

Eye Tracking and its application in educational multimedia

Rolf Wipfli

August 2007

Course *Attention et fonctions exécutives*
under the direction of Prof. Dirk Kerzel
University of Geneva

Contents

1	Introduction	3
1.1	Historic view	3
1.2	Types of Eye trackers	4
1.3	Types of application	5
1.4	Characteristics of eye movement	5
1.5	Perceptual Span	6
1.6	Visualizing eye movements	7
2	Controversies	8
2.1	Control of eye movement	8
2.2	Extent of parafoveal processing	9
2.3	Serial or parallel processing	9
3	Eye movement and Mental Imagery	9
4	Knowledge Acquisition	10
4.1	Integration of text and diagram	11
4.2	Model of Text and Diagram Processing	12
	References	15

1 Introduction

Studies on eye movement are numerous and the use of eye trackers is popular in current research. Although, it is wrong to assume that the study of eye movements is a recent approach. The first research on eye movement dates from 1879, when Javal, an ophthalmologist, analyzed eye movements during the lecture of text by simple observation. One of the first findings in these early studies was the fact that eye movements during lecture are not continuous, but are characterized by a rapid succession of saccades and fixations. Since these primary findings, research has made constant progress and due to the application of information technology it is more accurate and cheaper than ever. Eye tracking can nowadays be found in a large range of research fields.

In the first part of this present paper I would like to give a basic introduction to eye tracking. I will describe the basic characteristics of eye movements and the most popular paradigms which are used to analyze them. The focus is set on analyzing eye movements in the processing of text and the examination of pictures. In a next step I would like to address how the data of eye tracking analysis can help to understand the underlying cognitive processes and what controversies still exist regarding the interpretation of these data.

In a second part I would like to address the use of eye tracking in the research of text and picture processing in a pedagogical context. Learning materials often use pictures and text to explain a topic, when it is difficult to understand because of its complexity. I would like to illustrate how eye trackers are used to understand the process of learning, when the learner integrates information from both text and picture.

1.1 Historic view

Rayner (1998) summarizes the evolution of eye tracking and identifies three eras in the history of eye movement research. The first era starts in 1879, as mentioned before, with the first analysis of eye movements by Javal. He found that eye movements in reading are not continuous, but are a succession of fixations and saccades in the text. These first studies are conducted by simple observation of eye movements without any measuring equipment. It was Huey in 1908 who used the first technical installation to measure eye movements. This first apparatus, which was directly attached to the cornea, drew the eyes movements directly on a piece of paper (see Malchau, Märtner, Majer, Redel, & Sauer, 2005 for a short overview). A less intrusive method was developed by Buswell in 1922, when he introduced an installation with a light beam which was reflected by the cornea. The reflected light was captured on photosensitive paper, where the traces could later be analyzed.

The second era started on 1930 and lasted until 1958. This period was characterized by a behaviorist approach in research. Not surprisingly, there was little interest in the underlying cognitive processes during this period, but progress was made in the experimental setting of such tests.

The third era was marked by the arriving of information technology in the mid 70s, which offered new ways of collecting and analyzing data of eye tracking sessions. A

large number of research was done to see how the immensity of digitized data can be interpreted and what kind of conclusions regarding cognitive processes can be made from these data. With information technology a new application of eye tracker became possible: interactive eye tracking systems. These systems can feed information of eye movements back to the displaying system, which allows interactive application, where the displayed image can be influenced by eye movements. These systems are known as eye-contingent displays.

Since 1998, we are in the fourth era of eye tracking and this era is characterized by a further development of eye tracking systems, especially regarding their portability and affordability. Eye tracking systems can now be used in head-mounted systems which allows more ecological research of eye movements. Whereas before the collecting of eye movement data was only conducted in a laboratory it is now possible to research eye movement in daily life situations. Eye tracking systems have also found their way in a number of applications as for example in diagnostic systems in marketing or usability or in interactive systems to help disabled persons to regain some of their capabilities. A typical application, where eye-contingent eye trackers are helpful for disabled persons are eye typers, which are basically an on-screen keyboard, which can be used by eye movements.

