (very easy). Besides, subjects had to choose which element(s) ('texts', 'pictures' or 'both') gave them the most help in understanding and memorizing the materials.

The results are shown in Table 8. Most of the subjects rated the comprehensibility of the texts and pictures as easy (both with median = 4). Texts and pictures together were regarded as the essential element for understanding, whereas pictures were preferentially chosen as the aid for memorizing the content of instruction.

Comprehensibility of		Crucial element(s) for	
text	pictures	understanding	memorization
60% easy (median = 4)	**60%** easy (median = 4)	**75%** both text and pictures	**52.5%** pictures

Table 8: Results of the questionnaire shown in accordance with the SMP- and the DMP-conditions

8.5 Discussion

8.5.1 Learning results

In Experiment 2, I attempted to examine the hypotheses that subjects can only learn more efficiently with dual modality presentation (DMP) if the pictorial information is neither too easy nor too complicated; if the information conveyed by the pictures is very easy, the DMP is not necessarily better for learning than the SMP because the capacity of the learners' working memory (especially the CE) in the SMP condition is not exhausted, and therefore, they can process pictorial and textual information without being less efficient. In contrast, if the pictorial information is very complicated, the

capacity of the CE would be more likely exceeded by the DMP. In this case, learners will fail to process the verbal information even though the text is presented auditorily.

The results of this experiment were consistent with most of the hypotheses. As predicted, DMP was not superior to SMP when pictorial information was very easy or very complicated. When the demand for visual information in the pictures was high, the error rate of the DMP-conditions was substantially higher than that of the SMP-conditions. This evidence goes against the ubiquitous belief that DMP or a multi-modality presentation should be superior to SMP. On the contrary, DMP or a multi-modality presentation sometimes can be detrimental to learning because any extra modality used for presentation will additionally require a certain amount of cognitive resources to process the information presented in that presentation mode. Furthermore, the hypothesis that DMP might facilitate the processing of text and picture integration if the picture complexity is at a medium level is not confirmed in this experiment. Although the error rate of Condition (med-audi) was slightly lower than that of Condition (med-visu), the difference between them was not significant. The reason for this result might lie in the fact that the picture complexity did not reach the "real" medium level.

The assumption that DMP might facilitate information processing if working memory is not overloaded by the total amount of information that has to be processed simultaneously is in line with the following result, namely, that the error rate of Condition (ani-audi) was substantially lower than that of Condition (com-audi). It is assumed that, if animation can guide subjects' visual attention toward the relevant information shown in the complex diagrams, the superiority of DMP can be restored. That is, animation is supposed to reduce the complexity of the complex diagrams and to release CE from being overloaded by the large amount of information to be processed at the same time. Finally, the results showed that when animation was presented with visual text, subjects' performance was the worst among the conditions, which was a quite robust result throughout my research. The reason for this result is the same as the one I mentioned in Experiment 1. Since subjects cannot view the animation and read the text simultaneously, they have to either deal with these two sources of information consecutively, or to try to process these two sources of information by switching their gazes between text and animation both rapidly and frequently. According to the eye-movement data recorded on the video tapes, the subjects in relation to Condition (ani-visu) did employ the speedy eye-movement strategy. However, the speedy eye movements obviously cannot integrate the text and animation information successfully.

Moreover, the error rate of Condition (comp-visu) was much lower than that of Condition (comp-visu), which indicates that the cognitive load induced by Condition (comp-visu) is not nearly as heavy as that caused by Condition (comp-audi), which contradicts the modality effect suggested by the cognitive load theory.

8.5.2 Learning behavior

In addition to the results of the analyses of learning, the factors that influenced the learning time and the section time were also investigated. The reason for examining learning time or section time was to find out how and why different combinations of text presentation mode and picture complexity affected the amount of time subjects required for processing the information in the instruction. Further factors such as the number of propositions and the number of objects, etc. were also taken into account because they can yield more detailed information about how learning time changed as the quantity of information to be processed varied.

As the results indicated, the mean learning time was independent of the combinations of picture complexity and text mode. However, the mean section time was significantly affected by the combinations of picture complexity and text mode, the number of propositions, the quantity of visual distractors as well as by the quantity of relevant objects in pictures. The reason why the experimental conditions did not have an effect on the mean learning time but rather on the mean section time was that the subjects were allowed to view the instruction as long and as often as they wanted. Therefore, as subjects repeated the instructional sections with varying frequency, the experimental conditions were not able to affect the mean learning time but instead affected the mean section time and the mean section frequency.

According to the data, the mean section time of Condition (med-visu) was significantly shorter than that of the other conditions. I assume that the diagrams shown in Condition (med-visu) depicted all the possible positions that the Chinese chess pieces would occupy. In this case, the subjects did not need to infer all the possible positions by themselves. In contrast, subjects in Condition (simp-visu) could only see one or two possible positions in the diagrams. Although there were all in all fewer objects in the diagrams, subjects in Condition (simp-visu) conceivably tended to infer the rest of the possible movement positions by themselves, which is probably why the section time was longer.

With respect to the effect of the number of propositions and the number of objects on the mean section time, as one might predict, the mean section time was a

positive function of the number of propositions and the number of objects. The more information that was to be processed, the longer that the processing time was.

Due to a negative correlation between the mean section time and the mean section frequency ($r = -0.154$; $n = 680$; $p < 0.001$), subjects who viewed every single instructional section quickly tended to repeat each section more frequently. However, certain experimental conditions, like Condition (comp-audi) or Condition (ani-visu), made learning more difficult. As a result, the error rate of those conditions was high even though subjects repeated the sections more often or viewed each single section for a longer period of time. It appears that the inappropriateness of the two presentation conditions impaired learning to such a degree that it could not simply be recovered by longer section time or higher section frequency.

Animation, on the other hand, takes more time to view. Subjects might get the impression that they understood and remembered the information displayed by means of animation very well during the long viewing time, and might, therefore, repeat instructions less frequently. Moreover, the mean section frequency was negatively related to the number of propositions and the number of objects. This is a sort of parallel effect that resulted from the negative correlation between the mean section time and the mean section frequency because mean section time is positively related to the number of propositions and the number of objects. The reason for this could lie in the fact that the more time subjects spent on a section, the better they could memorize the information to be learned, which in turn leads to less repetitions of the instructional sections.

8.5.3 Interpretation of the eye-movement data

In this section, I would like to discuss the eye-movement data while considering one by one the effects of text mode, picture complexity, the number of propositions, the number of objects as well as the quantity of visual distractors and quantity of relevant objects on subjects' eye-movement behavior.

