Sequential display of diagrams: how does it affect exploration and memorization?

Rolf Wipfli

Mémoire présenté pour l’obtention du Master MALTT
Master of Science in Learning and Teaching Technologies

TECFA,
Faculté de Psychologie et de Sciences de l’Éducation
Université de Genève

Juin 2008

Jury
Prof. Dr. Mireille Bétrancourt professeur, TECFA Directeur
Dr. Daniel Schneider MER, TECFA Examinateur
Cyril Rebetez doctorant, TECFA Examinateur
Mirweis Sangin doctorant, ÉPFL - CRAFT Examinateur
## Contents

1 Introduction

2 Theories of multimedia learning
   2.1 Integrative models
      2.1.1 Dual coding theory of mind
      2.1.2 A cognitive process model of multimodal information integration and comprehension
      2.1.3 Cognitive theory of multimedia learning
   2.2 Instructional principles
      2.2.1 The multimedia principle
      2.2.2 Split-attention effect
      2.2.3 Segmenting principles and sequential display

3 Eye tracking and visual exploration
   3.1 Characteristics of eye movements
   3.2 Eye tracking devices
   3.3 Analysis and visualization of eye movements
   3.4 Controversies on eye tracking
   3.5 Eye tracking in multimedia learning

4 Experiment and hypotheses
   4.1 Visual exploration
   4.2 Learning performance

5 Method
   5.1 Participants
   5.2 Design
   5.3 Instrument
   5.4 Material
   5.5 Procedure
   5.6 Independent variables
   5.7 Dependent variables
      5.7.1 Spatial questions
      5.7.2 Temporal questions
      5.7.3 Functional questions
      5.7.4 Eye movement data
## Contents

### 6 Results
- 6.1 Data processing ........................................ 45
- 6.2 Eye movements ........................................... 45
- 6.3 Learning time ............................................. 47
- 6.4 Learning performance .................................... 47
- 6.5 Recall ..................................................... 48
  - 6.5.1 Recall performance .................................... 48
  - 6.5.2 Recall order ............................................ 50
  - 6.5.3 Clustering in recall task .............................. 53
- 6.6 Recognition ............................................... 53
- 6.7 Subjective evaluation of the material .................. 56

### 7 Discussion
- 7.1 Visual exploration ........................................ 57
- 7.2 Learning performance ..................................... 58
- 7.3 Recall performance and recall order .................... 59
- 7.4 Clustering in recall task .................................. 60
- 7.5 Recognition ............................................... 60
- 7.6 Subjective evaluation of the material .................. 61

### 8 Conclusion ................................................. 63

### Bibliography .................................................. 65

### A Testing material ............................................ 71
- A.1 Pretest on prior knowledge ............................... 71
- A.2 Pre- and posttest questions .............................. 71
  - A.2.1 Function questions .................................... 71
  - A.2.2 Troubleshooting questions ............................ 72

### B Program code ............................................... 75
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Experimental design</td>
<td>39</td>
</tr>
<tr>
<td>6.1</td>
<td>Mean numbers and standard deviation of referential and representational connections, Mean Levenshtein distance and standard deviation</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>for inspection of textual and pictorial elements.</td>
<td></td>
</tr>
<tr>
<td>6.2</td>
<td>Mean gain and standard deviation in score between the pre- and post-test questionnaires</td>
<td>48</td>
</tr>
<tr>
<td>6.3</td>
<td>Mean number and standard deviation of recalled elements in immediate and deferred recall task</td>
<td>50</td>
</tr>
<tr>
<td>6.4</td>
<td>Mean Levenshtein distance and standard deviation between the effective order of recall and the optimal order of recall</td>
<td>51</td>
</tr>
<tr>
<td>6.5</td>
<td>Mean number and standard deviation of pairings of corresponding textual and pictorial elements in the recognition task</td>
<td>56</td>
</tr>
<tr>
<td>6.6</td>
<td>Mean score and standard deviation on a six-point Likert-scale (1–6) in the subjective evaluation of the material for its interestingness, its learnability and its enjoyability</td>
<td>56</td>
</tr>
</tbody>
</table>
List of Tables
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The dual-coding model (Mayer &amp; Anderson, 1991)</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Model of multimedia learning (Mayer, 2001)</td>
<td>19</td>
</tr>
<tr>
<td>3.1</td>
<td>An example of a scan path visualization of eye tracking data. From Lund University (in public domain)</td>
<td>32</td>
</tr>
<tr>
<td>3.2</td>
<td>A 3-pulley system, copied from Hegarty &amp; Just, 1993</td>
<td>35</td>
</tr>
<tr>
<td>5.1</td>
<td>Learning material showing the static version</td>
<td>41</td>
</tr>
<tr>
<td>6.1</td>
<td>Mean gain in questions between pre- and post-test (Func: Function retention, Proc: Process retention, Ret: Retention (function &amp; process), TS: Troubleshooting, Tot: Total, St: Static condition, Sq: Sequential condition)</td>
<td>49</td>
</tr>
<tr>
<td>6.2</td>
<td>Mean Levenshtein distance and confidence interval (95%) for different types of recalled elements (Imm: Immediate recall, Def: Deferred recall, Proc: Process element, Obj: Object element, Tot: Total (processes and objects), St: static condition, Sq: sequential condition)</td>
<td>52</td>
</tr>
<tr>
<td>6.3</td>
<td>Hierarchical clustering of recall order in the static condition, AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) values</td>
<td>54</td>
</tr>
<tr>
<td>6.4</td>
<td>Hierarchical clustering of recall order in the sequential condition, AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) values</td>
<td>55</td>
</tr>
</tbody>
</table>
Abstract

Computers and multimedia presentations have found their way into classrooms and conference rooms. Research in multimedia learning offers models and guidelines on the effectiveness of the different presentation forms. We aim to provide some empirical evidence of how the sequential display of a material can be beneficial in multimedia presentations, as they offer a means to convey complex graphics by reducing their complexity and by offering a coherent order of the learning material. Thirty-two participants have learned either a static presentation or a sequential presentation on how a e-voting system works. The participants were tested before and after they learned the material in an immediate test and they were tested again one week later to evaluate the persistence of the learned information over time. The results of this experiment provide evidence that a sequential display of information fosters the retention of information measured in a questionnaire, as well as the coreferenciation of pictorial and textual information in a recognition task. The sequential presentation had also a significant effect on the performance and the order of recall. Additionally, eye tracking measures was conducted to study the inspection order of the participants. It was found that participants make more referential connections between the corresponding textual and pictorial elements in the sequential display.

Keywords: Multimedia learning, sequential display, eye tracking
Résumé

Les ordinateurs et les présentations multimédias ont trouvé leur place dans les salles de classe et de conférence. La recherche dans le domaine de l’apprentissage multimédia a développé plusieurs modèles et des consignes quant à l’efficacité de différentes formes de présentation. Notre objectif est de fournir des preuves empiriques pour les avantages de la présentation séquentielle d’un document à même de réduire la complexité initiale d’un matériel d’apprentissage et offrir un ordre cohérent dans l’apprentissage. Trente-deux participants ont appris, soit une présentation statique, soit une présentation séquentielle d’un document sur le fonctionnement d’un système de e-voting. Les participants ont passé des tests avant et après avoir appris la matière lors d’un essai immédiat et ils ont passé des tests une semaine plus tard pour évaluer la persistance des informations apprises. Les résultats de cette expérience confirment les hypothèses que la présentation séquentielle d’un document favorise la rétention de l’information mesurée par un questionnaire, ainsi que le coréférenciation des éléments imagés et textuels dans une tâche de reconnaissance. La présentation séquentielle montre également un effet important sur la performance et l’ordre de rappel. En outre, les mouvements oculaires des participants ont été enrégistrés pour évaluer l’ordre des inspections. On pouvait constater que les participants dans la condition séquentielle faisaient plus de connexions référentielles entre les textes et les images correspondants.

Mots de clé : apprentissage multimédia, présentation séquentielle, oculométrie
1 Introduction

It has been more than 30’000 years ago since humans started using paintings, 6000 years since writing, 450 years since printing, 100 years since producing animated pictures and it is only 30 years ago we have first introduced interactive computers with graphical interfaces. The purpose of all these inventions was to store and transmit information. Today, all of these different medias are used in different ways and combinations, may it be to broadcast the news, to express art, or to entertain people.

Education has also made use of these different medias and in recent time we adopted the new possibilities which are offered by information technology. Multimedia presentations are now commonly used in class rooms and during conferences. The conventional wisdom assumes that multimedia presentations are superior to simple textbooks as they are more motivating and easier to learn. But in how far is that true? Are animations superior to graphics? Are images sufficient to convey a learning material, is it better to have spoken text or written text?

All these questions have been explored by research in multimedia learning in the last two decades, in which the effect of the numerous forms of presentations on the learning performance was evaluated. But even if researchers were in general optimistic about the improvement of learning by multimedia and interactive presentations, they soon had to assert that learning with these new technologies does not always provide the expected results. Studies have shown that even if we can integrate more information in a presentation (i.e. conveying movements by using animations), it does not necessarily mean that the learner can adequately process the information on a cognitive level.

Cognitive scientists had to state that “the developments in media and authoring tools have, unfortunately, far outpaced concomitant developments in the theory how to design such documents for optimal comprehension.” (Narayanan & Hegarty, 1998). In order to ameliorate learning with multimedia presentations, several cognitive models were developed which aim to explain how we process the different types of presentations and several guidelines were written on how a multimedia learning material should be designed to be adapted to the learners cognitive capacity.

One central question in the research on multimedia learning is how complex graphics should be conveyed, particularly when they include changes over time. Temporal or causal relations can be represented by arrows in a diagram, but such a diagram can get easily overcharged with information. A natural way of presenting changes over time are animations, but the results of their effectiveness are mixed (Tversky, Morrison, & Bétrancourt, 2002). Animations have the risk, that the speed of the animation is unadapted to the learners capacities, which can have a detrimental effect on the learning success.
An intermediate way of presenting a complex system, somewhere in between a static diagram and an animation, was proposed with sequential displays. In sequential displays a diagram is progressively constructed, adding additional information at each sequence. This way of presenting information has been used for a long time in class rooms and without the aid of computers, when teachers developed a schema on a blackboard, explaining the new elements while drawing them. Research on sequential display has already provided some support for the effectiveness of sequential displays, but the number of studies in this field are still relatively limited.

In addition to provide further empirical evidence for the effects already identified in sequential displays, this paper would like to analyse the on-line processing of sequential diagrams with eye tracking measures. In particular we would like to evaluate the effect of the presentation order on the order of visual exploration of the material, which has in turn a possible effect on the mental organization of the learned material.

We will construct a theoretical basement by introducing some of the multimedia learning principles on the one hand and giving some methodological background on eye tracking techniques on the other hand.
2 Theories of multimedia learning

2.1 Integrative models

In science we create models which aim to explain the reality in which we live. Models help us to understand complex phenomena as they give us a simplified way to explain and predict such phenomena. Quite often there are several models which describe the same question from different angles. This is not different in the research on multimedia learning, where there are several models which help us to understand how text and images are treated on a cognitive level and what outcome (learning outcome in this case) we have to expect from different types of learning material.

2.1.1 Dual coding theory of mind

In what form do we store information in memory? How is the information represented? Not only psychologists and neuroscientists did research on this topic, also philosophers sought to find an answer to this question. Different theories have been developed in this research field. One of them assumes that we store our memories in form of pictures. Two thousand years ago, Aristotle argued already that “the soul never thinks without a mental picture” ([Paivio] 2007). We know from our own experience that we can imagine pictures and scenes, we have previously seen. But then, are images sufficient to store our experiences and can we reason with images alone? Another theoretical approach argues that our memories are largely language based. But is it sufficient to store the word “dog” in memory, in order to have a complete representation of what a dog is? We could probably assume then, that we need both text and images to think. Yet another theory assumes that we store information in some kind of abstract “mentalese”, an idea which was introduced by Plato when he described the nature of knowledge as a universe of ideas. We automatically access these ideas when we are confronted with a situation, which implicates this idea. When we see a circle we would access the idea of a prototypical circle, an idea which was already present in our mind.

Allan Paivio (2007) explored the question of mental representations during decades and he concluded, based on neurophysiological and psychological findings, that we have two different and independent systems of processing and representing information in our memory: a verbal system and a non-verbal system. The verbal system uses language as a form of representation. When we read or hear a word, we can access to a large collection of related information to this word. The word “dog” for example could activate other words like “animal”, “Labrador” or even “cat”. These associations are again verbal representations. The other mental system is the non-

13
verbal system, where we store information in pictorial form. When we imagine that
we stand in the doorway of our house, we can easily imagine the view we would have
of the entrance room. We could even walk mentally through the different rooms and
“seeing” different objects in our house. The new images activated by the first image
are again non-verbal, visual representations.

When mental representations are activated by external stimuli (verbal and non-
verbal), Paivio speaks of representational connections. Representational connections
include visual recognition processes and phonological selection processes, but they
also describe the activation of associated representations, as described above. Even
if the processing of verbal and non-verbal stimuli is independent, it is possible to make
cross-references between these two channels of cognitive processing. The picture of
a dog may activate the word “dog”, as well as the word “dog” let us imagine the
picture of a dog. The visual mental presentation can be a dog we know, or it can be
a prototypical representation of a dog. These kinds of cross-system connections are
called referential connections.

Whereas Paivio’s theories describe mental representations in general, Mayer and
Anderson (1991) were interested in its application in multimedia learning. They
conducted experiments which aimed to test the dual coding theory in a multimedia
learning situation. The researchers prepared an experiment in order to evaluate,
how the two channels, the verbal and the non-verbal, are linked together. There
were three hypotheses on how this relation could be: (1) The single-code hypothesis
states that there is only one channel for words and pictures; (2) The separate dual-
code hypothesis posits that there are separate channels for words and pictures; (3)
and finally the integrated dual-code hypothesis which states that there are separate
systems, but which do interact which each other. As expected, Mayer and Anderson
found based on experiments the third hypothesis to be true: subjects which were
presented a text and a picture simultaneously outperformed subjects in a recall and
retention test, who had text or pictures only, and they outperformed also subjects
which were presented both presentations, but consecutively. This led the authors
to the conclusion, that there must be an integrating process, when learning a text-
image presentation. In Fig 2.1, we present the adapted schema of Paivio’s dual-coding

The notion of representational and referential connections is most useful for us,
because we can measure this kind of learning behaviour with eye-tracking equipment,
which we discuss in a later discussion. When learners switch their focus between
different items in the textual or in the pictorial presentation, we can argue that
representational connections are built between these elements. When on the other
hand learners switch between the textual and the pictorial representation, we can
assume that referential connections are made between the two forms of presentation.
2.1.2 A cognitive process model of multimodal information integration and comprehension

Narayanan and Hegarty (1998) proposed a cognitive model of multimedia comprehension, which also was build on the idea of a dual system of pictorial and textual information. Their model describes the understanding and integration in a sequence of five stages of cognitive processing. The authors focused especially on diagrams involving complex mechanical systems, which do not only convey information about the configuration of the system, but also about kinematic processes when the system is in function. The authors tested for example representations of pulley systems, flushing cisterns or steering mechanisms in ships.