1.2 Types of Eye trackers

Eye movements are rapid (a saccade takes between 20-200 ms) and the analysis of these movement need accurate measurement systems. There are several eye tracking system in use and they use different physical characteristics of eye movements. Eye tracking systems rely either on (a) electrodes which are attached to the skin close to the eye, (b) systems which measures 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 methods is made on the base of the principal characteristics of the eye movement which is to be analyzed. The measurement of the electric skin potential (electro-oculography) is a good method to analyze saccades, but the method is less adequate to determine the gaze direction (it is still possible though).

Scleral Search Coils are contact lenses which have magnetic coils embedded. Thus, eye movements result in changes of the magnetic field, which offers a accurate way to measure time and direction of eye movements. This type of measure is intrusive and the participant can wear the contact lenses only for maximum an hour per week (Malchau et al., 2005).

Video-Oculography is the most often used method for eye movement analysis. It is the less intrusive method and it depends either on video analysis of the eye movement or on methods which rely on reflections of infrared-light on the eye. Light is reflected on different levels on the eye (two reflections on the cornea and two reflections on the lens). These four reflections permit to calculate the exact position of the eyes. The disadvantage is that the head often needs to be fixated to assure a precise measurement.

1.3 Types of application

Duchowsky (2002) proposes a hierarchy of eye tracking systems. He first distinguishes two main different groups of eye trackers: eye trackers for diagnostic purposes and eye trackers for interactive applications. Diagnostic eye trackers analyze eye movements without an retro-action of the eye movements on the system. This type of eye tracking is primarily used to do research in cognitive psychology and to analyse the reaction of a person to visual stimuli. This latter field of application is also used in marketing or in usability testing of graphical user interfaces.

The other group of eye trackers are systems which are interactive and allow a person to influence a system (voluntarily or not). Its main use is in clinical applications (to help disabled persons) or in new forms of human-computer-interaction. Duchowsky differentiate between two types of interactive systems: selective- or gaze-contingent systems. The first one describes interfaces which let the person select elements on the screen (similar to the usage of a computer mouse), the latter describes systems where the display changes in function of eye movements.

Eye-contingent eye trackers are as well an interesting application in computer sciences as they offer a new way of presenting high density visual information. For example, it still demands high computational power to display 3D landscapes. A gaze-contingent eye-tracker can determine which part of the landscape is looked at and the system has only to render the corresponding region of the landscape.

1.4 Characteristics of eye movement

The visual field has different degrees of acuity. We distinguish between the center of vision, which is called fovea, 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 try to focus the point on the 2 degrees of foveal vision. The foveal region is surrounded by a region which extends 5 degrees from the center of vision called parafoveal. The rest of our field of vision which extends up to 150 degrees is called peripheral.

To inspect a target object in detail we have to focus on it. When we get information on a new object in our extrafoveal vision and it attracts our attention we do automatically move our eyes to get the object in our foveal vision for detailed inspection. Extrafoveal vision is used to plan our next eye movement (saccade). Even if a detailed view of an object is only possible in foveal vision, we are able to identify an object in the parafoveal region and thus without moving our eyes.

As Javal already described in 1879, reading a text is a succession of saccades and fixations and not a linear movement. This characteristic is not only true for reading, but for all type of vision. Saccades are a very rapid movement of the eye with a velocity of up to 500 degrees per second. A typical saccade in reading would take about 30 milliseconds and move for about 2 degrees. Eye movements in scene perception take slightly longer with 40-50 milliseconds for a movement of 5 degrees. During a saccade we can not perceive visual information, the movement is too rapid 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 when the eye is relatively still. Mean fixation durations are from 225 ms from silent reading until 400 ms when typing a text (Rayner, 1998). The eye is never completely still, as the eye has to continuously correct its position to keep the fixation. These corrections are called micro-movement and they compensate for shifts, due to oculomotor faults.