8.5.3.1 Text mode

The effects of text mode are briefly summarized in Table 9. When text was presented auditorily, subjects' eye movements were slower whereby their fixation duration was longer and saccade length was shorter. This can be attributed to the fact that subjects had to perform the visual and auditory tasks at the same time. Since they had to pay attention to the auditory text, their attention resource for processing the pictorial information was

accordingly less than that of the subjects in the visual-only conditions. This kind of tradeoff between the auditory and visual tasks was clearly revealed in the eye-movement data which can be characterized by a longer fixation duration, a lower fixation rate and shorter saccade length. Similar results were obtained in a study by May et al. (1990), indicating that the range of saccadic extent decreased significantly as the complexity of the auditory task (tone counting) was increased.

Text mode	Mean fixation duration	Mean fixation duration in picture	Mean number of fixations	Mean number of fixations in picture	Mean fixation rate	Mean fixation rate in picture	Mean saccade length	Mean saccade length in picture
auditory	longer	longer	smaller	greater	lower	lower	shorter	shorter
visual	shorter	shorter	larger	smaller	higher	higher	longer	longer

Table 9: An overview of the effects of text mode on the eye movements

8.5.3.2 Picture complexity

The discussion of the effects of picture complexity focuses on the eye-movement data collected *in the picture region* of the instructional display because they can yield pure information about how picture complexity influences the way subjects view the pictures. Since the effects of picture complexity are not linear, it is difficult to discuss them in general. Therefore, I will discuss the effects one by one.

1) *Mean fixation duration*:
The mean fixation duration was longer when 'animation' was presented. Due to the interaction between picture complexity and text mode, the mean fixation duration of 'simple' diagrams was significantly longer than that of 'medium' and 'complex' diagrams when text was presented auditorily, whereas this effect did not exist when text was presented visually. This is quite an interesting and essential phenomenon that has been noticed, and it gives rise to the difference between the parallel and sequential processing of textual and pictorial information. When auditory text is presented, the textual and pictorial information have to be processed simultaneously. Fixation durations are significantly shorter when there are many objects in the pictures because

196

subjects try to integrate the auditory information with the information from the objects that are scattered in different locations in the picture. Rapid eye movements could be a more efficient strategy when subjects have to process the auditory information and view pictures with many objects in parallel. Since there are only a few objects in the simple diagrams, the mean fixation duration of Condition (simp-audi) is longer than that of Conditions (med-audi) and (comp-audi).

On the contrary, when text is presented visually, the textual and pictorial information processing are carried out more sequentially. In other words, the integration of textual and pictorial information is executed by switching gazes between text and pictures. Hence, subjects' eye movements in the picture region are not much affected by the parallel processing of textual information. As the data indicated, the differences in the mean fixation durations between the three levels of picture complexity ('simple', 'medium', and 'high') were not significant when visual text was presented.

The mean fixation duration of 'animation', on the other hand, was significantly longer than that of the experimental conditions with static pictures, irrespective of the presentation mode of text. The reason for this was probably that subjects most of the time followed the motion of the chess piece moving on the chess board when they were viewing the animation. The speed of the subjects' eye movements was controlled by the speed of movement of the chess piece, which was certainly much slower than the normal eye-movement speed when the subjects regarded static pictures. It should be noted that the measurement of fixation duration in the animation was probably not confounded by the effect of smooth pursuit because the movements shown in the animation were not based on slow or smooth motion but simply showed a series of alternative positions for a chess piece moving from one point to another. Therefore, it is unlikely that the eye-tracker will merge the fixations in rapid succession to a longer one as might occur when tracking very slight eye movements (smooth pursuit).

2) *Mean number of fixations*:

The mean number of fixations in relation to the 'simple' diagrams was significantly smaller than those in regard to the 'medium' and 'complex' pictures as well as 'animation'. By taking the effect of text mode into account, when text was presented visually, significant differences were found to exist between 'simple' and 'complex' as well as between 'simple' and 'animation'. Basically, the data pointed out that the number of fixations increased as the picture complexity (or the number of objects, respectively) increased. A straightforward explanation is that the more objects were scattered in the diagram, the more fixations were required to "pick up" the information.

3) *Mean fixation rate*:

The mean fixation rate was lower when animation was presented. The reason for this was conceivably that subjects followed the motion of animation, and therefore, their eye movements were substantially slower than when they were viewing static pictures. When visual text and static pictures were presented, the mean fixation rate in the picture region did not differ between complexity levels, whereas the mean fixation rate of 'complex' pictures was substantially higher than that of 'simple' pictures when auditory text was presented. The reason for this, as I mentioned before, could be that subjects must employ rapid eye movements to integrate textual and pictorial information if there are many objects in the picture, or if textual and pictorial information have to be processed at the same time. This can be regarded as a kind of "time-pressure" effect. In contrast, if textual and pictorial information can be processed sequentially, the "time-pressure" effect disappears.

4) *Mean saccade length*:

The mean saccade length was independent of picture complexity, but there was an interaction between picture complexity and text mode. As the results showed, the mean saccade length of Condition (simp-visu) was greater than that of Condition (ani-audi). For Condition (simp-visu), there were 2 to 5 objects in a diagram, which were sparsely arranged on the chess board. On the other hand, subjects often switched their gazes between the text and picture regions. Thus, many long saccades were made. For Condition (ani-audi), subjects only had to follow the movements of a chess piece, and there were no alterations of gaze positions between the text and pictures. The motion shown in the animation comprised a series of rapid movements for different distances (about 96.71 pixels in a move on average). When subjects followed the motion with their eyes, they did not necessarily follow every single motion foveally to catch the information because the movements were usually regular and symmetric, and therefore easy to predict. Subjects' eye movements that were recorded on the videotapes showed that subjects sometimes even made proactive saccadic eye movements to the positions the chess piece was about to reach. However, most of the time, subjects did not move their eyes as far as the chess piece moved. The reasons just mentioned might explain why the mean saccade length of Condition (ani-audi) was less than that of Condition (simp-visu).

Furthermore, the mean saccade length in the text region was affected by picture complexity as well. The data pointed out that the mean saccade length in the text was

greater when animation was shown. There were no significant differences in the mean saccade length between different levels of picture complexity when static pictures were presented. Based on the observation of subjects' eye movements recorded on the videotapes, subjects made long saccades when reading the text in order to obtain some text information as soon as possible, so that they could have time to view the animation which was running while they were reading. Since the time for running the animation was limited, subjects had to "scan" the text fast enough, which might account for the long saccades in the text region.

5) *Number of gaze changes between the text and the picture region*:
When static pictures were shown, there was no significant difference between the three levels of picture complexity ('simple', 'medium', and 'complex'). When animation was presented, subjects switched their gazes between visual text and animation considerably more often than when static pictures were displayed. These results indicate that subjects employed different viewing strategies according to the different picture types presented in the instructions. As I have mentioned earlier, when animation was shown, subjects moved their eyes between the text and pictures as quickly and as frequently as possible to integrate the text and animation information because of the "time-pressure". In contrast, when static pictures were presented, subjects moved their eyes between text and pictures less frequently possibly because they were not constrained by time.