Stage 1: machine decomposition by diagram parsing

The first step in the comprehension process of a diagram is its decomposition in basic elements. At this stage visual recognition routines are used to recognize known elements or geometrical shapes. At this stage we already depend on top-down processes using our prior knowledge, as it can be difficult to identify the boundaries in a system, when we have little knowledge of it.
Stage 2: constructing a static mental model; making representational connections

Second, we look for information we already have about this object in our long time memory. We can find information about similar systems we have seen before and we can make predictions about the present system we are studying. At the same time we analyze the single objects, we integrate them into the context of the system, as we reason about functional or spatial connections between the objects. Both of these mental connections are called representational connections, as they implicate the same type of representations.

Stage 3: construction of a static mental model; making referential connections

In a further step, information coming from pictorial and textual stimuli is combined by making referential connections. At this stage, the learner combines corresponding textual and pictorial information. When a instructional text describes the interaction of two gears, we would integrate the description of kinematic processes in the text with the schematic configural information we find in the diagram. In section 3.5 we will see an experiment, which analyzes the text-diagram integration with an eye tracking device (Hegarty & Just, 1993).

Stage 4: determining the causal chain of events

We already made referential and representational connections between the different textual and pictorial elements, but Narayanan and Hegarty argued that we need to determine additionally a causal chain of the system on a more global level. Causal chains do not necessarily implicate spatial relatedness. Sometimes the configurational representation of a system can even be misleading and does not directly convey the inherent function of it (i.e., electrical circuit schemas).

Stage 5: constructing a dynamic mental model by mental simulation and rule-based inference

The final stages involves the integration of all acquired information about configuration and kinematics of the system and we can build a “runnable mental model” of the system. At this stage we can mentally simulate the functioning of single elements and their interaction, which is required if we want to reason on a higher level about the functioning of the system (i.e., for answering troubleshooting questions, which ask for measures in case of failure).

Empirical evidence for the model

Narayanan and Hegarty conducted several experiments to test their model (Narayanan & Hegarty, 2002). In an experiment they compared a conventional printed presentation of a flushing cistern system to a conventional hypermedia presentation and to
a hypermedia presentation which is designed using the guidelines derived from the model described above. As the authors have expected, participants in the group which received the cognitively designed hypermedia material outperformed participants in the group with a printed version of the material. Comparing the conventional hypermedia presentation to the cognitively designed material the authors did find a positive effect for causal descriptions and function questions in the cognitively designed material, but they did not find a significant effect on troubleshooting questions. The results of the experiment showed no difference between the printed version and the standard hypermedia version of the instructional material. The authors concluded that “it is the content and structure of instructional materials, and not the media and modalities in which they are presented, that is important for comprehension of complex devices.” (Narayanan & Hegarty 2002).

It was argued that the model could be limited to mechanical systems and therefore Narayanan and Hegarty repeated the experiment with a material on the merge sort and quick sort algorithm used in computer sciences. As mechanical systems, programming algorithms can be hierarchically decomposed. The algorithm can be explained in subsequent steps, in order to lower the initial complexity of the material. But in contrast to mechanical systems, algorithms have no spatial relations and there are no kinematics involved in the process. Algorithms are rather characterised by different states of data structures and the modification of the data structure over time. The second experiment used the same experimental design and could confirm the findings of the first experiment with the flushing cistern.

2.1.3 Cognitive theory of multimedia learning

One of the most cited models in educational multimedia theories is the “Cognitive Theory of Multimedia Learning” of Richard Mayer (2001). Mayer’s theory is an integrating approach, which is based on existing theories of the human memory system (Baddeley 2000) and on a number of experiments in this research field. Mayer makes three basic assumptions building the basement of his theory, which we will shortly present in the following paragraphs.

The Dual-Channel assumption is that humans have two separated channels for treating multimedia information. It is based on the dual coding theory of Paivio (see Paivio 2007). There is a visual-pictorial channel for depictive information, as for example diagrams or pictures and there is a auditory-verbal channel for acoustic information, as for example narration or nonverbal sounds. The theory says that information in these two channels is mainly processed independently and is not integrated until the end of the information processing.

The Limited Capacity Assumption states that for both channels there is a limit on how much information can be processed at the same time. This assumption is based on models of working memory (Miller 1956; Baddeley 2000) and the cognitive load theory (Sweller 2005). Regarding the visual treatment of information, the learner can focus his attention only on some visual elements at the same time and this is valid as much for the processing on the perceptual level as the processing of novel
information in working memory. Regarding aural information, the limit of words, that can be kept in the phonological subsystem of the working memory is a couple of words.

The Active Processing Assumption is that the learner needs to be actively engaged in the learning process, which includes paying attention, organizing incoming information and integrating incoming information with prior knowledge. This opposes a behaviouristic view of the learning process, where information is passively consumed and stored similar to a recording on a tape recorder.

In Figure 2.2 we can see a graphical presentation of the model of Mayer. The boxes stand for different representations of the information at different stages in the information process. Arrows indicate active processing of information. The two main channels of information are present: the auditory-verbal channel in the top frame of the diagram and the visual-pictorial channel in the bottom frame. Pictures and words are sequentially treated by the sensory memory and the working memory and they result in a separated verbal and a pictorial mental model. Finally the two models are integrated with prior knowledge from the long-term memory. The product of this integration is an integrated mental model which holds a multimodal representation of the concept the learner has constructed from the learning material.

Let us take a closer look at the processes. As we can read from the diagram, the first active process is selecting words or images. As stated by the limited capacity assumption, we have limited resources on how much information we can treat simultaneously. An important step in treating visual or aural information is thus to select the relevant information for the present task. When we consult a manual for the installation of a cupboard, we would have to find for example the place of a specific screw in the diagram (upper-left corner) and we need to extract the important words in the instructions (the name of the parts which have to be combined).

The second processes comprehend the organization of selected textual and pictorial elements in two independent models: a verbal model and a pictorial model. When we have a textual description, and we have selected the relevant parts of it, we do have to build a structure, on how the words are related to each other. Often in learning material this would correspond to building a cause-and-effect chain between the different information given in a text. The same process occurs when images are organized: parts of an image have to be put in relation to each other.

We can also see that there are two possible exchanges between the auditory-verbal and visual-pictorial channels. The first indicates a possibility for words to be treated by the eyes. The explanation for this finding is not surprising: we can not only hear and speak words, we can read them as well. But as the limited capacity assumption predicts, there will be a trade off when textual and depictive information are presented visually at the same time: the two sources of information are treated in the same channel, which leads to a competitive use of the limited cognitive resources. The second possible exchange between the two channels is located in the working memory. This kind of exchange occurs typically when reading a text. The visual information from the written words has to be translated back into a phonological representation of the text for further semantic and grammatical processing.
2.2 Instructional principles

One of the main goals of instructional design is to construct the learning material in such a way, that it is adapted to the cognitive capacities of the learner. The most prominent theory on the nature of cognitive learning capacity is the cognitive load theory, which we will present in the following section.

The cognitive load theory finds its first origins in the works of G.A. Miller (1956), who suggested that we have a limited capacity to process novel information. Miller found that our working memory can hold a “magic” number of seven elements, where a deviation of plus or minus two elements can be observed in experimental testings. Miller also suggested that one could increase the capacity of elements to be processed by chunking different elements to meaningful groups of information. Thus we could memorize the telephone number 0442051976 as a sequence of ten digits or we can chunk them to meaningful pieces of information: the 044 could be the area code of a geographical region, 1976 could be the birth year of somebody we know etc.

Nevertheless, even if we can apply some strategies to improve the effectiveness of our working memory, its capacity stays limited. The limitation concerns the numbers of elements we can hold in our working memory as well as the time we can keep them activated in working memory without auto-repeating them. After about 20 seconds almost all the contents are lost in working memory (Peterson and Peterson 1959 cited in Sweller, 2005). Richard Mayer (2005) defines cognitive capacity as “the total amount of processing that can be supported by the learner’s working memory at any one time”.

A limited working memory has a great impact on how we can process novel information. The learner has to select a small number of information pieces and keep them in working memory, in order to process them. When all relevant elements of information can be processed simultaneously in working memory, we can speak of successful understanding (Sweller 2005). Seven elements may seem to be little, but the cognitive work we accomplish is less to keep the elements in our working memory, but to relate these elements to each other (i.e., causal-effect-relations).

What are the implications for the design of instructional material, when we have a limited capacity of processing novel information? Generally, we could conclude that we should limit the number of information we present simultaneously to the
learner and secondly, that we should only add information which is relevant for the understanding of the material in order to prevent an excessive cognitive load (overload). Although, the contrary is also true. When we present less information to the learner than he could process in working memory, his learning performance would decrease as well (underload). The goal of instructional design is therefore to adapt the learning material in such a way, that it is aligned to the cognitive capacity of the learner and will neither underchallenge nor overchallenge him. Whereas increasing the complexity of a learning material should not pose a lot of problems, the contrary is often true. How do we limit the cognitive load when we have a complex system to explain? We cannot always remove the complexity as it would also remove the essential facts of the system we want to convey.

Cognitive load theory tries to dismantle the complexity of a learning material as it distinguishes between three different types of cognitive load (Sweller 2005; Paas, Renkl, & Sweller 2004).

Intrinsic Cognitive Load is the cognitive load which is imposed by the number of elements and their interactivity. The term interactivity refers to the grade of connections which exists between the different elements. It is due to the natural complexity of the material.

Extraneous Cognitive Load or ineffective load is the cognitive load which does not contribute to the process of learning. Extraneous load is related to the form of presentation. A typical example is the extraneous load which is caused by the action of scrolling the window to continue to read.

Germane Cognitive Load or effective cognitive load is as the extraneous cognitive load a result of the form of presentation, but this time it contributes to the learning process. A typical example would be the additional cognitive load induced by the presence of different types of representations of the same content (textual and descriptive) to foster a deeper understanding.

The cognitive theory states that these types of cognitive load are additive. In consequence, when the intrinsic cognitive load can not be altered by instructional design, extraneous cognitive load should be eliminated as much as possible and the germane cognitive load should be increased as much as it does not overload the total cognitive capacity of the learner. Respectively, another consequence of that theory would be that when the intrinsic cognitive load of a material is low, even a material which is badly designed (i.e., with a high extraneous load) can teach successfully the material.

This limited capacity affects only the novel information coming from the sensory system, the activation of stored knowledge in the long-term memory is not affected by this limitation (Ericsson, Kintsch, Sweller cited in Jamet 2006).

Even if the elements and their interactivity are linked to the natural complexity of the material, we still can lower the intrinsic cognitive load in changing the number of elements we show simultaneously. This principle, called segmenting principle, will be presented in section 2.2.3.
2.2.1 The multimedia principle

The conventional wisdom on learning with multimedia assumes that pictures are more powerful in conveying information in a learning material than texts. Even our proverbs teach us that “a picture is worth a thousand words”. We could also follow the logic of John McCarthy [1] who is quoted that “As the Chinese say, 1001 words is worth more than a picture”. McCarthy invented the Chinese saying, but he was probably right in his humorous statement, that we cannot measure the advantages of images and texts by a simple comparison. The question should rather be, in which cases pictures are suited to help the learner in understanding a learning material rather than text and vice versa.

Studies have shown that graphics do also have draw-backs compared to a textual presentation. One major negative characteristic is that “in contrast to text, that is linear, there is no standard way of processing diagrams.” (Bétrancourt & Tversky submitted). When a teacher writes a text-book he structures the text with intent to progressively introduce the reader to the topic. Starting with the basic ideas he prepares the learner for the more complex parts of the topic. This kind of guidance does not exist in a static picture and a novice learner would likely be overwhelmed by having only a pictorial representation of a complex topic. We do rarely see diagrams alone in a learning material for this reason. Rather diagrams are accompanied by written or spoken text, in order to explain the pictorial elements in a diagram and to guide the learner in reaching the learning goal.

Still, images are a powerful tool to present visuospatial information, as given for example in maps or portraits, or information which is not inherently visible, but which presents relations on an abstract level, as for example in graphs or organization charts (Tversky et al., 2002). Depictions are especially suited to draw inferences, descriptions are more suited to represent subject matters (Schnotz & Bannert, 2003). Also, graphics are especially helpful as they can be used as an external memory aid, which stores information on different elements and their relations. When the pictorial presentation has the disadvantage of being non-linear, it proves to be an advantage when we integrate information provided by the textual information with the pictorial presentation. We can access specific elements in the diagram as we need them, freeing cognitive capacity as we do not have to keep them in working memory. Eye-tracking studies (Hegarty & Just, 1993) have shown how people integrate progressively textual information with their corresponding pictorial elements. These studies show also that we integrate periodically the different sub-parts in the diagram into their larger context. This progressive integration indicate that the learner follows the aim of the instructional designers building his knowledge in a coherent order.

Further, images are beneficial as they offer an alternative method of presenting the same information. According to the dual coding theory (Paivio, 2007), images and words are a complementary form of presenting information, as they involve separate cognitive mechanisms. As consequence, presenting information in showing a textual and a pictorial version can improve learning performance. Mayer (2001; 2005) and

Schnottz (Schnotz & Bannert, 2003; Schnotz, 2005) both based their integrative model of multimedia learning on this theoretical foundation.

Mayer (2001) summarizes a large number of empirical testings in his Multimedia principle, stating that one does better learn from words and pictures, than from words or images alone. Still, it has to be emphasized, that this statement is not always true for all types of pictures in every context. The use of pictures is only helpful if it is relevant for solving the task. Purely decorative pictures, which do not support directly the learning process failed to prove a positive effect on learning performance or motivation. Rather they interfere with instructional aims (Kalyuga, Chandler, & Sweller, 1999).

But what exactly do we mean by multimedia. The term is used in different contexts and asks for some clarification. According to Mayer (2001) the term multimedia can be viewed in three ways:

the delivery media view is a techno-centered view, which emphasizes on the different technological devices which can be used to deliver multimedia content (i.e., a computer, a screen, a projector, a video recorder etc.)

the presentation modes view is a learner-centered view, which emphasizes on the different types of a presentation, for example the same information can be presented in a verbal or a pictorial form

the sensory modalities view is also a learner-centered view, which emphasizes on the sensory systems which process the multimedia information, for example a textual information can be listened to (narration) or it can be read (written text)

The first definition is of little interest in a psychological perspective and will not be used in this paper. Our experimental design, which is presented in a later section, addresses only the visual channel and therefore the sensory modality definition is less prominent, but should be considered during the theoretical part of this paper. These definitions also exclude another misleading definition of multimedia: a multimedia presentation does not imply that a technological device is used to display the information. A text-book with illustrations is also a multimedia presentation from an educational engineers perspective.