A third characteristic of eye movement 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. It is still uncertain what happens in this time of planning and what cognitive activities take place during this phase.

1.5 Perceptual Span

When testing the perceptual span three paradigms are used to analyze its characteristics. As mentioned above, eye-contingent eye-trackers allow us to change the display in function of eye movements. This technology is used in the three paradigms of eye movement experiments: the *moving window*, the *foveal mask* and the *boundary paradigm*.

In the moving window paradigm, a fixed region around the fovea is either completely blackened, blurred or replaced by some pattern. When the participant moves his eyes the windows follow his view, so that the fixated area is visible, whereas the surrounding areas are always hidden. In contrast, the foveal mask has the opposite effect: the focused area is in some way distorted, whereas the parafoveal and peripheral areas are visible. The purpose of these paradigms is to analyze the perceptual span by changing the size of the window. The third paradigm is the Boundary paradigm, which is used in reading tasks. As soon as the reader's saccades crosses over an invisible prespecified boundary, a target word is replaced by another. When the reader finally focuses the target word, it is evaluated whether the replaced word has a negative effect on the reading performance.

It was found that in reading the perceptual span is 12-14 character spaces to the right of fixation for adult and experienced readers. To the left of fixation, 3-4 character spaces are processed. When more characters are hidden, reading performance decreases rapidly. It is also important to mention that the perceptual span changes in function of the text difficulty. In the case where words to the right of our fixation are predictable, they can be skipped by saccades without negative effect to text comprehension.

But how is information from parafoveal vision integrated in the reading process? It was found that the view of an oncoming word in the parafoveal region (up to 5 degrees) can facilitate the recognition of this word when it is read later. It seems that most information about a oncoming word is retrieved from the first three letters of the word. The effect seems to be a phonological processing or a simple recognition of letter codes, whereas studies have rejected the ideas that there is as well an effect of semantic or morphological processing.

When reading, the eye does not always move forward. Fifteen to twenty percent of eye movements are regressions. Most of the regressions are due to difficulties to integrate a word and regressions include only a few letters. Regressions of several words indicate

a problem of text understanding. Experienced readers are very accurate in finding the text which caused them troubles.

The results of eye tracking analysis of scenery perception are more controversial. Contrary to reading, there seem to be no apparent strategy for scene perception. A common paradigm to test scene perception is the tachistoscopic presentation, where elements in a scene are quickly flashed. The duration of the presentation of the elements is too short for the eye to make a saccade and fixate the object. In this manner it can be analyzed how extrafoveal information in scenery perception is processed. Rayner (1998) is sceptical about this method and argues that findings have shown that most information in a scenery is processed at the first look and extracted mainly from extrafoveal information.

Hendersen (1992 cited in Duchowski, 2002) also criticises eye movement in scenery perception: first, gazes on an object represent rather a post-identification process than the actual recognition. Second, in contrast to reading, there is no “natural” way of scene perception. The findings are different in function of the scene presented and the task that was done to the participant. Moreover, gazes are typically longer on elements, which are not consistent with the surrounding elements and thus are more informative.

When talking of scenery perception, we can not generalize these findings for all kind of pictorial information. When eye movements are analyzed in a context where participants are given a task to solve, we can observe patterns in eye movements. Tasks would include the interaction with a graphical interface or inspecting a diagram in a learning situation. We will discuss that later in this paper.

1.6 Visualizing eye movements

There are a number of different ways to visualize eye tracking data. Some of the more common ones I would like to briefly present here. A often used representation are visualisations of the scan path (for an example see Figure 1). In these scan path diagrams a line represents a saccade and a dot represents a fixation. To indicate the duration of the fixation dots can made larger in function of the time. This kind of representation is very useful in cases where the sequence of the eye movements is important. That would be the case for reading, but is often less important in scenery perception. A disadvantage of representations by scan paths is that they can only represent one subjects’ eye movements. A comparison or an aggregation of different scan paths is difficult to accomplish with scan path visualizations.