6) *Percentage of time spent in the picture and the text regions*:
Subjects in Condition (simp-visu) spent less time in viewing pictures than did subjects in other experimental conditions, whereas subjects in Conditions (comp-visu) and (ani-visu) spent considerably more time in the picture region. With animation, it appears that the subjects' visual attention was attracted more by the motion of the chess piece, so that they spent more time viewing the animation. With 'complex' diagrams, there were more objects depicted in the diagrams than in the 'simple' diagrams, and therefore there was more information to be processed. Certainly, subjects had to spend more time viewing the diagrams. As to the percentage of time that subjects spent in the text region, the results showed that subjects who viewed diagrams with 'simple' or 'medium' complexity spent substantially more time reading the text than did subjects who viewed 'complex' diagrams or 'animation'. These are the results that accompanied those just discussed. The more time subjects spent viewing the pictures, the less time they spent reading the text accordingly.

8.5.3.3 Number of propositions

The mean section time was positively related to the number of propositions while the mean section frequency was inversely related to the number of propositions. If there is more information to be processed in a section, subjects certainly need more time to study that section which accounts for the longer section time. On the other hand, there was a negative correlation between section time and section frequency. Accordingly, the mean section frequency decreased as the number of propositions increased. However, what was the reason why the subjects repeated the sections with more propositions less frequently than those sections with fewer propositions? A possible reason is that subjects could remember the information better after they had spent more time studying it. Therefore, more repetitions were not necessary.

As to the effects of the number of propositions on subjects' eye-movement behavior, the results indicated that the mean number of fixations in the picture region increased as the number of propositions increased, whereas the mean saccade length in the picture region decreased as the number of propositions increased when text was displayed auditorily. The positive relationship between the number of fixations in the picture region and the number of propositions indicates that if there are more propositions conveyed by the text, subjects require more fixations to pick up the corresponding information in the picture, regardless of how text is presented (visually or auditorily).

The negative relationship between the mean saccade length in the picture region and the number of propositions when text was presented auditorily can be related to a study by May et al. (1990). The results of this study show that under the condition of performing a dual task, where one visual and one auditory task have to be performed at the same time, the subjects' saccadic extent decreased as cognitive workload increased (for more details see Chapter 6). Nevertheless, the same effect did not exist when text was presented visually, which may be an indicator that the visual-only presentation mode induces relatively less mental workload than does the audio-visual one within the same span of time because the verbal and pictorial information is processed sequentially. The effect of the number of propositions on the mean saccade length was not significant when animation with auditory text was presented, which points out that animation could have a much stronger influence on subjects' eye movements than auditory text.

Finally, the mean number of gaze changes between the text and the picture region decreased as the number of propositions increased. Consider the two sections

with the largest number of propositions—the 'horse' and the 'soldier'. The rules in relation to the two pieces are similar to (but not the same as) those for the 'knight' and the 'pawn' in European chess, respectively. Subjects with a little chess experience may be more or less familiar with the basic movement rules of these pieces, but they are not familiar with the Chinese chessboard and the additional rules that distinguish the 'horse' from the 'knight' as well as the 'soldier' from the 'pawn'. In order to understand the rules for the 'horse' and the 'soldier', subjects would have to rely more on the information depicted in the pictures. According to the data, subjects did spend much more time viewing pictures rather than reading the texts when they viewed those two sections. This might explain the low frequency of gaze changes between the text and the picture because subjects concentrated more on the pictorial information.

8.5.3.4 Number of objects, quantity of visual distractors, and quantity of relevant objects

The mean section time was positively related to the number of objects and the quantity of relevant objects. The explanation that I provide is the same as the one that expounds the effect of the number of propositions on the mean section time. The more objects that are depicted in a section, the more time that is required for processing the information. The quantity of visual distractors had no effect on the mean section time or the mean section frequency, which is consistent with the effects of picture complexity on the mean section time and the mean section frequency, because picture complexity is (partly) confounded by the quantity of visual distractors.

The effects of the number of objects, the quantity of visual distractors as well as the quantity of relevant objects on subjects' eye movements were almost the same. Thus they are discussed here together. The results of this experiment show that:

1) The mean number of fixations in the picture region was positively related to the number of objects irrespective of the presentation mode of the text or pictures. The reason for this result is quite simple. Basically, if there are more objects in the diagrams, more fixations are accordingly required to scan the objects, regardless of how the text and pictures were presented.

2) The mean fixation duration in the picture region was inversely related to the number of objects only when auditory text was presented, but not when visual text was presented. Similarly, the mean fixation rate in the picture region was positively

related to the number of objects only when auditory text was presented. However, the number of objects had no effect on the mean fixation rate in the picture region when animation was displayed. The same results were also obtained with the quantity of visual distractors and the quantity of relevant objects.

In my opinion, this kind of eye-movement behavior can be explained by the "time-pressure" phenomenon. Since the auditory information can only be held briefly in working memory (for about 2 seconds) while waiting to be processed, the information will be gone if it is not processed in time. When subjects try to process the auditory and pictorial information simultaneously, the "time-pressure" phenomenon can be observed in their eye movements: As the number of objects increases, subjects have to speed up their eye movements to search or to scan the objects that convey information that is relevant for establishing reference to the text. However, when visual text was presented, subjects' eye movements were insensitive to the increase in the number of objects because there was no "time-pressure" phenomenon in this case. Subjects were free to process the textual and pictorial information at their own pace.

3) When auditory text and static pictures were presented, the mean saccade length in the picture region became larger as the number of objects or the quantity of relevant objects increased. These results can be explained by the "time-pressure" phenomenon effect as well. While subjects were listening to the text, they needed to scan the objects quickly, in order to establish a referential bond between the verbal and the pictorial information. Since the objects that carry relevant information are not restricted to a small area but are distributed throughout the picture, subjects often needed relatively long saccades to find the relevant objects. The more relevant objects a picture contained, the more long saccades were possibly made. Moreover, when animation was shown, the effect of the number of objects was not significant probably because subjects' eye movements were mainly controlled by the motion of the chess pieces.

By comparing the effect of the number of propositions with the effect of the number of objects when text was presented auditorily, the mean saccade length in the picture region was found to be much more strongly affected by the number of propositions, which points out that subjects did pay attention to the auditory text, and their eye movements were more controlled by the text than by the pictures.