What about computer supported learning? Is learning more efficient when using a computer? In the last decades, learning with computers has been object of research and some interesting new possibilities of presenting information have been investigated. First, computers offer new ways of displaying information, as visual information can be display as animation (see i.e., Tversky et al., 2002) or 3D-displays. The second change computers brought to educational design was the new way of interacting with the learning material. The use of hypertext was one of the first approaches which was explored, but also the use of pop-ups (i.e., Jamet & Erhel, 2006; Bétrancourt & Bisseret, 1998), controlling the speed of animations (i.e., Mayer & Chandler)
A simple but very powerful advantage of electronic documents is the possibility to search them quickly. We can handle large amounts of information at our finger tips. But electronic documents lack some features, which we appreciate when using paper-based documents: they offer an easy navigation (turning pages), we can annotate information just by writing or marking text and finally paper-based documents still offer a better readability in terms of speed and tiredness (Bétrancourt, Dillenbourg, & Montarnal, 2003).

### 2.2.2 Split-attention effect

Imagine if you read a book with an illustration of a mechanical device on one page and the explaining text on the next page, which obliges you to turn the page each time to compare the element and the corresponding description. This situation would probably be annoying (it will likely happen in the present paper as well). The example illustrates what is known as the split-attention effect. The split-attention effect occurs in situations when corresponding elements are temporally or spatially disparate in a learning material. The learner has to split his attention between two sources of information and has to integrate the disparate elements. This integration comes with a cost in form of extraneous cognitive load, a cognitive load which lacks for an effective processing of the material.

When Ayres and Sweller (2005) speak of one single effect, Mayer (2005) distinguishes between two types of conditions, this time formulated in a positive way. A material has a good spatial contiguity, when the different corresponding elements are presented in a spatially integrated form. A spatial integration would be when the textual description is written right next to the element which is described. But there is another remedy for the split-attention due to separated visual source of information: the textual description could be presented acoustically by adding an auditory description of the pictorial elements. The use of two different modalities (visual and acoustic) seem in general to be superior to only visual presentation (i.e., Mousavi, Low, & Sweller, 1995).

Although, the split attention effect only occurs when “...the logical relation between sources of information is critical for the split attention effect. The effect can only be obtained when multiple screens of information are essential for understanding and so cannot be understood in isolation.” (Ayres & Sweller, 2005). As we have seen in the cognitive load theory, a negative effect of extraneous cognitive load is only to be expected, when there is a high intrinsic cognitive load, as for example when the material is high in element interactivity.

Sometimes, it is not practical to include the description directly in the diagram, as it would become visually overloaded. A possible solution has been proposed in providing the corresponding information in a pop-up as soon as the element has been clicked. It seems that this method can indeed ameliorate the recall of elements and
the deeper understanding of the material compared with a material where the textual information was separated. Although, the presentation in pop-ups seem not to be superior to the integrated display per se (Bétrancourt & Bissereet, 1998).

There is a second form of split-attention, which concerns elements which are temporally disparated. This effect occurs for example when first the textual description is shown and in a second step the according diagram is shown. This effect is also possible, when the material is presented in a bi-modal (visual, aural) form. In his design principles Mayer (2001) recommends that the temporal contiguity principle should be adopted: “Students learn better when corresponding words and pictures are presented simultaneously rather than successively”. The positive effect of temporal contiguity for example has been shown in experiments of Mayer and Anderson (1991) for a bi-modal presentation of a bicycle pump.

2.2.3 Segmenting principles and sequential display

Longtime it was assumed that the intrinsic cognitive load can not be altered, as the number of elements and their interaction are the very concept of a material. Elements can not be simply omitted, as the system would be overly simplified and misinterpreted. Although, another technique was proposed to ease the complexity of system: a progressive display of the elements over time. In this way learners could build progressively their mental model, which would lower the cognitive load, as the learner has to process less elements simultaneously. Researchers have suggested different solutions to design a progressive presentation.

A first, quite natural way of presenting images over time lies in animations. Especially in situations, where physical movement is involved in the presentation, they could be beneficial in conveying a complex process. However, results from animations are ambivalent (Tversky et al., 2002). One factor which seems to influence negatively the learning process is the continuous play-back of the presentation, which can lead to an overwhelming feed of information, which is not adapted to the individual learning speed. Another possibility is offered by segmenting a diagram into different frames and present them progressively. This type of display would be halfway between a static display and animation.

Mayer and Chandler (2001) proposed to segment a multimedia presentation consisting of an animation and a corresponding acoustic explanation. The presentation conveyed the formation of lightening in sixteen steps, a material which is characterized by strong causal effects between the different steps. The system was user-paced, which means that the learner could decide, when he wanted to proceed to the next segment of the presentation. It was assumed that the segmented display alone would foster the understanding of the different steps, while having the negative effect of reducing the integration of the material in the larger context. The authors designed their experiment that four experimental groups would see either the presentation two times in one single sequence (whole-whole), or two times in the segmented form (part-part), or one time in segmented form then one time in a single sequence (part-whole) and finally vice-versa one time in a single sequence and one time in a segmented form.
(whole-part). It was found that the subjects in the part-whole presentation outperformed significantly the subject in the whole-part presentation regarding questions which needed a causal reasoning (transfer question). Similarly, it was shown that in a part-part presentation subjects showed higher scores in transfer questions than subjects in the whole-whole group. In contrast, no significant differences could be found for questions which only demanded a recall of the presented elements (retention).

Pollock, Chandler and Sweller (2002) did an experiment, where electrical engineering students had to learn how to conduct a test on insulation resistance. The authors argued that a method must be found to facilitate learning material, which is characterized by high-element interactivity. They proposed to present the material in two steps: first the sub-parts of the learning material were presented as isolated elements, where no interaction between the sub-parts was shown and in a second step all sub-parts were presented in their totality (interacting). The aim was to reduce the cognitive load as the complexity of the material was split into two different learning steps. The second group just viewed a presentation where only the integrated form of the presentation was shown. The results revealed that questions which required to understand relations between the sub-parts were better solved by the group, which had two different steps in their presentation (isolated-interaction). No significant effect could be found for questions where only knowledge about the processes in the sub-parts was necessary. Students in the isolated-interaction group also reported that they found the material less demanding than the interaction-only group.

Instead of segmenting the presentation in different parts, and showing them one after the other, it was proposed to present a diagram in a progressively growing way. This method is known as sequential display and offers a guided explanation of a complex system as it conveys “the organization and inherent logic of the instruction, just as a teacher draws a schema on a blackboard in a carefully chosen sequence” (Bétrancourt et al., 2003).

A sequential display of a diagram should facilitate the learning process for different reasons. According to the model of Mayer (2001), presenting information sequentially make it possible for the user to construct first a local mental model, which can be later integrated in a coherent mental model. Further, sequential display should be beneficial, for the following reasons (Hidrio & Jamet, n.d.; Jamet, 2006) as:

- it lowers the cognitive load presenting information gradually, because among other reasons
  - there are less visual elements visible at the same time, which should make it easier to process the information;
  - the learner needs less time for visual search;
- a sequential presentation offers a coherent order of information processing;
- the sequential presentation can be used to regroup the elements of the same category in time, even if they are spatially distant;
• when presenting a dynamic system, a sequential display facilitates the construction of a functional mental model.

Additionally, the sequential displays could provoke a constant vigilance caused by the appearance of new stimuli (Jamet & Arguel, in press).

Wright, Hull and Black (1990) introduced in their experiment a sequential display showing different companies and their business relations. The experiment revealed that subjects had equal scores in the subsequent true-false quiz compared to subjects with the static display, but they were significantly faster (8.4 min instead of 15.4 min) in learning the material. Their experiments showed that a diagram proves helpful, however, learners did not consult it often when they could choose.

Bétrancourt and Bisseret (1995) made the hypothesis that the order of presentation in a sequential display affects the order in an immediate and deferred recall and recognition task. They used the figures of Levelt (1982), which they presented in different forms of linearization (no linearization, route pattern, square pattern, cross pattern, reading pattern). The authors could demonstrate an effect of order on the recall task depending on the pattern used in the presentation. No main effect could be found for the score in the recognition task (same or different task), but response time differed significantly between the experimental groups. Nevertheless, the recognition of patterns was improved when they corresponded to the linearization of the presentation. It was not expected though, that the group with no linearization of the material (static presentation) had a higher performance in recalling elements in a first try.

Bétrancourt, Bisseret and Faure (2001) presented an experiment where two different materials were used to illustrate the effect of a sequential display. The authors used their findings of 1995 and identified three strategies how a material could mentally be structured:

• a route strategy, where elements are related by their spatial proximity. The subject would follow a trail of the different elements on the diagram;

• a hierarchical partitioning strategy, where the elements are organized by different characteristics (by their appearance or function);

• a schema application strategy, where elements are organized using a picture-independent linearization, for example by processing them from left to right.

The authors predicted that a sequential presentation which follows one of these linearization strategies would influence the organization of the subject’s mental representation.

The first presentation showed a city map, where elements on this map were presented gradually in three different forms: according to their functionality (every-day life criteria), according to their spatial hierarchy (proximity and symmetry), and according to their linear horizontal organization (display starting from top-left in a
occidental reading order). The second material was a library map with the same type of sequentialization.

The results showed that even if the sequential display did not lead to a better performance compared to the static display, the sequential display did significantly influence the recall order (except for the group with a spatial hierarchy). The experiments would suggest that we can influence the organization of elements in the mental model, when we use a corresponding order in the presentation of a sequential display.

Another experiment on sequential displays was conducted by Bétrancourt, Dillenbourg and Montarnal (2003). The material consisted of a lesson out of a textbook on accounting. The lesson explained how to transform an accounting balance sheet into a financial balance sheet. The authors introduce a novel element in the experiences, when they allowed the learner to control the order of the segmented material. It was already shown that a simple control of when to proceed to a subsequent step can improve learning performance (Mayer & Chandler, 2001). Giving the control over the order of presentation has two different factors to be accounted for: on the one hand a free exploration of the material can help to engage the learner in an active learning process as it is proposed by Mayer (2001), but on the other hand findings from this type of exploratory learning have not shown the expected success (De Jong & Joolingen, 1998). There were three experimental conditions in their experiment: (1) a sequential non-interactive condition, which had no control expect to continue to the subsequent step; (2) a sequential interactive condition, where students could control the order of the steps; (3) a static condition, where the textual description was presented in in an adjacent window. It was found that the learners with a sequential interactive display were fastest, followed by the students with a static and non-interactive display. Both groups with a sequential display outperformed the group with a static display, but there was no main-effect of type of display in general. However, when covariated with time, a main-effect of the display type could be determined. A surprising fact was also, that the learners rated the sequential displays as having of less “pedagogical value”.

Éric Jamet (2006) explored also the possibilities of sequential displays. One of his research questions was whether sequential displays prove only helpful for materials with an inherent temporal order or if sequential displays facilitate also learning with other, more conceptual materials. In two experiments he evaluated the effect of sequential displays for two different materials: a diagram which presents the different types of language disorders on a two-dimensional map and a diagram which explains the cerebral hemisphere lateralization. Both materials had no inherent temporal order, but were segmented in time by presenting them sequentially. He found that the sequential display is superior to the static display and this both for questions which asked for a recall of information presented in the material, as well as for questions which had to be deducted from the given information (transfer questions).

In a later study (Jamet & Arguel, in press) sequential displays were tested with a material where first aid measures were explained. This material seems to be particularly adapted to a realistic scenario as this kind of material is not meant to be
read in the situation where it is used. For the first time in the research of sequential displays, the diagram was accompanied by narration instead of written text, a form of presentation which should favour additionally the construction of a mental model (see [Mayer] 2001 or section 2.1.3). Furthermore, the authors made the hypothesis, that the sequential should be superior to a static display as it respects the congruence principle and the guidance principles. The congruence principle states that a material “should correspond to the desired structure and content of the internal representation” ([Tversky et al.], 2002). The guidance principle describes the beneficial effect when the attention of the learner is guided in a way, which should result in the construction of a consistent mental model. The experiment showed that subjects in the sequential presentation group had better scores in recall of the facts and the procedures. Also the learners who learned with a sequential presentation had better scores in transfer questions, however no significant effect could be found for the evaluation of dangerous actions the subjects proposed in their first aid measures.

Jamet, Gavota and Quaireau (2008) expanded the method of displaying a sequential display by additionally highlighting the elements which are presented. The highlighting of elements acts as another means of guiding the attention to the corresponding elements, preventing an excessive visual search in the diagram, which would in turn result in a raise of extraneous cognitive load. The authors designed the experiment with four different types of presentation: a static presentation, a sequential presentation, a static presentation with highlighting and finally a sequential presentation with highlighting. The results were partially in accordance with the author’s hypothesis: the sequential display resulted in a better performance in diagram completion tasks and function retention tasks. However, there was no significant effect on transfer tasks or process retention tasks. The authors argued that “the sequential element-by-element presentation or its highlighting might have reinforced the individual processing of the elements at the expense of a more global processing of the diagram.” ([Jamet et al.], 2008). In contrast to earlier studies ([Bétrancourt et al.], 2001), no effect on the organization of the learned material in memory could be observed. The authors made the supposition that the explanation could be the difference in modality between these two studies. The presence of spoken guidance may turn the guidance by the sequential display unnecessary. The second factor highlighting had an overall positive effect on learning performance, except for the transfer tasks. It had even more an additive effect with the sequential display.
3 Eye tracking and visual exploration

The effect of different multimedia presentations is usually evaluated by off-line tests. Typically participants fill out questionnaires at the end of the learning phase, where their ability to answer to retention or inference questions on the topic is evaluated. Additionally participants can be asked to evaluate the material itself for its difficulty or enjoyability. However, these test do not reveal information on the learning process itself, they just evaluate the result of it. In contrast, there are tests which are called “on line”, which designate methods that evaluate behaviour during the experiment itself. A known technique is to use a secondary task in the experimental design. It can be expected that the performance in a simple task (i.e., counting backwards from hundred to one) is lowered when the primary task (the actual learning task) demands a high cognitive load. Modern and direct ways of measuring cognitive activity are brain imagining techniques like functional magnetic resonance imaging (fMRI). But the learning process is complex and is composed of different mental processes, what makes it difficult to deduct learning processes by measuring levels of brain activation in different brain areas.