A popular way to visualize eye tracking data are heat maps. They are easily understandable without deeper knowledge of eye tracking metrics. Heat maps are a color-coded visualization of aggregated eye tracking data. Red color areas would signify a high overall fixation time, whereas areas in blue would mean that participants looked rarely at them. The heat maps are overlaid to the image. The disadvantage of this visualization is that it includes no more information about the order of the gazes.

Areas of interest (AOI) are predefined areas, which are distinguished by their semantic or graphical information. Often the areas are defined according to the testing hypothesis. When we take the example of usability testing, an area of interest is typically a navigation menu or the header of a page. Areas of interest have the advantage that a hypothesis

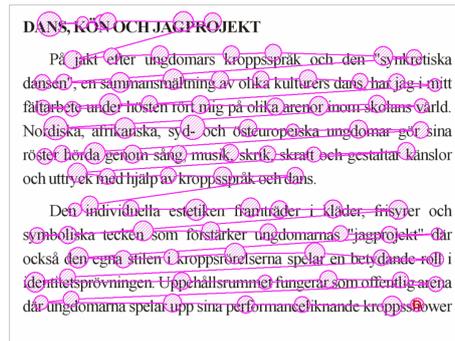


Figure 1: An example of a scan path visualization of Eye tracking data. From Lund University (in public domain)

can be directly tested, as the data is available in a quantitative and comparable form. Typical metrics would be gaze percentage per AOI or number of fixation per AOI (see Duchowski, 2006 for a case study).

When the goal is to evaluate the interaction with a number of screens (typically in usability testing), the methods above do not provide the information needed. Narayanan and Crowe (2002) propose a visualization, which includes the time passed on different areas of interests and multiple screens. Also included in the representations are haptic actions made by the participant (clicking with the mouse, pressing a key). The authors call their type of visualization Attention-Action Timeline.

2 Controversies

Even if technological advances in eye tracking methods help to gather and analyze eye movement data better than ever before, the interpretation of the data is still a methodological challenge. Starr and Rayner (2001) summarize some of the controversies about eye tracking in reading.

2.1 Control of eye movement

There are two models of the nature of eye movement control: the oculomotor model and the processing model. The first model postulates that eye movement are mainly influenced by visuomotor factors and are only little affected by linguistic factors. This theory is supported by studies which show that the eye focuses on specific points in a word, regardless of linguistic characteristics of the word itself. A reader fixates usually between the middle and the beginning of a word. This model explaining eye movement behaviour is also called the 'strategy tactics model'.

In contradiction to this model, Rayner et al. found that words with higher lexical frequency are less long fixated than lower lexical frequency. The supporter of a process model tend to emphasizes the influence of linguistic factors on eye movement. The time

of saccades and the duration of fixations are according to the process model controlled by higher cognitive functions as linguistic processing. Although, even in the process model low level oculomotor processing influences the position of the next saccade.

2.2 Extent of parafoveal processing

Another issue which is subject to present discussions is the importance of parafoveal processing in text reading. Researchers agree that information can be extracted from parafoveal vision, but it is still not clear what kind of information this includes. Studies have shown that short words of high frequency are often skipped. This would be an indicator that the word has been already recognized in parafoveal vision. It also seems that participants skip the first letters of a word when it is providing little information. The word *underneath* for example, which is not recognizable before the end of the word, is fixated at the end of word, whereas a word like *quarantine* is easily recognized by its first letters and could thus be skipped, for as the point of uniqueness is within the first four or five letters.

2.3 Serial or parallel processing

The third question of recent research is whether the fixated word and information in parafoveal vision can be processed in parallel or not. The classic serial processing theory, the assumption that text is processed in a strictly serial way, is questioned by some results found in recent research. It seems that information on the right parafoveal area can facilitate the recognition of a target word, when the information is contextually related with the target word, still the effect found is minimal. When text reading would be a parallel processing of information in foveal and extra-foveal vision, current computational models (like the 'E-Z Reader model') would have to be reevaluated.