4) When text was presented visually, the percentage of time subjects spent in the picture region was positively related to the number of objects, the quantity of visual distractors, and the quantity of relevant objects. The more objects a picture contained, the more time that subjects spent viewing them. This could be explained along the same lines as has been done with regard to the mean section time. People attempt to exploit the available information to the full, attending to each relevant object at least once in the course of processing.

9. General discussion

9.1 Issues regarding dual-modality vs. single-modality presentations

In this thesis, I have systematically investigated how people learn or process multimedia-based learning materials containing texts and pictures. Two experiments were conducted in which subjects learned multimedia instructions with regard to how to assemble a cube puzzle or play Chinese chess. Throughout the learning process, the subjects' eye movements were measured by means of an eye tracker.

The results of Experiment 1 revealed that the effect of the picture-presentation modes (static or moving) on learning efficiency was not significant, and thus it could not confirm the overall superiority of moving pictures in demonstrating the procedures for assembling the cube puzzle. Nevertheless, with regard to the frequency of repeating instructions, moving pictures were found to exert a positive effect in that subjects who viewed moving pictures had to go over the instructions substantially less often than did subjects who viewed static pictures. Besides, the superiority of moving pictures in terms of enabling the visualization of complex movements was demonstrated by the fact that subjects who viewed static pictures had to repeat the instructional sections that comprised complicated rotations of the puzzle parts more frequently than did the subjects who viewed moving pictures. The text presentation modes (written or spoken), in contrast, had a significant effect on learning efficiency. Spoken text, together with moving pictures, resulted in the highest level of learning efficiency, whereas written text with moving pictures resulted in the lowest level. However, this positive effect of the spoken text was only observed with moving pictures. When static pictures were displayed, there was no difference in learning efficiency with respect to the text modes. Therefore, the widely-held belief in the superiority of dual modality presentation (DMP) was not confirmed in the present experiment when static pictures were presented.

In Experiment 2, the factors determining the superiority of DMP were investigated further. It was hypothesized that when static pictures are employed in the instructions, the superiority of DMP in terms of information processing would only be confirmed if the pictorial information presented in the static pictures is neither too simple nor too complicated. If pictorial information is visually highly-demanding, DMP might even be worse than SMP. The reasoning behind this hypothesis is based on the dual-task paradigm: because the capacity of the central executive is limited, the attentional resource must be divided when faced with a dual

task, which results in a trade-off between the two tasks. In contrast, there is no such trade-off if the two tasks are performed in succession. Furthermore, it was hypothesized that if the pictorial information is visually demanding, animation could be used to guide subjects' visual attention, thus reducing the load on the visuo-spatial sketchpad in working memory. The results showed that there were no significant differences in the error rate between the DMP-condition and the SMP-condition when the pictures were simple. The same held for pictures of medium complexity. However, when the pictures were complex, the error rate was significantly higher in the DMP-condition than in the SMP-condition. In contrast, the error rate in the SMP-condition was higher than in the DMP-condition when the complex pictures were animated.

Overall, the results of these two experiments provide evidence that contradicts cognitive load theory in terms of the split-attention effect and the modality effect. The superiority of DMP as a tool for learning could only be observed when animation was involved. This does not necessarily mean that DMP cannot be effective when static pictures are involved. In fact, several empirical studies (see Chapter 5) did attest to the superiority of DMP when static pictures were used. As a consequence, it is unlikely that the picture mode is able to determine whether DMP is better. The positive effect of DMP observed in those studies may have been the result of the pictorial materials employed in those experiments being neither visually demanding nor particularly simple. Although Experiment 2 failed to yield results that clearly demonstrated a positive effect of DMP with static pictures of medium complexity, a slightly lower error rate for the DMP condition may be an indication that DMP has the potential to be beneficial. It goes without saying that further experiments are required to clarify this point. All in all, the results of Experiment 2 suggest that the capacity of working memory and the way in which working memory processes information both play a crucial role in determining whether or not DMP facilitates information processing.

The two experiments I conducted have demonstrated that both the split-attention effect and the modality effect require further modification. When static graphics were displayed together with texts, SMP was not less effective than DMP, a result which blatantly contrasts with the claims of the cognitive load theory. The widely-accepted view that DMP should be superior to SMP can no longer be upheld in light of the evidence. As a consequence, the use of SMP should not be underestimated. What is to be kept in mind, though, are the limits of the human cognitive system. Due to the limited capacity of the central executive (or the episodic buffer), i.e. the component in working memory that integrates information from multiple sources, the amount of information that can be processed at the same time is

restricted. The overall attentional resource cannot simply be enlarged by presenting information in different modalities. However, it can be exhausted better if the information is well coordinated in the presentation and when the quality as well as the quantity of information to be processed does not overburden working memory.

9.2 Eye movements and information processing

In both experiments I analyzed eye movements to investigate how subjects regarded the instructions presented in accordance with different experimental conditions because the eye-movement behavior can give insight into the cognitive processes and the dynamics of attention involving learning in the given multimedia-based learning scenarios. There were some correspondences in the results of the two experiments: 1) the number of fixations was positively related to the number of propositions as well as the number of objects; 2) the fixation duration was longer when animation or auditory text was presented, but decreased as the number of objects increased; 3) the fixation rate was substantially lower when auditory text or animation was displayed, but it increased with the number of objects. More detailed eye-movement analyses were especially conducted in the second experiment to examine the eye movements in the picture regions of the instructional displays under DMP-conditions versus under SMP-conditions. For instance, the effects of the number of objects on the fixation duration, the fixation rate, and the saccade length in the picture regions were significant only when text was presented auditorily but not visually.

Basically, the eye-movement behavior suggests that cognitive strategies are fairly flexible and adaptable to different multimedia displays. In general, eye movements are more rapid when dealing with the visual-only format of instructions and are much slower when dealing with the audio-visual format. This indicates that DMP actually gives rise to a split-attention effect because the attentional resources must be divided to process the simultaneously-presented auditory and visual information, and thus the eye-movement speed is reduced. Saccade lengths are informative with regard to the connection between mental workload and eye movements. The mean saccade length in the picture region was found to decrease as the number of propositions increased, provided that the text was presented auditorily. Besides, the fixation rate was positively related to the number of objects when static pictures were presented with auditory text, but not with visual text. All these results suggest that DMP induces a heavier cognitive load than SMP.

The qualitative analysis of eye movements in the second experiment demonstrated how subjects' visual attention was allocated to process the instructions. When a subject was inspecting a complex picture while listening to the verbal instructions, there was a noticeable delay in the oculomotor reaction, whereas a subject could instantaneously react to the auditory text when the picture was less demanding. This is an indicator of a tradeoff between the visual and the auditory task. Obviously, when the central executive devoted more attentional resources to deal with the information processed by the visuo-spatial sketchpad, the attentional resources left for handling the information processed by the phonological loop became less and vice versa. Moreover, when visual text was employed, gaze shifts between text and pictures could often be observed at the phrase boundaries. This is not only in line with the immediacy principle put forward by Just and Carpenter (1980, 1987) but also indicates that the processes of integrating textual and pictorial information are incremental.