In this chapter we focus on another on-line technique to evaluate learning behaviour: eye tracking. Studies on eye movements are numerous and they are popular in different research areas. Eye trackers are used to evaluate the way we read medias like newspapers (Holmqvist, Holsanova, Barthelson, & Lundqvist, 2003), in usability evaluations of computer interfaces (Duchowski, 2006), in marketing research and in many other fields of research (see Duchowski, 2002 for an extensive overview of eye tracking applications). Eye tracking can give us information how learners process the visual information, in particular the order of how they inspect the learning material. We start by giving a short introduction of different eye tracking techniques and their applications. We also want to address some controversies about the interpretation of eye tracking data and finally we look at experiments in multimedia learning which used eye tracking measurements.

3.1 Characteristics of eye movements

Even if the so far mentioned studies are quite recent, it is wrong to assume that the technique of eye tracking is an invention of modern times. The first research of eye movement dates back to 1879, when Javal, an ophthalmologist, analyzed eye movement during the lecture of text by simple observation. It was Javal who discovered the basic findings on eye movement, in particular that eye movement is not continuous as it was assumed, but that eye movements consist of a rapid succession of fixations and saccades.
Saccades are called the movements of the eye, which take between 20-200 ms. In text reading a typical saccade would take about 30 ms and move for about 2 degrees. In scene perception saccades take slightly longer with a duration of 40-50 ms for a distance of 5 degrees. During a saccade we cannot perceive visual information, the movement is too rapid (about 500 degrees a second) and the image would be blurred. This effect is known as saccadic suppression, which means that we are actually blind during eye movements. A saccade is followed by a fixation where the eye is relatively still. Mean fixation durations when reading a text range from 225 ms in silent reading until 400 ms when typing a text (Rayner, 1998). A third characteristic of eye movements is the saccade latency. Even if the target of the next saccade is clearly defined, it takes the eye 150 ms – 175 ms to plan the next saccade.

Another characteristic of visual perception is the perceptual span. The visual field has different degrees of acuity. We distinguish between three regions of vision: the foveal region, the parafoveal region and the peripheral region. The fovea is the region with the highest resolution of cones and with very good acuity. When we look at a point we focus on it, in order to have it in the 2 degrees of foveal vision. The foveal region is surrounded by a 5 degree circle of parafoveal vision, which is characterized by a lower acuity, but it still provides important visual information. The region which extend to the total of 150 degrees of our visual field is called peripheral.

Whereas foveal vision is used to inspect a visual object in detail, parafoveal vision is used mainly to plan next saccades. Although, it seems that some information can already be extracted from parafoveal vision, for example the characteristics of a subsequent word in text reading (Rayner, 1998). The peripheral region of vision provides little information about the visual characteristics of an object, but the area is sensitive to motion, which let us guide our visual focus, when a salient new information comes into our visual field. When reading, the eye does not always move forward. Fifteen to twenty percent of eye movements are regressions. When regressions span over a few letters they are often a sign that the reader has difficulties to integrate a word. Regressions of several words on the other hand indicate usually a problem of text understanding.

Whereas characteristics of eye movements in text reading are specified in detail, the data on image perception is less conclusive. The major difference between text and images are that text has an inherent order of processing (in western cultures we read from left to right and from top to bottom). Images do not feature an unambiguous way of processing. It was Yarbus who has shown in his experiments that the visual processing of an image is highly dependent on the task the participant was given. When no specific task is given, salience of visual elements guide preferably our attention (Yarbus, 1967).

3.2 Eye tracking devices

Eye trackers require to have a high spatial and temporal resolution to be able to measure the rapid eye movements. In order to accomplish that, measurements can
be obtained from different physiological characteristics related to eye movements. Eye tracking systems rely either on (a) electrodes which are attached to the skin close to the eye, (b) systems which measure the reflection of infrared-light on the cornea, (c) video-based eye trackers, (d) Purkinje images which are generated by a number of reflections on lens and cornea, and (e) search coils which are attached like contact lenses to the surface of the eye (Rayner, 1998).

The choice between these different eye tracking mechanisms is based on the characteristics of the eye movement which is most important for the effectuated study. Whereas electro-oculographic methods are most accurate when duration of saccades have to be evaluated, scleral search coils are most effective when determining the direction of the gazes, as they measure the change of the magnetic field when the eye moves. Both systems are rather intrusive as measuring equipment has to be attached to the eye or on the skin around the eye.

Today’s most commonly used eye tracking apparatuses are infrared corneal reflection eye trackers. In these systems infrared light is projected on the eyes, where it is reflected on different levels (two reflections on the cornea and two reflections on the lens). The reflected light is then recorded by video cameras sensitive to infrared light. The four different reflections permit to calculate the position and direction of both eyes. As there are individual differences in the anatomy of the eyes (distance between eyes, but also differences of the eye itself), systems which depend on reflection of light have to be calibrated to the participants anatomy to assure a precise measurement.

3.3 Analysis and visualization of eye movements

An eye tracking device measures in general the coordinates and the duration of fixations. But eye movements can be quite erratic and a fixation does not mean that an object was processed at a cognitive level. There are for example micro-movements who compensate for shifts due to oculomotor faults. Also, saccades do not find the targeted object on the first try, but they often approximate the targeted object. For this reason fixations are usually aggregated into gazes, which are typically defined to a fixation of at least 100 ms in an area of maximum 30 pixels.

An often used visualization method for eye movement data are scan path diagrams. They use dots whose radius indicate the fixation duration and lines to indicate length and direction of saccades (see Fig.3.1). This kind of eye tracking representation proves especially useful if the order of eye movements is important. A draw back of this method is that only one subject’s eye movements can be visualized.

Another method of visualization of gazes are heat maps. Heat maps are color-coded representations of aggregated eye tracking data. Red colored areas would for example indicate regions with high overall fixation times, whereas blue areas are rarely looked at. The advantage of this kind of representation is that eye tracking data of several subjects can be aggregated into one single heat map, allowing to make some basic generalizations about the visual processing of the material. Also
heat maps are easy to interpret and do not need deeper understanding of eye tracking metrics. A trade-off of heat maps is that all information about the gaze order is lost.

Scan paths and heat maps are well suited to visualize eye movement data, but they proof to be hard to interpret, when used in quantitative research. Often gaze data is compared to area of interests (AOI) to make them easier to quantify. Areas of interest are predefined areas, which have a meaning according the research question. Typically an AOI consist of a paragraph of text or a pictorial element which does carry some semantic information. AOI allow us to make conclusions about how often a specific object was looked at and in what order the different objects were inspected. Typical metrics would be gaze percentage per AOI or number of fixations per AOI.

In studies of human-computer action, we have additionally the problem, that visual stimuli change over time and we need to take into account the actions of the participant when using the system. The methods above do not match these requirements. Narayanan and Crowe (2002) proposed a visualization named “Attention-Action time line”, which includes the time passed on different areas of interest and multiple screens. The visualization also included information about haptic actions by the participant (i.e., mouse clicks, key presses).

3.4 Controversies on eye tracking

Eye tracking devices are commonly used in today’s research and they provide an accuracy which is sufficient for most applications. Nevertheless, there are still some controversies about the validity of the results provided by eye-tracking data. Some critics concern the interpretation of the data on the lower cognitive level, which we would like to address briefly in the next paragraphs (for a complete discussion see Starr & Rayner, 2001).

It is commonly accepted, that eye movements are influenced by two control mech-
anisms. On the one hand there are automatic visuomotor factors (the oculomotor model). When reading a text for example, we usually focus the next word at a position between the beginning and the middle of the word, regardless of linguistic characteristics of the word itself. On the other hand there are factors which influence the eye movement in a top-down process (process model). For example it was shown that we fixate words of a low lexical frequency longer than words with a high lexical frequency. The first controversy consists in the question on which of these two models explains primarily our eye movement behaviour.

Another topic under discussion is to what extent we can process information from parafoveal vision. It is known that we use parafoveal vision to plan our next saccade, but some studies show that we can already extract information from this visual area. When for example the point of uniqueness in a word is located in the first letters of a word (as for example in the word “quarantine”) the rest of the word can be skipped, as the word is already recognized at “a first glance”.

In eye movement studies, it us usually assumed that the processing of visual information is sequential. Although, this assumption is not fully accepted. It has been shown in text reading, that the forthcoming word in the right parafoveal region can facilitate the recognition of the word currently read. These findings challenge the idea of sequential processing and support the idea of a parallel processing of visual information.

There are other critics of eye tracking data, which address the influence of high level cognitive processing on eye movements. As Javal already has shown, eye movements are highly task-dependent. This is somewhat true for text reading, but it is very much the case for image inspection. Therefore, when we analyse eye movements, we also have to consider the type of instructions given to the participant, as well as the expertise of the participant. Respectively, it is still a matter of controversy of how eye movement data let us make conclusions about higher cognitive processes like learning.

Still, some findings in cognitive research seem to indicate, that visual perception and higher cognitive processing of visual information are highly connected processes. Neurophysiological studies have shown that mental imagery and visual perception share the same areas in visual cortex (Kosslyn, 1994), which let us assume that the same processes are involved. It has also been shown, that when participants imagine a picture they have previously seen, they move their eyes as if they would percept the image (Johansson, Holsanova, & Kolmqvist, 2005).

### 3.5 Eye tracking in multimedia learning

As we have seen, research has provided us with a few models which seem to be adequate to predict the outcome of different types of presentations on learning performance. An important aspect of multimedia learning research lies in the question how we integrate these different types of presentations (in particular pictorial and textual). In order to gain more empirical evidence for the integration process, eye
tracking measurement was proposed.

Hegarty and Just (1989) were one of the first who did eye tracking investigations in the field of multimedia learning. The authors identified three characteristics of diagrams which can aid the learner in building a mental model of the system: (a) “A diagram can depict spatial and visual properties of a device that have also been verbally described in the text.”; (b) “a diagram can act as a memory aid to reactivate the representation of information that has been previously read and presented”; (c) “A diagram can be a source of new information that is not given in the text”. According to these assumptions, the authors have defined a priori three different types of inspections, which they expect to see in the eye tracking measurements (see also Wipfå 2007).

**Formation inspections** are inspections of the diagram, which aim to look for information in the diagram for a text passage the participants have just read. The learner wants either to verify if he has understood correctly the description in the text or he refers to the diagram when he has difficulties imagining the description given in the text.

**Reactivation inspections** are inspections of the diagram where the learner looks at components of the system, which he has read about previously. These references help the reader to reactivate a representation he has already constructed, but which is no longer activated in working memory.

**Elaboration inspections** are inspections of the diagram which aim to acquire new knowledge, which is not present in the text.

As learning material the authors choose a 2-pulley-system and they assigned the participants either to a group which had a longer textual description or to a group which had a shorter textual description of the system. Both groups viewed the same diagram with the pulley system. The results showed that most diagram inspections were made at linguistic boundaries in the text. Participants preferably finished to read a clause before they switched to the diagram to inspect the corresponding element(s). This would be in favour of the theory that participants progressively integrate the textual information with their pictorial correspondences. In contrast to their hypothesis, reactivation inspections (gazes on parts of the system, which did not correspond to the text they just have read) were numerous. It was therefore argued that reactivation process could also have the purpose of integrating the different parts of the system.

In 1993, Hegarty and Just conducted a further experiment (Hegarty & Just, 1993), where a text-image presentation of a 3-pulley-system was used (see Fig 3.2 for the pictorial part). The authors changed their initial classification of eye movement behaviour and now distinguished between local and global inspections.

**Local inspections** are generally shorter inspections which include not more than three components. They serve to establish co-references between expressions
in the text and the corresponding parts in the diagram. Local inspections also include the adjacent parts of the described part in the text.

**Global inspections** are longer and include more than three components. They are used to integrate the different parts of the system in order to construct a mental model of the components and their interaction.

![A 3-pulley system](image)

**Figure 3.2:** A 3-pulley system, copied from Hegarty & Just, 1993

The authors made two hypothesis: “When they (the participants) look to the diagram after a reading episode, they should inspect it globally . . .” and “If the subject integrates the information in the text and diagram on a local level, they should inspect the diagram more often, their inspections should occur in mid text and not just at the end of the text, and their inspections should be focused on those particular regions of the diagram that are related to the most recently read section of text.” (Hegarty & Just, 1993).

Their findings confirmed the results of their previous research and their hypothesis was corroborated. Global inspections of the diagram occur with preference when having finished a paragraph (M = 53.7%, SD = 23.2). In the case of local inspections, 36% of diagram inspections concerned elements which they have just read and a further 44% percent of inspections included the next two most recently read clauses.

Tabbers (2002) conducted an experiment with a learning document where the effect of modality and user interaction on the visual exploration of the document was analyzed. The document was a diagram accompanied by either an oral explanation or textual explanation. In the case of the textual explanation the presentation was either system- or user-paced. It was found that learner had less, but longer fixations when the descriptions were presented acoustically. Not surprisingly learners spent more time on the diagram in the audio condition (98% of fixations) compared with the conditions where the textual descriptions were presented visually (44% in system-paced, respectively 38% in user-paced condition). In contrast to the assumption of the authors, the visual exploration did not differ in the system-paced and the user-paced condition.
An interesting question is also the effect of learners prior knowledge on the exploration behaviour while studying the document. It is already known that a presentation form, which can be beneficial for novices can be detrimental for learner with a good expertise on the subject (Kalyuga, 2005; Kalyuga, Chandler, & Sweller, 1998), but it has not been yet evaluated what differs experts from novices according to their learning strategies. Boucheix, Lowe and Soirat (2006) evaluated the on-line processing of the piano mechanisms with three different experimental groups: (a) piano makers and repairers; (b) pianists of a music teacher school; (c) psychology students. Not surprisingly the scores in an off-line test differed significantly between the three groups. The focus of this experiment was more the strategy of exploration and the authors could find the expected differences in eye movement patterns. Experts did inspect more often the functional and essential parts of the piano mechanisms, whereas novices (both pianists and psychology students) fixated rather on the most salient parts in the animated mechanism (the ones who moved most).

Schneider (2007) and Schneider, & Boucheix (2007) investigated the effect control and signaling on visual investigation of the material. As material they chose an adapted version of the three pulley system as it can be found in the experiments of Hegarty and Just (1993). They presented the material either in an animated or a static format.

For the analyses of the eye movements the authors distinguished between three types of visual transitions:

- global transitions are defined as transitions between two non-neighbourhood AOI
- local transitions are defined as transitions between neighbourhood AOI
- causal chain transitions are defined as transitions between at least three neighbourhood AOI

In their experiment they found more local transitions than global transitions, which was again higher than the number of causal chain transitions. Overall there were more transitions when the material could be controlled by the learner. As we have already seen in the general introduction on eye movements, the visual inspection of a material depends on the task which is given. In the case of the experiment of Schneider and Boucheix, the type of task (orientation on functional elements, local kinematic elements or configuration elements) did influence the eye movement behaviour. There were less local and chain transitions in the configuration elements orientation.
4 Experiment and hypotheses

The aim of our experiment was to investigate the effect of a sequential display on the visual inspection and the memorization of a technical text-picture document. We constructed a learning material of the functioning of an e-voting system. The participants viewed either a static version or a sequential version of the same document. The participants in the sequential condition could display the consequent chunks by pressing a button. The learner’s eye movements were recorded while studying the diagram, in order to measure the inspection order and the number of representational and referential connections. Also, the learner’s performance was tested in a pre- and post-test questionnaire.