3 Eye movement and Mental Imagery

Even if the connection between eye movements and higher cognitive processes is still under discussion, there are some interesting results in research of eye movements and mental imagery. In contrast to see eye movements as a way to gather information, eye movement also seem to be the result of cognitive processes. Brandt and Stark showed that eye movements also occur when participants imagine a scene they have looked at before (see Johansson, Holsanova, & Kolmqvist, 2005). When imagining the scene they fixated the same points on the imaginary landscape as when the scene was actually there. When participants were asked to keep focused on a spot when imagining the scene, their performance to recall the scene was significantly decreased. It even seems that eye movement is related with speaking, similar to gestures, but this is still subject to further research.

Johansson et al. (2005) argue that an explanation to this phenomena could be the representation in a *visual buffer*, which holds a mental image of the scene as proposed by Kosslyn. In a mental model objects are represented with information about the spatial

relation to each other. The mind can traverse the different objects when imagining the scene. Eye movements would follow the traversal of the object in the mental model. Previous researches used simple displays to test this effect: a grid and geometric objects. Johansson et al. (2005) verified the previous findings with a more complex scenery. To control the fidelity of their tests, the researchers applied a categorization of two levels of acceptance. They used the term 'high correspondence' for eye movements which match the direction and the final position of the object, and 'low correspondence' where only the direction of the eye movement was the same as it was in the learning phase with the real picture. In this experiment two experimental groups were created: a group which had to recall a picture they have previously examined and a second group which had to retell the description of a scenery which was told them. It was found that there existed a correspondence between eye movements and mental imaging for both experimental groups and for both, low and high correspondence evaluations.

4 Knowledge Acquisition

In a context of educational multimedia the analysis of knowledge acquisition is an interesting application of eye tracking technology. When learning we have different kind of media at hand: written text, audio recordings, schemas, pictures, or animations. In general, it is considered to improve learning when the material is presented in different ways of representation. Although, when using different medias, it is important that the information given is coherent in both representations. This enables the learner to integrate the information of both media and to construct a mental model of the information given. These findings and other guidelines in the design of multimedia learning can be found in the work of Richard Meyer (2001). He summarized his theory about multimedia learning in principles of multimedia learning design, of which I present eight:

Multimedia Principle: People learn better from words and pictures than from words alone.

Modality Principle: People learn better from animation and narration than from animation and on-screen text.

Coherence principle: People learn better when extraneous words, pictures, and sounds are excluded rather than included.

Redundancy principle: People learn better from animation and narration than from animation, narration, and on on-screen text.

Signaling principle: People learn better when the words include cues about the organization of the presentation.

Spatial contiguity principle: People learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.

Temporal contiguity principle: People learn better when corresponding words and pictures are presented simultaneously rather than successively.

Individual differences principle: Design effects are stronger for low-knowledge learners than for high-knowledge learners. Design effects are stronger for high-spatial learners than for low-spatial learners.

These guidelines give advice on the visual design of information and are thus a possible subject of eye movement analysis. The other guidelines can be found in Mayer (2001).

Pictorial representations are especially helpful in teaching materials where spatial relations or movements are an important part of the information. Mechanical documentation goes in this category and uses diagrams, which help the learner to integrate the different parts of a system in a mental model. Understanding machines require the learner to know the components of the machine, their functions and the interaction between them.

4.1 Integration of text and diagram

Hegarty and Just (1989) analyzed in several experiments the learning behaviour when reading technical instructions in text-diagram manuals. In such a manual the reader has to look back and forth between text and diagram to understand the processes which involves the different components in the system.