Aside from the correspondences in the results described above, there is an inconsistent result between the two experiments that I would like to mention here. It is concerned with the number of gaze switches between text and moving pictures. In the first experiment, I employed video clips combined with visual text. Subjects often switched their gazes between the text and the video region while the video was not playing. This was not the case in the second experiment. In Experiment 2, animation was used for the instructions rather than video clips. As the videotaped eye-movements show, subjects characteristically shifted their gazes between the text and the animation region while the animation was playing. The reason for this difference was probably that the animation started playing automatically when an instructional section was just loaded, whereas the video clips in Experiment 1 were played only when subjects clicked on a button to start them. As motion is likely to capture the beholder's visual attention automatically, subjects who received animation were encouraged to watch it right at the beginning. However, animation alone was not easy to understand without reading the text. Hence, speedy eye movements between the text and the animation region were observed. What is peculiar to me is that subjects did not replay the animation many times, even though they knew they were allowed to replay it as often as they would like. In contrast, subjects who received video clips (which did not start automatically) tended to read the text first, and then watch the videos. After the first or the second visit to the same instructional section, subjects were inclined to concentrate on viewing the videos. Despite the different viewing strategies employed by those subjects, their performance in both experiments was the worst compared with that of the subjects in other experimental conditions. While DMP has proved to be no better than SMP in

many situations, animation (or moving pictures) presented with visual text is the very form of SMP that remains less efficient than DMP because information presented by means of both animation and visual text is difficult to integrate.

9.3 Applications of this research

The current research is a kind of basic research from which several recommendations that might apply to instructional design in general arise. First, instructional designers should consider their choices. Moving pictures are particularly efficient in visualizing complex processes or actions. Animation, on the other hand, can be used to guide the recipients' visual attention or to emphasize alterations of states within a dynamic system. However, both moving pictures and animation will only facilitate information processing when the accompanying verbal instructions are presented auditorily. Visual texts should be avoided in this case because of the danger of structural interference. Second, instructional designers should be careful in using multimodal presentations, bearing in mind that the capacity of recipients' working memory is limited. While multimodal presentations are attractive to many people, poorly-designed multimodal presentations might either impose a heavier cognitive load on the recipients or unnecessarily distract their attention from relevant content. Third, instructional designers should take learners' characteristics into account. There is no way that a multimedia presentation can be designed in such a way that it is suitable for everyone. Learners with high domain-specific knowledge or aptitude should be addressed differently from learners with low domain-specific knowledge or aptitude. Age and learning experience are also factors that need to be considered. Finally, instructional designers should adopt appropriate measures to evaluate the effectiveness of multimedia presentations. Only a suitable evaluation can guarantee with a reasonable probability that a particular multimedia presentation will have the desired effect on learning.

References

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research,* Vol. 38, 403-414.

Ahern, S. K. (1978). *Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilities.* Unpublished doctoral dissertation, University of California, Los Angeles.

Alba, J., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin,* 93: 203-231.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, Mass.: Harvard University Press.

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

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning: An introduction to school learning.* New York: Grune & Stratton.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston.

Ausubel, D. P. (1977). The facilitation of meaningful verbal learning in the classroom. *Educational Psychologist,* 12, 162-178.

Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology,* Vol. 18, 362-365.

Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology,* Vol. 18, 302-309.

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

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

Baddeley, A. D. (1995). Working memory. In M. S. Gazzaniga et al. (Eds.), *The Cognitive Neurosciences.* Cambridge: The MIT Press.

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

<u>References</u>

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

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences,* Vol. 4, No. 11, 417-423.

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

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working Memory and the Control of Action: Evidence From Task Switching. *Journal of Experimental Psychology: General,* Vol. 130, No. 4, 641-657.

Baddeley, A. D., Baddeley, H., Bucks, R., & Wilcock, G. K. (2001). Attentional control in Alzheimer's disease. *Brain,* Vol. 124, 1492-1508.

Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease: A longitudinal study. *Brain,* Vol. 104, 2521-2542.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device, *Psychological Review,* Vol. 105, 158-173.

Baddeley, A. D., Grant, S., Wright, E., & Thomson, N. (1975). Imagery and visual working memory. In P.M.A. Rabbitt and S. Dornic (Eds.), *Attention and Performance V* (pp. 205-217). London: Academic Press.

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

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 521-539). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General,* Vol. 103, 518-540.

Baddeley, A. D., Logie, R., Bressi, S., Della Sala, S., & Spinnler, H. (1986). Dementia and working memory. *Quarterly of Journal of Experimental Psychology,* Vol. 38A. 603-618.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior,* Vol. 9, 176-189.

Bannert, M. (2002). Managing cognitive load—recent trends in cognitive load theory. *Learning and Instruction,* Vol. 12, 139-146.

<u>References</u>

Bartlett, F. C. (1932). *Remembering.* Cambridge: Cambridge University Press

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, Vol. 91, No. 2, 276-292.

Bennett, L. T., & Dwyer, F. D. (1994). The effect of varied visual interactive strategies in facilitating student achievement of different education objectives. *International Journal of Instructional Media,* Vol. 21, No. 1, 23-32.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology,* Vol.103, 597-600.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology,* Vol. 49A, 525-545.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology,* Vol. 11, 177-220.

Boynton, R. M. (1960). Summary and discussion. In A. Morris, & E. P. Horne (Eds.), *Visual search.* Washington, D.C.: National Academy of Science.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior,* Vol. 8, 323-343.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology,* Vol. 10, 12-21.

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature,* Vol. 232, 456-458.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125-148). Oxford, England: Elsevier.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review,* Vol. 14, No. 1, 5-26.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition andIinstruction*, Vol. 8, No. 4, 293-332.

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

References

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology,* Vol. 10., 151-170.

ChanLin, L. J. (1998). Animation to teach students of different knowledge levels. *Journal of Instructional Psychology,* Vol. 25, No. 3, 166-175.

ChanLin, L. J. (2000). Attributes of Animation for learning scientific knowledge. *Journal of Instructional Psychology,* Vol. 27, No. 4, 228-238.

Christianson, S. A., Loftus, E. F., Hoffman, H., & Loftus, G. R. (1991). Eye fixations and memory for emotional events. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 17, 693-701.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, Vol. 3, No. 3, 149-210.

Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, and W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 385-407). Hillsdale, N. J.: Erlbaum.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review,* Vol. 82, 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior,* Vol. 8, 240-247.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley.