The questions in the questionnaires were of three different types, which are briefly explained:

**Function retention questions** are questions which concern the functioning of elements described in the document.

**Process retention questions** are questions which concern the temporal aspects of the flow of information in the system.

**Troubleshooting questions** are questions which demand to solve or explain a possible problem in the system. The answers to these questions can not be directly found in the document, but they have to be deducted from a mental model on the functioning of the system.

Finally, we conducted recall and recognition tasks, one time in an immediate test after the learning phase, and one time in a deferred test one week later.

4.1 Visual exploration

Our first prediction is that the participants in the sequential display group will switch their visual attention less often between the different pictorial elements of the diagram (representational connections) compared to the participants in the static display group. In the sequential condition elements are presented gradually and thus the diagram offers less possibilities to make representational connections in a given time span.

Second, for the same reason as mentioned in the first prediction, we expect less switches of visual attention in the sequential display group between the different textual elements as compared to the static display group (also representational questions).
Third, we expect a higher number of switches between corresponding pictorial and textual elements in the sequential display group as compared to the static display group. In the sequential condition, the participant is guided to make more referential connections, as representational connections are less likely (see first and second prediction).

Finally, we predict an order of visual inspection which is more congruent to the inherent order of the material for participants in the sequential display group compared to participants in the static display group.

4.2 Learning performance

In general, we expect a higher score for retention and troubleshooting questions in the sequential display group compared to the static display group, as the intrinsic cognitive load in the sequential condition is lowered. This assumption is based on cognitive load theory (Sweller, 2005) and its derived implications for sequential displays (Jamet, 2006).

In addition to the general assumption stated above, we make two competitive hypothesis regarding the performance in process retention questions. If the comprehension of temporal processes is fostered by making representational connections, as the learner compares different steps in time, we should expect a decreased score in process retention questions in the sequential condition. If on the contrary, the comprehension of temporal processes is fostered by the “effect of mental organization”, as the elements are provided in a meaningful order (Bétrancourt et al., 2003), we should expect an increased score in process function questions for the sequential display group.

Further there should be more correct pairings of corresponding textual and pictorial elements in a recognition task when participants viewed the sequential display as compared to participants who viewed the static display. This is expected, as the participants in the sequential group make more referential connections.

Another assumption is that the sequential display will lead to a recall order of elements which is congruent to the inherent order of the material, as the chunking of information in the sequential display offers a coherent order of information processing (as has been argued for example in Bétrancourt & Bisseret, 1995). The recall order of elements in the static condition in contrast will tend to be random.

Also the number of recalled elements is expected to be higher in the sequential display group than in the static display group, as the sequential display shows less elements at the same time, facilitating the processing of information.

Our last prediction is that the participants in the sequential display group recall the elements in clusters, which corresponds to the chunking as it was presented in the sequential display, whereas the chunking of elements in the static display group tends to be random. This assumption is based on the findings by Bétrancourt, Bisseret and Faure (2001).
5 Method

5.1 Participants

The participants were thirty-two students recruited at the University of Geneva, of which 62.5 % were women. Their participation in this experiment was voluntary. Eighteen participants were undergraduate students, five were graduate students and nine of them were post-graduate students or PhD students. They came from different faculties of the university. Mean age of the participants was 23.72 (SD = 4.81).

5.2 Design

The experiment consisted of a mixed two-factor design (see Tab 5.1). The first factor was the type of display, which was manipulated between subjects. In one condition, the type of display was static, meaning that the diagram was displayed in its full complexity since the beginning of the learning phase. In the second condition, the type of display was sequential, meaning that the same diagram was constructed progressively, showing the different steps in a meaningful, temporal order. Participants were randomly assigned to one of the two conditions, although we ensured a balanced male-female ratio in both conditions. The second factor was the testing time, which was manipulated within subjects. The participants were tested a first time (immediate) and returned for a second time for a further evaluation (deferred).

5.3 Instrument

The testing configuration consisted of two 17 inches monitors, each one with a resolution of 1280×1024 pixels. The right screen was used to present the web-based instructions and questionnaires. The left screen was a Tobii 1750 corneal reflectance eye tracker, in front of which the participants studied the learning material. The visualization and processing of eye tracking data was made using Clearview 2.71.

<table>
<thead>
<tr>
<th>Type of display</th>
<th>Testing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Immediate N=16</td>
</tr>
<tr>
<td>Sequential</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Experimental design
analysis software from Tobii. According to the user manual of Clearview 2.7\(^1\) the eye tracker has an average accuracy of 0.5 degrees, which corresponds to a margin of error of 0.5 cm at a distance of 50 cm from the screen. This margin of error has been respected in the construction of the material, as different AOIs had at least a distance of 1 cm between them (effective measure on the screen). The eye tracker was operated at a frequency of 50 Hz, which was constant during all recordings. The validity filter for gaze aggregation was set to 100 ms and 30 pixels.

5.4 Material

The learning material consisted of a network map, which explained the different steps in a typical e-voting process. The construction of the material was inspired by an animated presentation\(^2\) of the e-voting system “pnyx” by scyt\(^3\), a system which is also used by the State of Geneva for their e-voting solution\(^4\)\(^5\)\(^6\). The material has been heavily adapted, in order to have a static (not animated) presentation which is suited to novice learners, who do have no expertise in information technology. The learning material presented a diagram with accompanying text explaining the different objects and the different processes which are implicated in the e-voting process (see Fig.5.1). The descriptions were integrated into the diagram in accordance to the spatial contiguity principle (Mayer, 2001). The material was constructed with the vector drawing software inkscape v.0.45.1 and the graphics were exported in the GIF format for further use in Clearview.

The distances in the present computer network map represent temporal-functional relations and they do not provide information about distances in real world, a characteristic which it shares with other materials like electronic circuit maps (Kalyuga et al., 1998). Arrows indicate the flow of information, but the order of the steps could only be extracted from the textual description, in particular by the numeration of processes. Textual information which describe processes are marked by numbers (1–11), whereas descriptions of objects were marked by letters (A–L). Counting all depictive and textual elements of the material, the network map consisted of a total of 35 elements, each element corresponding to an area of interest (AOI).

In the static display group the text-picture diagram was shown all at once. In the sequential display group, the material was segmented in single processes and their related objects. The learner could view additional segments by pressing the space bar. There were nine segments in total, so that when adding the ninth segment the learner viewed the complete diagram, which corresponded to the display shown to the static display group. We are aware that the interaction with an interface can

---

\(^1\)the manual can be found at [http://www.tobii.com/scientific_research/support_downloads/downloads.aspx](http://www.tobii.com/scientific_research/support_downloads/downloads.aspx)


\(^3\)[http://www.scyt1.com](http://www.scyt1.com)


\(^5\)[http://www.epractice.eu/cases/GIVA](http://www.epractice.eu/cases/GIVA)

\(^6\)[http://www.geneve.ch/evoting/](http://www.geneve.ch/evoting/)
Figure 5.1: Learning material showing the static version
have a positive effect on learning performance (Mayer & Chandler, 2001), a variable which we have not kept equal in both conditions. Instead of adding some artificial interaction mechanism to the static version, we have decided to keep both versions as close to a realistic learning material as possible, facilitating the adaption of our findings in real educational scenarios.

Further, a web-based questionnaire was designed. Besides questions on demographic information of the participant, the participant was asked on his experience with voting in general and a self-evaluation of his knowledge of the technical functioning of an e-voting system. Further, fifteen multiple choice questions were elaborated, of which eight could be answered using facts directly derived from the learning material (retention questions). From these eight retention questions, four were of temporal nature (e.g., “what do you receive after sending your ballot paper?”), which are designated as process retention questions, and four questions were without a temporal implication (e.g., “What does the mixing server do?”), which are designated as function retention questions. The remaining seven questions were troubleshooting questions, which demanded a reasoning where the answers had to be deducted from a mental model of the whole system (e.g., “why can’t you vote twice?”) (see all questions in appendix A). The order of the questions was randomized, as well as the order of the answers. The last part of the questionnaire consisted of three questions on the learners evaluation of the material. He could evaluate the interestingness, the learnability and the enjoyability of the material on a six-point Likert-scale.

To evaluate the learners performance in a free recall task, a A4 paper showing only the basic lines of the diagram was prepared. All the pictures and texts were omitted and also the arrowheads were truncated, so that there was no more information on the direction of the processes. The learner could write and draw on the blank schema reconstructing the diagram he has learned.

The same layout of lines was also used for a recognition task, but this time the learner was additionally provided with thirty-five pieces of paper where all the pictorial and textual elements were printed. He could place these pieces of paper on the blank schema similarly to completing a puzzle.

5.5 Procedure

The participants were passed individually in a small room, which provided just enough space for the desk with the apparatus, the experimenter and the participant. The participant was placed in front of the two screens and asked to start with the questionnaire on the right screen. When the learner answered, that he has studied computer sciences, he was asked to specify the kind of studies he has done. As the french word “informatique” is much broader than the English word computer science and includes also courses in software applications for a general public, this kind of specification was necessary. Participants which had an expertise of an engineer in computer sciences would have been excluded from the experiment, but none of the participants fulfilled this criteria. After completing the part on demographic infor-
mation and self-evaluation on the subject of e-voting, participants continued with the fifteen multiple choice questions. The participants were reminded that it will be difficult to answer the questions, but they should choose an answer even if they are not sure and select the answer which seem most probable to them.

After completing the pre-test, the participant was asked to sit in front of the screen with the eye-tracking system. First, the system was calibrated to the participant’s eyes using the built-in calibration tool of Clearview. The participant was then informed that he will see a instruction diagram, which he could learn as long as he wishes to. The participant was instructed to learn the diagram in such a way, that he could afterwards explain the functioning of the e-voting system to a fellow student. When the participant told the experimenter that he has learned enough, he was asked to sit in front of the right screen and to continue with the web-based questionnaire. The same fifteen multiple choice questions were asked again. The order of the questions and the order of the response was again randomized and was therefore different from the first pass of the questionnaire.

The participant could then evaluate the learning material for its interestingness, learnability and enjoyability.

As a next step, the participant received a paper in format A4 with a blank schema and a pencil and he was asked to recall as many objects and processes as possible and to write them down. He was informed that it is not necessary to make complete phrases and he could as well draw images if he wishes to. When the participant had finished, he turned the page where he had the same blank schema as before. He received the envelope with thirty-five pieces of papers with all the objects and texts he has previously seen in the learning material. He was instructed to put the pieces of paper on the blank schema, in order to reconstruct the schema previously learned.

After completion, the participant was thanked for his participation and the date for the second part of the experiment was fixed. He was kept naive on the tasks he would have to perform next time.

Approximately one week later the participant came by for the second part of the experiment. He was provided with a double-sided A4 paper with the two blank schemas and was asked to do the free recall and the recognition test as he has already done in the first pass of the experiment.

If the participant so wished, he was informed afterwards about the research questions of the experiment and he could view the visualization of his scanpath, which were exported from Clearview.

5.6 Independent variables

- type of display (static, sequential)
- testing time (immediate, deferred)
5.7 Dependent variables

5.7.1 Spatial questions

• Number of elements (textual and pictorial) written or drawn on a blank schema (free recall). Textual or pictorial elements were counted, when they could be assigned to one of the elements in the original diagram. This variable was measured in the immediate and the deferred recall task.

• Order of elements (textual and pictorial) written or drawn on a blank schema. The effective order is compared to the theoretical order using the Levenshtein distance algorithm (Levenshtein, 1966). Further a hierarchical cluster analysis is conducted to evaluate similarities in the chunking of the elements. This variable was measured in the immediate and the deferred recall task.

• Number of correctly paired elements (image and corresponding text) in the recognition task.

5.7.2 Temporal questions

• Gain in score in process retention questions between pre- and post-test (4 multiple choice questions)

• Decrease of correctly recalled elements in the recall tasks between the immediate and deferred free recall test

5.7.3 Functional questions

• Gain in score in function retention questions between pre- and post-test (4 multiple choices questions)

• Gain in score in troubleshooting questions between pre- and post-test (7 multiple choices questions)

5.7.4 Eye movement data

• Number of switches between pictorial elements (representational connections)

• Number of switches between textual elements (representational connections)

• Number of switches between related textual and pictorial elements (referential connections)

• The inspection order of textual elements

• The inspection order of pictorial elements
6 Results

6.1 Data processing

The measuring data of the present experiment came from four different sources. First all answers of the participants given in the interaction with the web interface, as well as the time they spent on the different parts of the questionnaire, were collected and stored in XML Files (you find them in the directory results/questions/ of the accompanying DVD). The web interface was written in PHP 5 and can also be found on the DVD (see webinterface/). The second source of information was the raw data of eye tracking data provided by the Clearview software (see results/rawdata/), where information about gazes and gazes per AOI are stored. AOIs were defined before in the Clearview software, marking the thirty-five textual and pictorial elements with surrounding rectangles, overlapping the element by approximately five pixels. Third, gaze plots of each participant were exported using the “Techsmith Screen Capture Codec” (see results/scanpath/). Finally, each participant filled out a A4 sheet when he completed the recall and the recognition task (one time in the immediate pass, a second time in the deferred pass of the test) (see results/sheets/).

For data aggregation we used the script language Perl version 5.8.8 as it is powerful in processing text files. We used Perl scripts among other reasons to extract a gaze order in the raw data files provided by Clearview or calculating the Levenshtein distance between an optimal and the effective order (see analysis/ for all scripts used in this experiment). For statistical calculations we used SPSS version 16.0.1 for the creation of diagrams we used R version 2.6.2 (R Project for Statistical Computing) including standard packages. For the hierarchical clustering with probability values we have additionally imported the pvclust package.

For all parametrical tests (in particular all analyses of variances), the heterogeneity of variances was analyzed using the Levene’s test. If not otherwise stated, analyses passed Levene’s test.

6.2 Eye movements

The measurements of referential connections confirmed our hypothesis. In the sequential display group we could determine more referential connections (M = 24.25,
Table 6.1: Mean numbers and standard deviation of referential and representational connections. Mean Levenshtein distance and standard deviation for inspection of textual and pictorial elements.

<table>
<thead>
<tr>
<th>Group</th>
<th>Referential connections</th>
<th>Representational connections</th>
<th>Levenshtein distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(text)</td>
<td>(pictures)</td>
<td>(text)</td>
</tr>
<tr>
<td>Static</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Sequential</td>
<td>16.00</td>
<td>9.85</td>
<td>3.62</td>
</tr>
</tbody>
</table>

SD = 8.40) compared to the static display group (M = 16.00, SD = 9.85) which proved to be significant (F_{1,30} = 6.490, p < 0.05).