Diagrams can be either realistic or schematic. Schematic representation offers the possibility to emphasize the important parts of a system, while the more superficial parts can be omitted. This helps the inexperienced learner to recognize the essential parts of a system. On the other hand, a realistic diagram can help the learner to easily apply his knowledge in a real situation. Often diagrams rely on conventions, which are unique to a technical domain. For example when reading circuit diagrams in electrical engineering, distances have usually, unlike in mechanics, no meaning.

Diagrams can also differ in their labeling. The labels can be directly written next to the part it is referring to or they can be numbered and summarized in a box next to the schema. Mayers spatial contiguity principle would indicate that the first option should be preferred as object and its referent are closer. When the information in a diagram becomes dense, in-picture labels could be problematic as they could lead to an overload of information and the cognitive capacity of the learner (overwhelming effect). A solution can be to split the information in several parts and present the information sequentially.

Another type of information which can be included in a diagram, are kinematics, which indicate how the different components move when the machine is in motion. In this case the advantage of diagram is particularly high, as different objects have to be processed in parallel as they interact with each other. Diagrams seem to lower the cognitive load as the diagram can serve as an external memory.

4.2 Model of Text and Diagram Processing

Hegarty and Just (1989) propose a model of text and diagram processing with the purpose to understand how people integrate information coming from multimedia presentations. The authors identified three characteristics of a diagram which can aid the learner in constructing 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.”

Following these characteristics the authors have categorized the different types of inspections:

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 difficulty 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 more activated in working memory.

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

Hegarty and Just used an eye tracker to test their hypothesis regarding the processing of information from diagram and text. For the experiment four experimental groups were formed. The participants were assigned to a low or high mechanical ability group based on the score of the Bennett Mechanical Comprehension test, which evaluates their prior knowledge of mechanics. These groups were then assigned to either a group which had a longer text or a shorter text explaining the functioning of a pulley system (see Figure 2 for an example). Both texts were accompanied by the same schema, which showed a schematic view of a 2-pulley-system without any labels. For more information on the theoretical background and the hypothesis please refer to Hegarty and Just (1989) as this present text focuses on the use of eye tracking in their work.

The researchers used a corneal-reflectance eye tracker by Gulf and Western to monitor the subjects when they inspected the learning material. The data was recorded both digitally and on video tape. The participants eye fixations on the text or the diagram were aggregated into gazes. Later, these gazes were categorized into the three different kind of inspections, which were mentioned earlier: formation inspections, reactivation inspections, and elaboration inspections. This categorization was done by two raters who agreed in 83 % of the cases. In the other 17 % the categorization was made according to the rating by one rater.

The results showed that most diagram inspections were made at linguistic boundaries in the text. It was not part of the hypothesis of the researcher, but in the context of

Location in Text	Longer text	Shorter text
End of page (paragraph)	16 (27.1 %)	15 (42.9 %)
End of sentence	15 (25.4 %)	3 (8.6 %)
End of clause	7 (11.9 %)	2 (5.7 %)
End of noun phrase	4 (6.8 %)	1 (2.9 %)
Other	17 (28.8 %)	14 (40.0 %)

Table 1: Locations in the text at which subjects inspected the diagram.

the present paper it is informative and I included the table, which shows the locations in the text at which subjects inspected the diagram (see Tab. 1).

Participants read preferably until the end of a clause or a sentence, and they seem to “fully interpret a sentence or clause of text before checking the representation of this unit of text against the diagram”. Also the data suggests that participants progressively construct a mental model as they do not finish reading the whole text before referring to the diagram. Most of the inspections of the diagram were classified as reactivation inspections, the remaining inspections were equally found to be formation or elaboration inspections.

The authors criticised their prior categorization in the discussion of their results. The reactivation inspections (gazes on parts of the system, which did not correspond to the text they just have read) were numerous and the participants looked at different parts of the diagram especially after having finished reading the complete text. The results showed also that reactivation inspections were considerably longer (3.4 seconds) than formation inspections (1.2 seconds) and the authors assume that it is probably not only a reactivation of already inspected parts of the diagram. The purpose of reactivation inspections could also serve to integrate information of different parts of the pulley system or to encode new information.