Cooper, G. A. (1998) Research into cognitive load theory and instructional design at UNSW. http://www.arts.unsw.edu.au/education/CLT_NET_Aug_97.HTML.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin,* Vol. 104, 163-191.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences,* Vol. 24, 87-185.

Craik, K. (1943). *The nature of explanation.* Cambridge: Cambridge University Press.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General,* Vol. 125, 159-180.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* Vol. 19, 450-466.

References

Desse, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports,* Vol. 5, 305-312.

Douglas, L. M., & Brian, H. R. (1992). *Cognitive Psychology*. Fort Worth etc.: Harcourt Brace Jovanovich College Publishers.

Dretzke, B. J. (1993). Effects of pictorial mnemonic strategy usage on prose recall of young, middle-aged, and older adults. *Educational Gerontology,* 19, 489-502.

Duffy, C. J. (1997). Implicit memory. Knowledge without awareness. *Neurology,* Vol. 49, 1200-1202.

Duffy, S. A. (1992). Eye movements and complex comprehension processes. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 462-471). New York: Springer-Verlag.

d'Ydewalle, G., Van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J.K. O'Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 313-321). Amsterdam: Elsevier Science Publishers B. V. (North-Holland).

Eichenbaum, H. (1997). Declarative memory: Insights from cognitive neurobiology. *Annual Review of Psychology.* Vol. 48, 547-572.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review,* Vol. 102, 211-245.

Fang, Z. (1996). Illustrations, text, and the child reader. What are pictures in children's storybooks for? *Read. Horizons* 37: 130-142.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior,* Vol. 5, 351-360.

Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. V. New York: Academic Press.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of Memory and Language,* Vol. 31, 129-151.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Vol. 11, 501-518.

<u>References</u>

Graf, P. (1994). Explicit and implicit memory: A decade of research. In C. Umiltà, & M. Moscovitch (Eds.), *Attention and Performance XV: Conscious and nonconscious information processing* (pp. 681-696). Cambridge, MA: The MIT Press.

Gould, J. D. (1983). Looking at pictures. In R. A. Monty, & J. W. Sender (Eds.), *Eye Movements and Psychological Processes* (pp. 323-369). New York: Lawerence Erlbaum Associates, Publishers.

Hays, T. A. (1996). Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *Journal of educational computing research,* Vol. 14, No. 2, 139-155.

Hegarty, M. (1992). 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 etc.: Springer-Verlag.

Hegarty, M.,& Just, M. A. (1989). Understanding machines from text and diagrams. In H. Mandl, & J. Levin (Eds.), *Knowledge acquisition from text and picture* (pp. 171-194). Amsterdam: North Holland.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory and Cognition,* Vol. 22, 411-430.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American,* Vol. 212, No.4, 46-54.

Hess, E. H. (1972). Pupillometrics: A method of studying mental, emotional, and sensory processes. In: N. S. Greenfield, & R. A. Sternbach (Eds.), *Handbook of Psychophysiology* (pp. 491-531). New York etc.: Holt, Rinehart and Winston, INC..

Jenkins, J. J., & Russel, W. A. (1952). Associative clustering during recall. *Journal of Abnormal and Social Psychology,* Vol. 47, 818-821.

Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology,* Vol. 17, No. 3, 329-343.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1988). *The computer and the mind.* London: Fontana.

Johnson-Laird, P. N. (1995). Mental models, deductive reasoning, and the brain. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences.* (pp. 999-1008). Cambridge: The MIT Press.

Jonassen, D. H. (1995). Operationalizing mental models: Strategies for assessing mental models to support meaningful learning and design supportive learning environments. *http://hagar.up.ac.za/catts/learner/magda/model.html*

References

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review,* Vol. 87, 329-354.

Just, M. A., Carpenter, P. A., & Wu, R. (1983). *Eye fixations in the reading of Chinese technical text* (Technical Report). Pittsburgh: Carnegie-Mellon University.

Just, M. A., & Carpenter, P. A. (1987). *The Psychology of Reading and Language Comprehension.* Boston: Allyn and Bacon, Inc.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology,* Vol. 47, No. 2, 310-339.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science,* 154, 1583-1585.

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australian Journal of Educational Technology*, Vol.16, No. 2, 161-172.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors,* Vol. 40, No. 1, 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing Split-attention and Redundancy in Multimedia Instruction. *APPLIED COGNITIVE PSYCHOLOGY*, Vol. 13: 351-371.

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

Kennedy, A. (1983). On looking into space. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 237-251). New York: Academic Press.

Kennedy, A. (1992). The spatial coding hypothesis. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 379-396). New York: Springer-Verlag.

Kennedy, A., & Murray, W. S. (1987). Spatial coordinates and reading: Comments on Monk. *Quarterly Journal of Experimental Psychology,* 39A, 649-656.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review,* Vol. 85, No. 5, 363-394.

<u>References</u>

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review,* Vol. 95, 163-182.

Kintsch, W. (1994). The psychology of discourse processing. In M. A. Gernsbach (Ed.), *Handbook of Psycholinguistics* (pp. 721-739). San Diego: Academic Press.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction,* Vol. 12, 1-10.

Koesling, H., Clermont, T., & Ritter, H. (2001). CRC 360 "Situated Artificial Communicators". University of Bielefeld, Germany.

Kosslyn, S. M. (1973). Scanning visual images: some structural implications. *Perception and Psychophysics.* Vol. 14, 90-94.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and brain. The resolution of the imagery debate.* Cambridge, MA: The MIT Press.

Kristjanson, A. F., & Antes, J. R. (1989). Eye movement analysis of artists and nonartists viewing paintings. *Visual Arts Research,* Vol. 15, 21-30.

Kunz, G. C., Drewniak, U., & Schott, F. (1989). On-line and off-line assessment of self-regulation in learning from instructional text and picture. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Large, A. (1996). Computer animation in an instructional environment. *Library & Information Science Research,* Vol. 18, No. 1, 3-23.

Large, A., Beheshti, J., Breuleux, A., & Renaud, A. (1996). The effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of the American Society for Information Science,* Vol. 47, No. 6, 437-448.

Levin, J. R. (1981). On functions of pictures in prose. In F. J. Pirozzolo, & M. C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading* (pp. 203-228). New York: Academic Press.

Levin, J. R. (1989). A transfer-appropriate-processing perspective of pictures in prose. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 83-100). Amsterdam: Elsevier.

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

Levin, J. R., & Mayer, R. E. (1993). Understanding illustrations in text. In B. K. Britton, A. Woodward, & M. Brinkley (Eds.), *Learning from textbooks* (pp. 95-113). Hillsdale, NJ: Erlbaum.