The number of representational connections between textual elements seem to be similar in the two conditions, as participants in the static condition made an average of 3.62 representational connections (SD = 2.82) and participants in the sequential condition made an average number of 3.31 representational connections between textual elements (SD = 2.08). The difference was not significant (F_{1,30} = 0.127, ns).

The results of the analysis of representational connections between pictorial elements goes in the same direction. The difference of representational connections between the sequential condition (M = 0.62, SD = 0.81) and the static condition (M = 0.19, SD = 0.40) is not significant (U = 91.500, ns). The variances in numbers of representational connections was not sufficiently homogeneous between the two conditions (F_{1,30} = 13.646, p < 0.001), to apply an analysis of variance.

The hypothesis that there are more representational connections in the static display group compared to the sequential display group has to be clearly rejected.

Regarding the order of inspection, we distinguished between the order of textual elements and pictorial elements, as mentioned before. In the sequential condition, the participants followed the inherent order of the textual elements as indicated by Levenshteins distance (M = 42.12, SD = 18.47) similarly to the participants in the static condition (M = 42.69, SD = 25.19). This difference was not significant (F_{1,30} = 0.005, ns). A similar result can be found for the Levenshtein distance and pictorial elements, where in the sequential condition a mean Levenshtein distance of 22.69 (SD = 6.94), respectively 21.00 (SD = 13.40) in the static condition could be found. A Mann-Whitney test was applied and showed that this difference is also not significant (U = 116.00, ns), the test of Levene indicating a heterogeneity of variances (F_{1,30} = 8.371, p < 0.01). A table of all measurements of eye movement indicators is given in Table 6.1.

The hypothesis that the participants in the sequential display group follow the order of elements more congruently to the inherent order of the material than the participants in static display group has to be rejected.
6.3 Learning time

The learning time was not constrained in this experiment and first a comparison of learning time was made between the two conditions. If a significant difference of learning time would exist between the groups, a higher performance could be attributed to a longer study time. The analysis showed that the participants passed an equal amount of time with the learning material. Participants in the static group learned the material for an average of 256 seconds (SD = 93 s) and the participants in the sequential group learned the material for an average of 247 s (SD = 79 s). Apparently the type of display had no effect on study time ($F_{1,30} = 0.085$, ns). Similarly, participants in the static condition did not pass more time ($M = 328$ s, SD = 86.52 s) on the questionnaire in the post-test than the participants in the sequential condition ($M = 317$ s, SD = 75 s) and the difference is indeed not significative ($F_{1,30} = 0.944$, ns).

6.4 Learning performance

For all tests regarding the learning performance measured by the questionnaires, we calculated the mean gain of score between the pre-test and the post-test individually for each type of questions (see Tab 6.2 or Fig 6.1 for an overview).

Regarding function retention questions, a small advantage of the participants in the sequential group could be found ($M = 1.69$, SD = 1.14) compared to the participants in the static group ($M = 1.44$, SD = 1.03). This difference is not significant ($F_{1,30} = 0.424$, ns). When the question implicated a temporal aspect, as for the process retention questions, the participants in the sequential group ($M = 2.31$, SD = 1.01) outperformed the participants in the static display group clearly ($M = 0.75$, SD = 1.44), a difference which proved to be highly significant ($F_{1,30} = 12.618$, $p < 0.001$). When these type of questions are combined to retention questions, the difference between the sequential group ($M = 4.00$, SD = 1.51) and static group ($M = 2.19$, SD = 1.76) was still significant ($F_{1,30} = 9.802$, $p < 0.01$).

In contrast, the gain of score in troubleshooting questions was minimal. The participants in the sequential group had a mean gain of $1.31$ in score (SD = 1.74), whereas the participants in the static group had a mean gain of $0.44$ (SD = 2.03) in score for troubleshooting questions. An analysis of variance showed that this difference between the groups was not significative ($F_{1,30} = 1.711$, ns). Combining all questions in a total score, the difference between gain in score for the sequential group ($M = 5.31$, SD = 2.39) and static group ($M = 2.62$, SD = 2.36) was significant ($F_{1,30} = 10.246$, $p < 0.01$) again.

These results let us reject the hypothesis, that a static display could foster the learning of process retention questions, as more representational connections are possible. First, as we have seen in the previous section, there were not significantly more representational connections made in the static display group. Second, the number of representational connections between textual elements seem to be detrimental for
the gain in score in process retention questions \( r = -0.433, p < 0.05 \).

### 6.5 Recall

For the analysis of the recall tasks, we applied three different measures. First, we counted the number of recalled elements both in immediate test and deferred test and for both recalled processes and recalled elements. Second, we measured the Levenshtein distance \(^1\) (Levenshtein, 1966) between the effective order of recall and the optimal order of recall, which corresponds to the steps provided in the material. Finally, we did a hierarchical clustering analysis of the effective recall order to detect similar clusterings of recalled objects.

In order to assign written or drawn elements to a corresponding element in the original diagram, we used following rules:

- A picture was mapped to the corresponding textual description element (recalled object), when the picture was clearly corresponding to it.
- A textual description was mapped to a recalled process, if there was a verb included in the description or if it referred otherwise to an action.
- If the text described no action, but only described an element, it was mapped to the recalled object.

The accuracy of the answers decreased in general between the immediate and deferred recall task. An element was counted, if it could be mapped at best guess to a corresponding element in the original diagram, unless it contained wrong statements.

#### 6.5.1 Recall performance

In the immediate recall task, participants of the sequential display group recalled more elements \( (M = 11.75, SD = 2.46) \) compared to the participants in the static display group \( (M = 8.94, SD = 3.61) \). An analysis of variance indicate that the difference in recall performance is significative \( (F_{1,30} = 6.639, p < 0.05) \). When the recalled elements where split up into recalled processes and recalled objects, we find for both types of elements an advantage for the sequential display group. We found a mean number of 3.00 recalled processes \( (SD = 1.71) \) in the sequential group compared to a mean number of 2.75 recalled processes \( (SD = 2.24) \) in the static group and a

---

\(^1\) The Levenshtein distance is a measure of the minimum number of single-character edits (insertions, deletions, or substitutions) required to change one word into the other word. It is used here to quantify the difference between the order of recall and the optimal order of recall.
Figure 6.1: Mean gain in questions between pre- and post-test (Func: Function retention, Proc: Process retention, Ret: Retention (function & process), TS: Troubleshooting, Tot: Total, St: Static condition, Sq: Sequential condition)
mean number of 8.75 recalled objects (SD = 1.84) in the sequential group compared to a mean number of 6.19 of recalled objects in the static group (SD = 3.31). Only the number of recalled objects seem to be significant in the immediate test (U = 68.000, p < 0.05), the difference in recall of processes fails to be significant (F<sub>1,30</sub> = 0.126, ns). A Whitney-Mann test had to be applied for the recalled objects, as a significant heterogeneity between the variances of recalled objects was found (F<sub>1,30</sub> = 9.397, p < 0.01).

In the deferred recall task, participants tended to recall in total more objects as in the immediate recall task. A mean number of 12.12 recalled elements in the sequential display (SD = 3.07) compared to a mean number of 9.19 of recalled elements in the static display (SD = 3.12) could be measured. This difference is significative (F<sub>1,30</sub> = 7.186, p < 0.05). On a qualitative level, it has been assessed that, even if the number of elements increased, the acuity (details in the descriptions) has decreased between the immediate and deferred recall tasks. If we look at the differences more in detail, we see that the difference for the recalled processes got even smaller, as the recalled processes are on average 3.31 in the sequential group (SD = 2.94) and 3.25 in the static group (SD = 2.49), a small difference which proves to be not significative (F<sub>1,30</sub> = 0.004, ns). The difference of recalled objects however increased over time. Participants in the sequential group recalled 8.81 objects (SD = 1.91), whereas participants in the static group recalled only 5.94 objects (SD = 2.35). This difference is highly significative F<sub>1,30</sub> = 14.440, p < 0.001.

Our hypothesis that the participants in the sequential display group recall more elements than the participants in the static display group has been corroborated. For an overview of the recall performance refer to Tab 6.3.

### 6.5.2 Recall order

The recall order was evaluated by the Levenshtein distance, which becomes smaller the more similar the effective recall order is to an optimal (theoretical) recall order.

When including both types of elements, the recall order in the immediate recall task reveals a recall order which is slightly more congruent in the static condition than in the sequential condition. The recall order of the participants in the static condition is characterised by a mean Levenshtein distance of 13.25 (SD = 3.97), whereas participants in the sequential group recalled the elements with an order with a mean Levenshtein distance of 14.69 (SD = 3.17). This would be in contrast to our hypothesis, but the difference proves to be not significant (F<sub>1,30</sub> = 1.277, ns).
When we look more in detail at the two types of elements in the recall task, we find that the participants in the static group recall the processes more congruently to the order of presentation (M = 5.06, SD = 3.71) than participants in the sequential group (M = 8.00, SD = 2.36). This difference of the Levenshtein distance is significant (U = 54.500, p < 0.01). When we look at the difference in recall order of objects, we have a contrary situation. Participants in the sequential group follow the theoretical order more likely (M = 6.69, SD = 2.06) than the participants in the static group (M = 8.19, SD = 2.50). As well, this difference of the Levenshtein distance is marginally significative (F\textsubscript{1,30} = 3.420, p = 0.074).

In the deferred recall test, we can observe a change in the recall order, which results in a general advantage of the sequential group compared to the static group. The difference in the combined Levenshtein distances is marginally significant for the sequential group (M = 13.81, SD = 4.04) and the static group (M = 16.31, SD = 2.98), (F\textsubscript{1,30} = 3.970, p = 0.055). In detail, we can see that the disadvantage for the sequential display group in the immediate task for process questions turned into a small advantage in the deferred test. The participants in the sequential group had a mean Levenshtein distance of 6.25 (SD = 3.61) compared to the participants in the static group which recalled elements with a mean Levenshtein distance to the optimal order of 7.56 (SD = 2.78). However, this advantage in the Levenshtein distance is very likely by chance (F\textsubscript{1,30} = 0.004, ns). In contrast, we can see that participants in the sequential group (M = 7.56, SD = 1.46) outperform clearly the participants in the static group (M = 8.75, SD = 1.18) when it comes to the recall order of objects, (F\textsubscript{1,30} = 14.440, p < 0.001).

The hypothesis that the recall order evaluated by the Levenshtein distance corresponds more likely in the sequential display group can only be partially corroborated. It seems that the participants in the sequential order respect mainly the order of objects and deviate more likely from the optimal order, when it comes to the process elements. Reciprocally, the participants in the static group tend to recall the process elements more congruently to the optimal recall order, recalling the elements in a more random order. An tabular overview of the descriptive statistics can be found in Tab 6.4.
Figure 6.2: Mean Levenshtein distance and confidence interval (95%) for different types of recalled elements (Imm: Immediate recall, Def: Deferred recall, Proc: Process element, Obj: Object element, Tot: Total (processes and objects), St: static condition, Sq: sequential condition)
6.5.3 Clustering in recall task

For the identification of clusters, we used a hierarchical cluster analysis, which measured the euclidean distance of the dissimilarities and an average linkage method has been applied. Approximately unbiased (AU) probability values were determined, which were computed by multiscale bootstrap resampling (bootstrap sample size was set to 10000). When we analyzed the order of recall of elements we encountered the problem, that a cluster analysis does not allow missing values for its computation (see [Wagstaff, 2004 for a problem description and some advices on solving the problem]). The participants did not recall all elements and thus we had to complete the data basis to effectuate the analysis. We choose to fill the missing values with random numbers, which do corresponds to the null hypothesis. When a participant recalled five elements, we assigned an order from six to twenty-three to the not recalled elements in a randomized order (randomization was done by the randomization function in OpenOffice version 2.4.0).

When analyzing the result of the hierarchical cluster analysis for the static condition (see Fig 6.3) and for the sequential condition (see Fig 6.4) we can make several statements. In both conditions the number of significant clusterings decreased between the immediate pass and the deferred pass of the recall task (clusters, which have an error $\alpha$ below 5 % are marked with a surrounding rectangle). In the static display group, the number of recalled elements which could be clustered with a high certitude decreased from 9 elements in 3 clusters to 7 elements in 3 clusters ($p < 0.05$). In the sequential display group the number of significant clusters dropped from 15 elements in 1 cluster to 8 elements in 2 clusters. If we take clustered elements as indicator for chunking of information, we can assume that participants in the sequential group have a higher degree of chunking of the information in the immediate recall task than the participants in the static group, but this chunking drops in both conditions to a similar level.

However, if the clustering in the recall tasks corresponds to the chunking of information in the sequential display is difficult to evaluate. We will give an explanation in the discussion of the present paper.

6.6 Recognition

The analysis of the recognition task reveals a clear advantage of the sequential display group over the static display group. In the immediate recognition task, participants in the sequential display group paired a mean number of 9.56 (SD = 1.36) elements correctly, whereas the static display group paired a mean number of 6.06 (SD = 3.60) correctly. In the recognition test one week later, the participants in the sequential

Please refer to the documentation of the pvclust package for a clue, what that could mean: [http://www.is.titech.ac.jp/~shimo/prog/pvclust/](http://www.is.titech.ac.jp/~shimo/prog/pvclust/)
Figure 6.3: Hierarchical clustering of recall order in the static condition, AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) values.
Figure 6.4: Hierarchical clustering of recall order in the sequential condition, AU (Approximately Unbiased) p-value and BP (Bootstrap Probability) values.
Table 6.5: Mean number and standard deviation of pairings of corresponding textual and pictorial elements in the recognition task.

<table>
<thead>
<tr>
<th>Group</th>
<th>Immediate recognition test</th>
<th>Deferred recognition test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Static</td>
<td>6.06</td>
<td>3.60</td>
</tr>
<tr>
<td>Sequential</td>
<td>9.56</td>
<td>1.36</td>
</tr>
</tbody>
</table>

Table 6.6: Mean score and standard deviation on a six-point Likert-scale (1–6) in the subjective evaluation of the material for its interestingness, its learnability and its enjoyability.

<table>
<thead>
<tr>
<th>Group</th>
<th>Interestingness</th>
<th>Learnability</th>
<th>enjoyability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Static</td>
<td>4.06</td>
<td>1.18</td>
<td>3.50</td>
</tr>
<tr>
<td>Sequential</td>
<td>4.69</td>
<td>0.79</td>
<td>3.75</td>
</tr>
</tbody>
</table>

The hypothesis, that the number of correct pairings of textual and pictorial elements are higher in the sequential display group than in the static display group has been confirmed. For an overview of the results regarding the recognition task please see Tab 6.5.