The strategy when reading a diagram seem to be different for low- and high-mechanical-ability subjects. High-mechanical-ability subjects spend more time inspecting the diagram to encode new information not present in the text. But formation inspections and reactivation inspections were not significantly different for the two groups of different mechanical-ability.

In 1993, Hegarty and Just conducted a new research with a similar testing disposition. Again they used a pulley system (Figure 2) as learning material. This time the authors did not categorize the diagram inspections as formation inspections, reactivation inspections or elaboration inspection, but they distinguished between local or global inspections.

Local inspections are generally shorter inspections which include not more than three components. They serve to establish co-reference 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

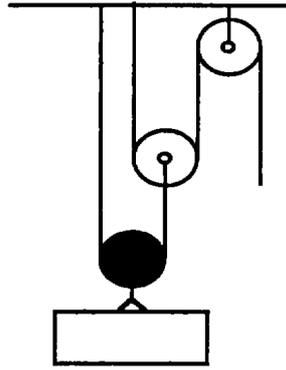


Figure 2: A 3-pulley system (copied from Hegarty and Just, 1993)

to integrate the different parts of the system in order to construct a mental model of the components and their interaction.

The authors agree that in reality the distinction between local and global inspections might not be dichotomic, but more like a continuum. Still, when using this method, it was found that global inspections occurred significantly more often when finished reading the paragraph ($M = 53.7\%$, $SD = 23.2$) than in mid-text ($M = 29.1\%$, $SD = 22.5$).

Regarding text processing, Hegarty and Just distinguished between two types of regressions in the text: regressions within reading episodes and regressions across reading episodes. The first mentioned might be used as attempts to integrate the information in the text before switching to the diagram (34% of the regressions were of this type). Regressions across reading episodes might serve to relocate the subject's place in the text as most regressions ($M = 72\%$, $SD = 18\%$) occurred just after they inspected the diagram.

Participants switched in 36% ($SD = 8\%$) of the cases during local inspections to referents of the most recently read clause. A further 44% of fixated components were referents to the next two most recently read clauses and finally 80% ($SD = 7\%$) were referents of the three most recently read clauses of the text. "A possible interpretation of this behaviour depends on the assumption that the unit of representation for the text base is the clause, whereas the unit of representation for the mental model is the component."

A presentation of text and diagrams seem to aid the learner in constructing a mental model, especially if the learner has low abilities in mechanics and mental manipulations in a spatial context are required. Diagrams prove to be especially helpful when a system has moving parts or when there are other complex relations between the components. It is important that text and diagram are available simultaneously so the reader can switch forth and back to construct a mental model from both medias. In all conducted tests diagram inspection was largely text directed. Diagram may also serve as an external representation of the mental model, which frees up working memory resources.

References

- Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, and Computers*, 1.
- Duchowski, A. T. (2006). High-level eye movement metrics in the usability context.
- Hegarty, M., & Just, M. A. (1989). Understanding machines from text and diagrams. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 171–194). North-Holland: Elsevier Science Publishers.
- Johansson, R., Holsanova, J., & Kolmqvist, K. (2005). *What do eye movements reveal about mental imagery? evidence from visual and verbal elicitations*. Proceedings of the Cognitive Science Conference in Stresa, Italy.
- Malchau, W., Märtner, S., Majer, C., Redel, B., & Sauer, H. (2005). *Eye tracking*. Downloaded 5th august 2007 from <http://www.evaluationstechniken.de/>.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge U.K.: Cambridge University Press.
- Narayanan, N. H., & Crowe, E. C. (2002). Integrating eye movement and haptic interaction data for comparing multimedia interfaces. *Journal of Applied Systems Studies*.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Starr, M. R., & Rayner, K. (2001, April). Eye movements during reading: some current controversies. *TRENDS in Cognitive Sciences*, 5(4), 156–163.