References

Lesgold, A. M., Roth, S. F., & Curtis, M. E. (1979). Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and VerbalBehavior,* Vol. 18, 291-308.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in cognitive sciences,* Vol. 4, No. 1, 6-14.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception & Performance,* 4, 565-572.

Loftus, G. R. (1983). Eye fixations on text and scenes. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 359-376). New York: Academic Press.

Logie, R.H. (1986). Visuo-spatial processes in working memory. *Quarterly Journal of Experimental Psychology*, Vol. 38A, 229-247.

Logie, R. H. (1996). The Seven Ages of Working Memory. In J. T. E. Richardson et al. (Eds.), *Working Memory and Human Cognition* (pp. 31-65). New York, Oxford: Oxford University Press.

Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). *Division of attention in Alzheimer's disease.* Paper presented at the Psychonomics Society meeting, Los Angeles, CA.

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models.* Cambridge MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, & J. L. McClelland (Eds.), *Parallel distributed processing.* Cambridge, MA: MIT Press.

Macworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic science* 3, 67-68.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics,* 2, 547-552.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology,* Vol. 88, No. 1, 49-63.

Marschark, M., & Paivio, A. (1977). Integrative processing of concrete and abstract sentences. *Journal of Verbal Learning and Verbal Behavior,* Vol. 16, 217-231.

Marshall, S. P. (1995). *Schemas in problem solving.* Cambridge: Cambridge University Press.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition,* Vol. 10, 181-187.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychologica* 75, 75-89.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology,* Vol. 81, No. 2, 240-246.

Mayer, R. E. (1993). Instructions that instruct. In R. Glaser (Ed.), *Advances in Instructional Psychology.* Vol. 4, pp. 253-284.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist,* Vol. 32, No. 1, 1-19.

Mayer, R. E., & Anderson, R. B. (1991). Animation need narrations: an experimental test of a dual-coding hypothesis. *Journal of Educational Psychology,* Vol. 83, No. 4, 484-490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology,* Vol. 84, No. 4, 444-452.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology,* Vol. 82, 715-726.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology,* Vol. 93, 187-198.

Mayer, R. E., & Moreno, R. (1998). A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing System in Working Memory. *Journal of Educational Psychology,* Vol. 90, No. 2, 312-320.

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

Mayer, R. E., & Moreno, R. (2002b). Animation as an aid to multimedia learning. *Educational Psychological Review.* Vol. 14, No. 1, 87-99.

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

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology,* Vol. 86, No. 3, 389-401.

<u>References</u>

Meili, R. (1955). Würfelabwicklungen. Bern: Hans Huber.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior,* Vol. 2, 1-21.

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

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt.

Milner, B. (1966). Amnesia following operation on the temporal lobes. In C. W. M. Whitty, & O. L. Zangwill (Eds.), *Amnesia* (pp. 109-133). London: Butterworths.

Minsky, M. L. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw Hill.

Miyake, A. & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues and future directions. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge, England: Cambridge University Press.

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

Moreno, R., & Mayer. R. E. (2000a). A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory. *Interactive multimedia electronic journal of computer-enhanced learning.* Wake Forest University. http://imej.wfu.edu/articles/2000/2/05/printver.asp

Moreno, R., & Mayer, R. E. (2000b). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology,* Vol. 93, 724-733.

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

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology*, Vol. 87, No. 2, 319-334.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception and Psychophysics,* Vol. 13, 180-184.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature,* Vol. 215, 1519-1520.

References

Najjar, L. J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia,* Vol. 5, 129-150.

Narayanan, N. H., & Schrimpsher, D. J. (2000). Extending Eye Tracking to Analyse Interactions with Multimedia Information Presentations. In People and Computers XIV- Usability or Else! Proceedings of HCI 2000 (pp. 271- 286). London etc.: Springer-Verlag.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: an approach to combine mental-effort and performance measures. *Human Factors,* Vol. 35, 737-743.

Paivio, A. (1967). *Meaning, mediation, and memory.* Invited address at the Annual Meeting of the Canadian Psychological Association, Ottawa, Ontario.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, Vol. 76, 241-263.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston. (Reprinted 1979, Hillsdale, NJ: Lawrence Erlbaum Associates.)

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory and Cognition,* Vol. 3, 635-647.

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

Paivio, A. (1991). Dual coding theory: retrospect and current status. *Canadian Journal of Psychology,* Vol. 45, No. 3, 255-287.

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

Peavler, W. S. (1974). Individual differences in pupil size and performance. In M. Janisse (Ed.), *Pupillary dynamics and behavior.* New York: Plenum Press.

Peeck, J. (1987). The role of illustrations in processing and remembering illustrated text. In D. M. Willows, & H. A. Houghton (Eds.), *The psychology of illustration: Vol. 1. Basic research* (pp. 115-151). New York: Springer.

Peeck, J. (1993). Increasing picture effects in learning from illustrated text. *Learning and Instruction,* Vol. 3, 227-238.

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language,* Vol. 24, 503-518.

References

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology,* Vol. 58, 193-198.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Pomplun, M. (1998). *Analysis and Models of Eye Movements in Comparative Visual Search.* Göttingen: Cuvillier Verlag.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction,* Vol. 12, 61-86.

Posner, M. I., & Raichle, M. E. (1999). *Images of Mind.* New York: Scientific American Library.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin,* Vol. 80, 1-24.

Pylyshyn, Z. W. (1979). The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory and Cognition,* Vol. 7, 19-28.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review,* Vol. 86, 16-45.

Pylyshyn, Z. W. (2001). Is the imagery debate over? If so what was it about? In E. Dupoux (Ed.), *Cognition: a critical look. Advances, questions and controversies in honor of J. Mehler.* Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (2002). Mental imagery: in search of a theory. *Behavioral and Brain Sciences,* Vol. 25, No. 2. xxx-xxx.

Rantanen, E. M., & Goldberg, J. H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics,* vol. 42, No. 6, 816-834.

Rayner, K. et al. (1980). Asymmetry of the effective visual field in reading. *Percept. Psychophys.* 27, 537-544.

Rayner, K. et al. (1982). The availability of useful information to the right of fixation in reading. *Percept. Psychophys.* 31, 537-550.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin,* Vol. 124, No. 3, 372-422.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology*: *Human Perception and Performance,* 22, 1188-1200.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review,* 3, 504-509.

References

Richardson, J. T. E., Engle, R. W., Hasher, L., Logie, R. H., Stoltzfus, E. R., & Zacks, R. T. (1996). *Working Memory and Human Cognition.* New York, Oxford: Oxford University Press.

Rickheit, G., & Sichelschmidt, L. (1999). Mental models: some answers, some questions, some suggestions. In G. Rickheit, & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 9-40). Amsterdam: Elsevier Science B.V.