6.7 Subjective evaluation of the material

The results of the subjective evaluation showed that participants favoured the sequential presentation, but the differences were not significant. In detail, the participants in the sequential display group rated the interestingness of the material with an average of 4.69 (SD = 0.79) on a six-point Likert-scale from 1 to 6. The participants in the static display group gave an average of 4.06 points (SD = 1.18). This difference is marginally significant ($t_{30} = 1.757$, $p = 0.089$, two-tailed). Regarding the learnability of the material, the participants in the sequential display group rated it higher (M = 3.75, SD = 1.29) than their colleagues in the static display group (M = 3.50, SD = 1.21), but this difference was not significant ($t_{30} = 0.565$, ns, two-tailed). Finally, participants in the sequential group seem to have enjoyed the material more (M = 4.31, SD = 0.95) than the participants in the static group (M = 3.75, SD = 1.48), but not on a significant level ($t_{30} = 1.279$, ns, two-tailed). For an overview of the descriptive data refer to Tab 6.6.
7 Discussion

7.1 Visual exploration

We remember that our hypothesis regarding representational switches of visual attention was “that the participants in the sequential display group will switch their visual attention less often between the different pictorial elements of the diagram (representational connections) compared to the participants in the static display group.” and that “we expect less switches of visual attention in the sequential display group between the different textual elements as compared to the static display group (also representational questions).”

For referential switches we expected “a higher number of switches between corresponding pictorial and textual elements in the sequential display group as compared to the static display group.”

The eye movement data in the present experiment provided mixed results. On the one hand we could demonstrate that there is a significant difference in referential connections between the static display group and the sequential display group, revealing a higher number of referential connections in the sequential group. On the other hand, we could not demonstrate the anticipated higher representational connections in the static display group. The number of representational connections proved to be very low with the method we applied to measure them. An explanation for this fact could be that the target of a saccade when moving to a next step in the diagram is usually a pictorial element. This could be due to the simple fact that pictorial elements are more salient than textual elements. However, in order to understand the process in a single step of the diagram, the learner preferably ends the exploration of the step in reading the text. This could explain the unexpected low number of representational connections, as a typical visual inspection pattern would go from pictorial element to a corresponding textual element and then again to the next pictorial element in successive steps through the learning material. This explanation is purely hypothetical and would have to be explored in a further experiment, but a rather superficial qualitative analysis of the scanpath movies provided by the eye tracker seem to give some evidence for this hypothesis.

If this finding could be confirmed, it would be in contrast with the inspection strategy in diagrams where text and diagram are separated. For a separated presentation of a diagram explaining a pulley-system, Hegarty and Just (1993) showed that diagram inspection is largely text guided.

In our hypotheses we predicted “an order of visual inspection which is more congruent to the inherent order of the material for participants in the sequential display group compared to participants in the static display group.”
The analysis of the inspection order failed to reveal any significant difference between the static and sequential display group. Again, this could be due to the reason mentioned above. In the inspection we distinguished order between a pictorial path and a textual path, because both paths provide an alternative route of processing the same information. This means that we measured the order of representational connections, which were very low for the reasons mentioned before.

The notion of representational and referential connections is used in different theories (Mayer & Anderson, 1991; Narayanan & Hegarty, 1998), but it has never been tried to demonstrate this kind of visual integration on a perceptual level. Rather, they were used to designate processes on higher cognitive levels, in particular to explain the construction of a mental model from different sources of information, as for example proposed in the model of Narayanan and Hegarty (1998; 2002). Researchers in the field of multimedia learning used preferably the distinction between local and global inspections (Hegarty & Just, 1989; Hegarty & Just, 1993; Schneider, 2007), which seem to be especially adequate to demonstrate integration processes in the learning process, an important aspect of visual inspection we have completely omitted in the present study, which should be made up for in a further study with the same material.

It seems difficult to make direct conclusions from the number of referential connections to a learning outcome. In the type of material we have provided, it is not essential to integrate the textual and pictorial information. The textual descriptions could have sufficed to understand the process of e-voting, as the pictures did not reveal new information which was not given in the text. Still, the pictorial elements probably supported the understanding as they provide an additional form of representation according to the multimedia principle (Mayer, 2005). The more essential information were probably the connections of the different elements by lines, which conveyed the functional, temporal relations between the elements. Lines would be difficult to analyze with eye tracking devices as they are practically one-dimensional.

### 7.2 Learning performance

In our hypotheses we expected in general “a higher score for retention and troubleshooting questions in the sequential display group compared to the static display group”.

The results in score for the retention questions confirmed our hypothesis: the participants in the sequential group were in general superior to the participants in the static group. Regarding the process retention questions, we had two competitive hypotheses, either the comprehension of temporal processes is fostered by making representational connections or it is rather fostered by an effect of mental organization induced by the meaningful chunking of information. It is clearly the second hypothesis who proved to be true: it is the chunking of information which foster the understanding of temporal processes as proposed by earlier studies in this field (Bétrancourt & Bisseret, 1995). Regarding representational connections, we have even
shown that there is a negative correlation between making representational connections and the performance in process retention questions in the subsequent post-test. The presence of a correlation does not have to mean that these two factors are causally linked, but we can assume with high certitude that representational connection do not have a positive influence on the comprehension of temporal processes.

The learning material seemed not to help the learners in answering the troubleshooting questions. The gain in score between pre- and post-test was quite low. Comparing the two conditions, no significant difference in gain of score could be found for these troubleshooting questions, which demanded a deduction from the information given in the material to correctly reply to them. The ability of responding to troubleshooting questions, or inference question as they are called more generally, demand a preliminary construction of an integrated mental model (Mayer 2001; Schnotz 2005; Narayanan & Hegarty 2002), which can be mentally animated, in order to see the effects of different actions. In the case of a network map we do not have physical motions, but we have a flow of information in a metaphorical sense (Narayanan & Hegarty 2002).

We could make several conclusions from the fact that there was no difference between the two conditions regarding troubleshooting questions. Either the factor type of display had no influence on the construction of a mental model or it could even be argued that a sequential presentation is detrimental for the construction of a mental model, as it “might have reinforced the individual processing of the elements at the expense of a more global processing of the diagram”. It is also possible that the chosen troubleshooting questions were a bad indicator for a successful construction of a mental model. In the case of processes which are linked in a functional-temporal way, the answering of troubleshooting questions could also depend on a general ability of making logical deductions.

7.3 Recall performance and recall order

In our hypotheses, we made the assumption that “the number of recalled elements is expected to be higher in the sequential display group than in the static display group”.

Further we made the assumption that “the sequential display will lead to a recall order of elements which is congruent to the inherent order of the material, as the chunking of information in the sequential display offers a coherent order of information processing. The recall order of elements in the static condition in contrast will tend to be random.”

Participants in the sequential group recalled in general more elements than the participants in the static display group, a finding which is congruent with our hypothesis. An interesting result was discovered when the recall of process elements and object elements was distinguished in the analysis. Only the number of recalled elements was significantly higher in the sequential group, but not the recall of process elements.
An explanation for this finding could lie in the fact that in the sequential display the process information is to some extent already inherent, resulting in an “effect of order”, which makes it less essential for the learner to recall the processes. When for example a process in the diagram describes the action “sending”, the information could be encoded visually by the new appearance of an arrow.

It is a vague hypothesis, but it could explain also the other interesting finding concerning the difference in the recall order of process elements and object elements. Participants in the static condition recalled the process elements in an order which was significantly more similar to the original steps as the order in which the participants in the sequential display group recalled them. And inversely, participants in the sequential display group recalled the object elements in order which was significantly more congruent to the original diagram than was the recall order of the participants in the static display group.

**7.4 Clustering in recall task**

Our prediction regarding clustering in the recall task was “that the participants in the sequential display group recall the elements in clusters, which corresponds to the chunking as it was presented in the sequential display, whereas the chunking of elements in the static display group tends to be random.”

The hierarchical cluster analysis provides an interesting indicator of how the sequential display fosters clustering of elements in the mental representation of the learner. The participants in the sequential display seem to cluster more elements in the immediate recall task than their fellow students in the static display group. However, this advantage seems to be lost one week later as the number of elements which are chunked with a high certainty becomes equal in both groups.

The comparison of the clustering in the recall task to the chunking of information in the sequential display seem difficult, when reviewing the results. In contrast to previous studies with hierarchical cluster analysis of sequential presentations (Bétrancourt & Tversky submitted), where this kind of correspondence could be identified, the present material is characterized by a high degree of element interaction and typically there were only three elements per cluster. The last element of one cluster was typically the first element of the successive cluster, which makes it difficult to identify a natural way of grouping the information in clusters. The ordering in a sequence seems more appropriate to evaluate the mental organization of information (as done by measuring the Levenshtein distance), than measuring the clusters.

**7.5 Recognition**

We expected in our hypotheses that “there would be more correct pairings of corresponding textual and pictorial elements in a recognition task when participants viewed the sequential display as compared to participants who viewed the static display.”
Indeed, we found that the participants in the sequential display group paired significantly more corresponding textual and pictorial elements than the participants in the static display group. As we have shown, the sequential display has a positive effect on the number of referential connections, which has in turn, according to our hypothesis, a positive effect on the recognition performance.

7.6 Subjective evaluation of the material

The results of the subjective evaluation do not provide us with new findings. In general, there was no significative difference between the sequential display group and the static display group regarding the evaluation of the material. Still, we could find a marginally significant difference of interestingness in favour of the sequential group. Other studies (Jamet et al., 2008) have also failed to show any effect of the segmented display on the subjective presentation, earlier studies (Bétrancourt et al., 2003) even showed a negative effect of sequential displays on the user evaluation (called “pedagogical value”).
8 Conclusion

The present paper could confirm some of the positive effects of sequential presentations. Notably, it has been demonstrated again, that the sequential presentation of information seem to foster the retention of the material. The experiment also revealed that there is an effect of a meaningful segmentation of information on the mental organization of the learned material in the learner’s mental model.

However, the analysis of the eye tracking data did confront us with some problems. First, representational connections, as we have measured them, were a bad indicator for the integration of different steps of the learning material. Nevertheless, the low number of representational connections opened new questions of how we explore an integrated display of the type of material we have provided. It would be interesting to analyze the integration of the elements more in detail by reinspecting the gaze-paths of the learners.

Second, the inspection order was little congruent when using our operalization of how an effective visual inspection order can be compared to the optimal inspection order, which was given by the sequence of processes in the diagram. A short review of the scanpaths of the participants would suggest that the participants did follow the given order in the diagram, especially in the sequential display condition. The data should be reevaluated and a more suitable operalization of visual inspection order should be applied.

Another approach which should be taken into consideration, is the analysis of visual inspections in terms of local and global inspections as proposed by several studies (Hegarty & Just, 1989; Schneider, 2007). It would additionally provide some evidence on integrative processes, which could be identified in evaluating the scanpath in a superficial review of the material. The analysis of global inspections could also give us some empirical data in order to investigate the lack of gain in performance for troubleshooting questions in the sequential display.

Some questions have been answered, some new questions are waiting for an answer.
Bibliography


Wipfli, R. (2007, August). Eye tracking and its application in educational multimedia. (Course work for the University of Geneva.)


Acknowledgements

Merci à Mireille Bétrancourt pour tous ses conseils et pour m’avoir introduit à ce domaine de recherche intéressant et ses chercheurs sympathiques.

A Testing material

A.1 Pretest on prior knowledge

– Êtes-vous déjà aller élire en Suisse ou dans un autre pays ?
– Avez-vous déjà voté en Suisse ou dans un autre Pays ?
– Si vous avez déjà donné votre vote, comment avez-vous le fait (plusieurs choix possible) ?
  – à l’urne
  – par courrier
  – par internet
  – par téléphone portable
– Comment jugez-vous vos connaissances du fonctionnement technique de l’e-voting ?

A.2 Pre- and posttest questions

A.2.1 Function questions

Non-temporal questions

Quels sont les documents que le serveur de vote vous envoie ?
– un bulletin de vote vide et une quittance de vote confirmée
– un login personnel et une enveloppe cryptée
– une enveloppe cryptée avec bulletin de vote vide
– une copie confirmée de vote

Qu’est-ce que les membres de la chancellerie font pour voir le résultat du vote ?
– Ils utilisent leurs mots de passe conjointement pour ouvrir les enveloppes cryptées.
– Ils assignent un membre de la chancellerie à ouvrir toutes les enveloppes cryptées.
– Ils séparent tous les bulletins de vote des quittances de vote.
– Ils m’envoient une quittance de vote.

Comment accédez-vous au serveur de vote ?
– par internet
– par un réseau “peer-to-peer”
– par un fil téléphonique sécurisé
– toutes les trois réponses sont vraies

Quelle est la fonction du serveur mélangeur ?
Il sépare le vote des données personnelles qui figurent sur la quittance de vote et il mélange toutes les données
Il mélange le réseau internet avec le réseau de la chancellerie
Il mélange seulement les données personnelles.
Il sépare les données personnelles des clés des membres de la chancellerie, pour avoir plus de confidentialité

temporal questions

Qu’est-ce que vous envoyez après avoir rempli votre bulletin de vote?
- une enveloppe cryptée avec bulletin de vote rempli et une quittance de vote préparée
- mon mot de passe et mon bulletin de vote rempli
- une quittance de vote confirmée
- un bulletin vide et personnalisé avec une quittance de vote confirmée

Quels sont les documents disponibles à la fin de la procédure d’e-voting?
- Tous les bulletins de vote remplis, toutes les quittances de vote et une quittance de vote confirmée pour chaque votant
- Tous les bulletins de vote remplis et toutes les quittances de vote
- Une quittance de vote confirmée pour chaque votant
- Un bulletin de vote et une quittance de vote

Qu’est-ce qui se passe quand la votation est terminée?
- toutes les enveloppes cryptées sont envoyées au serveur mélangeur
- tous les bulletins de vote sont envoyés à la chancellerie
- tous les bulletins de vote sont publiés
- une enveloppe cryptée est envoyée au serveur de vote

Qu’est-ce que vous recevez après que vous être identifié auprès du serveur de vote?
- un bulletin de vote vide et personnalisé
- un mot passe pour accéder à mon bulletin de vote
- une lettre avec ma carte de vote
- une quittance qui confirme que j’ai voté avec succès

A.2.2 Troubleshooting questions

La connexion entre serveur de vote et serveur mélangeur est découpée un jour avant la fin de la votation. Quelle est la conséquence?
- Aucune conséquence. La connexion est seulement utilisée le jour de la fin de la votation.
- On ne peut plus voter.
- Une partie de votes sera perdue, lorsque la connexion découpée fait perdre des données.
- Il n’y a pas de connexion entre le serveur de vote et le serveur mélangeur.