Rickheit, G., & Strohner, H. (1993). *Grundlagen der kognitiven Sprachverarbeitung: Modelle, Methoden, Ergebnisse.* Tübingen; Basel: Francke.

Rieber, L. P., & Kini, A. S. (1991). Theoretical foundations of instructional applications computer-generated animated visuals. *Journal of Computer-Based Instruction,* Vol. 18, 83-88.

Rieber, L. P., & Kini, A. (1995). Using computer simulations in inductive learning strategies with children in science. *International Journal of Instructional Media,* Vol. 22, No. 2, 135-143.

Roshal, S. M. (1961). Film mediated learning with varying representations of the task: viewing angle, portrayal of demonstration, motion and student participation. In A. A. Lumsdaine (Ed.), *Student response in programmed instruction* (pp. 155-175). National Research Council.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow, & A. Collins, (Eds.), *Representation and understanding* (pp. 211-236). New York: Academic Press.

Rusted, J. M. (1984). Differential facilitation by pictures of children's retention of written texts: A review. *Current psychological research and reviews,* Vol. 3, 61-71.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Exploration of comprehension beyond the sentence*. New York: Wiley.

Sanford. A. J., & Garrod, S. C. (1982). Towards a processing account of reference. In A. Flammer, & W. Kintsch (Eds.), *Discourse processing* (pp. 100-110). Amsterdam: North-Holland.

Sanford, A. J., Garrod, S. C., & Henderson, R. (1980). Topic shift as a variable in text cohesion: Experimental evidence from studies in reading time. Paper presented at the Meeting of the Experimental Psychological Society, Cambridge.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Vol. 13, 510-518.

Schnotz, W. (1985). Selectivity in drawing inferences. In G. Rickheit, & H. Strohner (Eds.), *Inferences in text processing* (pp. 287-326). Amsterdam: North-Holland.

<u>References</u>

Schnotz, W. (1988). Textverstehen als Aufbau mentaler Modelle. In H. Mandl, & H. Spada (Eds.), *Wissenspsychologie* (pp. 299-330).

Schnotz, W., & Bannert, M. (1999). Strukturaufbau und Strukturinterferenz bei der multimedial angeleiteten Konstruktion mentaler Modelle. In I. Wachsmuth, & B. Jung (Eds.), *KogWiss 99. Proceedings der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft,* Bielefeld 28. September - 1. Oktober, 1999 (pp. 79-85). Sankt Augustin: Infix.

Schnotz, W., & Preuß, A. (1999). Task-dependent construction of mental models as a basis for conceptual change. In G. Rickheit, & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 131-167). Amsterdam: Elsevier Science B. V.

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review,* Vol. 14, No. 1, 101-120.

Schank R. C., & Abelson, R. (1977). *Script, plans, goals and understanding.* Hillsdale, NJ: Lawrence, Erlbaum Associates Inc.

Schunk, D. H. (1996). *Learning theories: an educational perspective.* NJ: Prentice-Hall, Inc.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology,* Vol. 91, No. 4, 690-702.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology,* Vol. 22: 261-273.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science,* Vol. 171, 701-703.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, Mass.: Bradford.

Silverman, R. E. (1958). The comparative effectiveness of animated and static transparences. (Report No. NTDC-TR 78-1). Port Washington, NY: US Naval Training Device Center.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review,* Vol. 81, 214-241.

Spangenberg, R. W. (1973). The motion variable in procedural learning. *Audio-Visual Communication Review,* Vol. 21, 419-436.

References

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied,* Vol. 74, 1-29.
Sperling, G. (1963). A model for visual memory tasks. *Human factors,* Vol. 5, 19-31.

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, Vol. 18, 275-290.

Squire, L. R. (1987). *Memory and Brain.* New York: Oxford University Press.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of National Academic Science USA.* Vol. 93, 13515-13522.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of human intelligence.* Cambridge: Cambridge University Press.

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

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology,* 81, 457-466.

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

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

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction,* Vol. 12, No. 3, 185-233.

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

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychological Review,* Vol. 10, No. 3, 251-296.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science,* Vol. 268, 1632-1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using Eye Movements to Study Spoken Language Comprehension: Evidence for Visually Mediated Incremental Interpretation. In T. Inui, & J. L. McClelland, (Eds.).

References

Attention and Performance XVI: Information Integration in Perception and Communication. Cambridge: The MIT Press.

Thorndyke, P. W., & Yekovich, F. R. (1980). A critique of schema-based theories of human story memory. *Poetics,* 9: 23-49.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When Two Sensory Modes Are Better Than One. *Journal of Experimental Psychology: Applied.* Vol. 3, No. 4, 257-287.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Reasoning: Representation and process in children and adults.* Hillsdale, N. J.: Erlbaum.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review,* Vol. 69, 344-354.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.

Tulving, E. (1993). What is episodic memory. *Current Directions in Psychological Sciences,* Vol. 2, 67-70.

Tulving, E. (1995). Organization of Memory: Quo Vadis? In M. S. Gazzaniga et al. (Eds.), *The Cognitive Neurosciences.* Cambridge: The MIT Press.

Tulving, E. (2001). Episodic memory and common sense: How far apart? *PHILOSOPHICAL TRANSACTIONS- ROYAL SOCIETY OF LONDON SERIES B*, Vol. 356, No. 1413, 1505-1515.

Tversky, B. (1974). Eye fixations in prediction of recognition and recall. *Memory and Cognition.* Vol. 2, No. 2, 275-278.

Valcke, M. (2002). Cognitive load: updating the theory? *Learning and Instruction,* Vol. 12, 147-154.

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition, 3,* 199-222.

Waddill, P. J., McDaniel, M. A., & Einstein, G. O. (1988). Illustrations as adjuncts to prose: a text-appropriate processing approach. *Journal of Educational Psychology,* Vol. 80, No. 4, 457-464.

Waddill, P. J., & McDaniel, M. A. (1992). Pictorial enhancement of text memory: limitations imposed by picture type and comprehension skill. *Memory & Cognition,* Vol. 20, No. 5, 472-482.

References

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review,* Vol. 72, 89-104.

Weidenmann, B. (1988). *Psychische Prozesse beim Verstehen von Bildern.* Bern: Huber.

Weidenmann, B. (1997). Multicodierung und Multimodalität im Lernprozeß. In L. J. Issing, & P. Klimsa (Eds.), *Information und lernen mit multimedia* (pp. 64-84). Weinheim: Psychologie Verlags Union.

Wilson, R. A., & Keil, F. C. (1999). *The MIT encyclopedia of the cognitive sciences.* Cambridge: The MIT Press.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.