Comment pouvez-vous vérifier si votre vote était compté?
- Je consulte les quittances de vote publiées.
– Je dois demander quelqu’un de la chancellerie qui a une clé pour ouvrir mon enveloppe cryptée.
– Il n’y a pas de possibilité de le vérifier.
– Je peux vérifier avec ma copie confirmée de la quittance.

Qu’est-ce que vous pouvez faire si votre vote n’était pas compté ?
– Je peux montrer ma copie de quittance de vote à la chancellerie.
– Je ne peux rien faire.
– Je peux répéter mon vote.
– Je donne mes données personnelles à la chancellerie pour qu’ils puissent retrouver mon vote.

Un pirate informatique peut accéder à toutes les données du serveur de vote, mais il ne connaît pas de mot de passe. Qu’est-ce qu’il peut faire ?
– Effacer les enveloppes cryptées avec les votes dedans
– Ouvrir les enveloppes cryptées et changer les votes
– Ouvrir les enveloppes cryptées et voir les votes
– Il ne peut rien faire.

Pourquoi ne pouvez-vous pas voter deux fois pour la même votation ?
– Lorsque je me suis identifié au serveur de vote, le serveur de vote ne m’envoie pas un deuxième bulletin de vote à remplir
– Parce que j’ai déjà une quittance sur mon ordinateur. Je ne peux pas en avoir plus qu’une seule.
– On verra sur les quittances de vote publiées que j’ai voté deux fois.
– Il n’y a pas de moyen de vérifier si je vote pour la deuxième fois.

Pourquoi la chancellerie peut voir ce que vous avez voté ?
– Il n’y a pas de moyen pour la chancellerie de voir ce que j’ai voté.
– La chancellerie peut voir ce que j’ai voté, si tous les membres ouvrent mon enveloppe cryptée avec leurs mots de passe.
– Chaque membre de la chancellerie peut voir ce que j’ai voté, s’il a le mot de passe nécessaire.
– Lorsque les quittances de vote sont publiées, tout le monde peut voir ce que j’ai voté.

Après que vous vous êtes identifié au serveur de vote, votre ordinateur a une panne. On vous assure que vous pouvez refaire votre vote. Pourquoi ?
– Le serveur de vote n’a pas encore reçu mon bulletin de vote rempli.
– Le serveur de vote a déjà vérifié mon mot de passe.
– Le serveur de vote m’a déjà envoyé le bulletin de vote.
– Le serveur mélangeur n’a pas encore mélangé mon vote.
B Program code

Listing B.1: Perl script for analyzing representational and referential connections in AOI gaze data provided by Tobii Clearview

```perl
#!/usr/bin/perl

# the script analyzeTransitions.pl reads the text output of the Tobii eye tracker in ./results/raw_data/*AOI.txt and analyzes the sequence of AOI fixations. The result is written in XML files, where the number of and Levenshtein distance of referential and representational (text, image) are stored.

use strict;
use IO::File;
use XML::Writer;
use Data::Dumper;

# define referential connections
my %referential = {
    'text_A' => 'image_A',
    'text_B' => 'image_B',
    'text_C' => 'image_C',
    'text_D' => 'image_D',
    'text_E' => 'image_E',
    'text_F' => 'image_F',
    'text_G' => 'image_G',
    'text_H' => 'image_H',
    'text_I' => 'image_I',
    'text_J' => 'image_J',
    'text_K' => 'image_K',
    'text_L' => 'image_L',
    'text_1' => 'image_A',
    'text_1' => 'image_C',
    'text_2' => 'image_B',
    'text_2' => 'image_C',
    'text_3' => 'image_C',
    'text_3' => 'image_D',
    'text_4' => 'image_D',
    'text_5' => 'image_A',
    'text_5' => 'image_E',
    'text_6' => 'image_E',
    'text_6' => 'image_C',
    'text_7' => 'image_C',
    'text_7' => 'image_G',
    'text_8' => 'image_F',
    'text_8' => 'image_G',
    'text_9' => 'image_H',
    'text_10' => 'image_I',
    'text_10' => 'image_G',
    'text_10' => 'image_J',
    'text_11' => 'image_K',
}

# the following relation is minimal, we could add more

75
```
# define representational connection between textual elements
my %representationalText = {
    'text_1' => 'text_2',
    'text_1' => 'text_3',
    'text_1' => 'text_4',
    'text_1' => 'text_5',
    'text_1' => 'text_6',
    'text_1' => 'text_7',
    'text_1' => 'text_8',
    'text_1' => 'text_9',
    'text_1' => 'text_10',
    'text_1' => 'text_11',
    'text_2' => 'text_3',
    'text_2' => 'text_4',
    'text_2' => 'text_5',
    'text_2' => 'text_6',
    'text_2' => 'text_7',
    'text_2' => 'text_8',
    'text_2' => 'text_9',
    'text_2' => 'text_10',
    'text_2' => 'text_11',
    'text_3' => 'text_4',
    'text_3' => 'text_5',
    'text_3' => 'text_6',
    'text_3' => 'text_7',
    'text_3' => 'text_8',
    'text_3' => 'text_9',
    'text_3' => 'text_10',
    'text_3' => 'text_11',
    'text_4' => 'text_5',
    'text_4' => 'text_6',
    'text_4' => 'text_7',
    'text_4' => 'text_8',
    'text_4' => 'text_9',
    'text_4' => 'text_10',
    'text_4' => 'text_11',
    'text_5' => 'text_6',
    'text_5' => 'text_7',
    'text_5' => 'text_8',
    'text_5' => 'text_9',
    'text_5' => 'text_10',
    'text_5' => 'text_11',
    'text_6' => 'text_7',
    'text_6' => 'text_8',
    'text_6' => 'text_9',
    'text_6' => 'text_10',
    'text_6' => 'text_11',
    'text_7' => 'text_8',
    'text_7' => 'text_9',
    'text_7' => 'text_10',
    'text_7' => 'text_11',
    'text_8' => 'text_9',
    'text_8' => 'text_10',
    'text_8' => 'text_11',
    'text_9' => 'text_10',
    'text_9' => 'text_11',
    'text_10' => 'text_11',
    'text_A' => 'text_B',
    'text_A' => 'text_C',
};
'text_J' => 'text_L',
'text_K' => 'text_L'
);

# define representational connections between pictorial elements
my %representationalImage = (
    'image_A' => 'image_B',
    'image_A' => 'image_C',
    'image_A' => 'image_D',
    'image_A' => 'image_E',
    'image_A' => 'image_F',
    'image_A' => 'image_G',
    'image_A' => 'image_H',
    'image_A' => 'image_I',
    'image_A' => 'image_J',
    'image_A' => 'image_K',
    'image_A' => 'image_L',
    'image_B' => 'image_C',
    'image_B' => 'image_D',
    'image_B' => 'image_E',
    'image_B' => 'image_F',
    'image_B' => 'image_G',
    'image_B' => 'image_H',
    'image_B' => 'image_I',
    'image_B' => 'image_J',
    'image_B' => 'image_K',
    'image_B' => 'image_L',
    'image_C' => 'image_D',
    'image_C' => 'image_E',
    'image_C' => 'image_F',
    'image_C' => 'image_G',
    'image_C' => 'image_H',
    'image_C' => 'image_I',
    'image_C' => 'image_J',
    'image_C' => 'image_K',
    'image_C' => 'image_L',
    'image_D' => 'image_E',
    'image_D' => 'image_F',
    'image_D' => 'image_G',
    'image_D' => 'image_H',
    'image_D' => 'image_I',
    'image_D' => 'image_J',
    'image_D' => 'image_K',
    'image_D' => 'image_L',
    'image_E' => 'image_F',
    'image_E' => 'image_G',
    'image_E' => 'image_H',
    'image_E' => 'image_I',
    'image_E' => 'image_J',
    'image_E' => 'image_K',
    'image_E' => 'image_L',
    'image_F' => 'image_G',
    'image_F' => 'image_H',
    'image_F' => 'image_I',
    'image_F' => 'image_J',
    'image_F' => 'image_K',
    'image_F' => 'image_L',
    'image_G' => 'image_H',
    'image_G' => 'image_I',
    'image_G' => 'image_J',
    'image_G' => 'image_K',
    'image_G' => 'image_L',
);
referral connections can also be reciprocal, for example

text_L => text_A

# same for the two hashes of representational connections

my %back_representationalText = reverse %representationalText;
my %back_representationalImage = reverse %representationalImage;

my @effGazeOrderTextProcess;
my @effGazeOrderTextElement;

# optimal (theoretical) order of inspection provided by the diagram
# for processes
my @optGazeOrderTextProcess = ('text_1', 'text_2', 'text_3', 'text_4', 'text_5',
    'text_6', 'text_7', 'text_8', 'text_9', 'text_10', 'text_11');
# for objects
my @optGazeOrderTextElement = ('text_A', 'text_B', 'text_C', 'text_D', 'text_E',
    'text_F', 'text_G', 'text_H', 'text_I', 'text_J', 'text_K', 'text_L');

my $buffer;
my $refCount = 0;
my $repTextCount = 0;
my $repImageCount = 0;

my $filename = $ARGV[0];
if (!$filename) {
    print 'usage: analyzeTransitions.pl <filename>
';
    exit(1);
}

# looking for the identification number in the filename
$filename =~ /\d+\(\d+\)\sAOI\.txt$/;
my $idNumber = $1;

# Levenshtein Distance Algorithm: Perl Implementation
# by Eli Bendersky, published under the LGPL licence
# http://eli.thegreenplace.net/programs-and-code/
# minimally adapted to make it work for arrays instead of strings

sub levenshtein {
    # $s1 and $s2 are the two strings
    # $len1 and $len2 are their respective lengths
    my $type = shift (@_);
    my @s1;
    my @s2;
}
if ($type == 1) {
    @s1 = @effGazeOrderTextProcess;
    @s2 = @optGazeOrderTextProcess;
}
elseif ($type == 2) {
    @s1 = @effGazeOrderTextElement;
    @s2 = @optGazeOrderTextElement;
} else {
    # for testing purposes
    @s1 = (1,3,5,7,9,2,4,5,7,8);
    @s2 = (1,3,5,7,9,2,4,5,7,8);
}
my ($len1, $len2) = ($#s1, $#s2);

# If one of the strings is empty, the distance is the length
# of the other string
return $len2 if ($len1 == 0);
return $len1 if ($len2 == 0);

my %mat;

# Init the distance matrix
# The first row to 0..$len1
# The first column to 0..$len2
# The rest to 0
#
# The first row and column are initialized so to denote distance
# from the empty string
# for (my $i = 0; $i <= $len1; ++$i)
# { for (my $j = 0; $j <= $len2; ++$j)
# { $mat{$i}{$j} = 0;
# $mat{0}{$j} = $j;
# } $mat{$i}{0} = $i;
# }

# Some char-by-char processing is ahead, so prepare
# array of chars from the strings
my @ar1 = @s1;
my @ar2 = @s2;
for (my $i = 1; $i <= $len1; ++$i)
{
    for (my $j = 1; $j <= $len2; ++$j)
    {
        # Set the cost to 1 iff the ith char of $s1
        # equals the jth of $s2
        # Denotes a substitution cost. When the char are equal
        # there is no need to substitute, so the cost is 0
        # my $cost = ($ar1[$i-1] eq $ar2[$j-1]) ? 0 : 1;
    }
Cell \( \mat{i}{j} \) equals the minimum of:

- The cell immediately above plus 1
- The cell immediately to the left plus 1
- The cell diagonally above and to the left plus the cost

We can either insert a new char, delete a char or substitute an existing char (with an associated cost)

\[
\mat{i}{j} = \min(\mat{i-1}{j} + 1, \mat{i}{j-1} + 1, \mat{i-1}{j-1} + \text{cost})
\]

Finally, the Levenshtein distance equals the rightmost bottom cell of the matrix

Note that \( \mat{x}{y} \) denotes the distance between the substrings \( 1..x \) and \( 1..y \)

return \( \mat{\text{len1}}{\text{len2}} \);

# minimal element of a list

sub min {
  my @list = @{\$_[0]};
  my $min = $list[0];
  foreach my $i (@list) {
    $min = $i if ($i < $min);
  }
  return $min;
}

# each line in the AOI gaze data file is processed by this file

sub processLine {
  my $line = \$_;
  # file is of dos type, chop 2 times to remove newline characters
  chop($line);
  chop($line);
  my ($time, $duration, $AOIid, $AOIname, $file) = split(/\t/, $line);
  # AOIid = 0 means gaze was outside of the defined AOIs
  if ($AOIid > 0) {
    if ($referential{$buffer} eq $AOIname ||
        $back_referential{$buffer} eq $AOIname) {
      print $buffer . ' is referential to ' . $AOIname . ' \n';
      $refCount++;
    }
    elsif ($representationalText{$buffer} eq $AOIname ||
             $back REPRESENTATIONALTEXT{$buffer} eq $AOIname) {
      $repTextCount++;
    }
    elsif ($representationalImage{$buffer} eq $AOIname ||
             $back REPRESENTATIONALIMAGE{$buffer} eq $AOIname) {
      $repImageCount++;
    }
  }
}

81
if ($buffer ne $AOIname && $AOIname =~ /^text_\d*/) { push(@effGazeOrderTextProcess, $AOIname); } } 
if ($buffer ne $AOIname && $AOIname =~ /^text_\D*/) { push(@effGazeOrderTextElement, $AOIname); } 
$buffer = $AOIname;

my $fh = new IO::File $filename, "r"; 
if (defined $fh) { 
    while (<$fh>) { 
        # is first character a digit? 
        if ($_. =~ /^\d.*)/) { 
            processLine($_); 
        } 
    } 
    $fh->close;
} 

my $levTextProcess = levenshtein(1); 
my $levTextElement = levenshtein(2); 
#my $lev = $levTextProcess + $levTextElement; 
my $outputFile = new IO::File "./transitions/".$idNumber."_TRA.xml", "w";

my $XMLw = new XML::Writer (NEWLINES => 0, OUTPUT => $outputFile); 
#my $XMLw = new XML::Writer (NEWLINES => 0); 
$XMLw->xmlDecl("UTF-8"); 
$XMLw->startTag("subject"); 
$XMLw->startTag("id"); 
$XMLw->characters($idNumber); 
$XMLw->endTag; 
$XMLw->startTag("referential"); 
$XMLw->characters($refCount); 
$XMLw->endTag; 
$XMLw->startTag("representationalText"); 
$XMLw->characters($repTextCount); 
$XMLw->endTag; 
$XMLw->startTag("representationalImage"); 
$XMLw->characters($repImageCount); 
$XMLw->endTag; 
$XMLw->startTag("levenshteinViewProcess"); 
$XMLw->characters($levTextProcess); 
$XMLw->endTag; 
$XMLw->startTag("levenshteinViewElement"); 
$XMLw->characters($levTextElement); 
$XMLw->endTag; 
$XMLw->end; 
$outputFile->close(); 

exit(0);