

Chapter 9

Studying Eye Movements in Multimedia Learning

Huib K. Tabbers, Fred Paas, Chris Lankford, Rob L. Martens,
and Jeroen J.G. van Merriënboer

Abstract This paper discusses the usefulness of tools that enable the analysis of eye movement data in dynamic interfaces for investigating theoretical issues in the area of multimedia learning. One of these tools, *GazeTracker*TM, a program that links eye movement data to information about the internal computer processes and automatically combines the two for further analyses, is discussed. The functionality of the tool for studying the process of multimedia learning is illustrated with an experiment on the integration of text and pictures in a web-based lesson on instructional design. In the experiment, differences in fixation patterns between several presentation formats are investigated. It is concluded that tools like *GazeTracker*TM make it easier to study how people integrate text and pictures in dynamic interfaces like web browsers.

Keywords Eye movement · Integration · Picture · Text · Time constraints

9.1 Introduction

In educational psychology there is an on-going debate about how students learn with pictures and text, especially with the rise of multimedia learning environments where people have to integrate verbal and pictorial information. This has produced a large amount of empirical research on the effectiveness of different presentation formats (e.g., Mayer, 2001). However, as far as we know, hardly any research in this area has used eye-tracking methods to study looking behavior. Eye movement measures might be a very interesting addition to the research on multimedia learning, especially because the existing theories are partly based on assumptions about where people look when they are integrating text and pictures.

H.K. Tabbers
Institute of Psychology, Erasmus University Rotterdam, Woudestein, T12-39,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
e-mail: tabbers@fsw.eur.nl

In the field of eye movement research, there are numerous studies on reading behavior and on scene perception (see Rayner, 1998, for an overview), but only a few studies on the integration of text and pictures (Duffy, 1992). These are notable exceptions like studies by Hegarty on mental animation (Hegarty, 1992a,b; Hegarty & Just, 1989, 1993), the work of d'Ydewalle and colleagues on television subtitles (for an overview, see d'Ydewalle & Gielen, 1992), some work on the perception of cartoons by Carroll, Young, and Guertin (1992), a study on how people look at advertisements by Rayner, Rotello, Stewart, Keir, and Duffy (2001) and a study of sentence-picture verification tasks by Underwood, Jebbett, & Roberts (2004). Most of these studies used static images so that the gaze position of the participants could easily be related to the different elements of the scene. Dynamic interfaces substantially increase the complexity of the data analysis because changes on the screen have to be directly related to the eye movement data (e.g., Goldberg & Kotval, 1999; Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002). This makes eye tracking research in the area of multimedia learning not an easy task to do.

Fortunately, some interesting analysis tools have become available that integrate eye movement data with the dynamic processes that simultaneously take place on the computer screen (e.g., Crowe & Narayanan, 2000; Lankford, 2000b). In this article we will discuss the usefulness of these tools for examining theoretical issues related to the area of multimedia learning, and describe an experiment in which we applied one of these tools called *GazeTracker*TM.

9.2 The Added Value of Studying Eye Movements in Multimedia Learning

Recent theories on multimedia learning like Mayer's generative theory (2001) and cognitive load theory (Sweller, 1999; Sweller, Van Merriënboer, & Paas, 1998) are based on a number of assumptions about the learner's cognitive architecture. Both Mayer's theory and Sweller's cognitive load theory stress the relevance of limitations in working memory capacity for processing multimedia instructions and the differences in processing verbal and pictorial materials. According to both theories, learners who are presented with a picture and an accompanying (visual) text have to split their attention between both information sources, resulting in a possible overload in (the visual part of) working memory. To prevent this overload and to enhance learning, several design guidelines have been proposed that have been tested in a number of empirical studies. For example, one design guideline is to replace visual (written or on-screen) text with spoken text in multimedia instructions (the so-called *modality* principle). Applying this guideline has resulted in superior learning in terms of faster problem solving (Jeung, Chandler, & Sweller, 1997); Mousavi, Low, & Sweller, 1995), higher scores on retention and transfer tests (Kalyuga, Chandler, & Sweller, 1999, 2000; Leahy, Chandler, & Sweller, 2003; Mayer &

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

Although a great number of empirical studies support the design guidelines derived from the theories of Mayer and Sweller, none of these studies has actually taken a closer look at the process of multimedia learning. What are learners looking at when they are watching a multimedia instruction? And what exactly are people doing when they are trying to integrate text and picture? Eye tracking methods can give at least a partial answer to these questions, by providing information about the gaze position of the learner during the learning process. Moreover, an answer to these questions can help advance research on multimedia learning in at least two ways.

First of all, most researchers in the field of multimedia learning have developed their own multimedia materials for their experiments. They assume that both the textual and pictorial information in their materials are necessary for understanding (unless of course one is interested in the so-called *redundancy* effect). However, this assumption is not tested as measures like mental effort scales, time-on-task and learning results do not really tell if learners have actually looked at both pictures and text. After all, to mentally process an information source like a picture it will have to be perceived first. In order to know if learners treat the materials as real 'multimedia' instructions, measures of eye movements can provide the researcher with valuable information that can help in optimizing their multimedia materials for doing research.

The second advantage is that eye movement data can yield additional evidence for the theoretical rationale behind certain design guidelines. Different presentation formats of multimedia instructions do not only result in different cognitive processes (more or less cognitive load; more or less effective learning), but also lead to differences in looking behavior. For example, one of the guidelines deriving from cognitive load theory is that text should be physically integrated with a picture, in order to prevent unnecessary visual search (e.g., Chandler & Sweller, 1991, 1992). With eye-tracking research, the amount of visual search in the split-attention condition might be compared with the amount of visual search in the integrated condition. This way, the eye movement data can reveal if the underlying explanation of the guideline is supported or that alternative explanations are needed. For example in the case of split-attention, the crucial factor might be not the amount of visual search, but the fact that people in the split-format condition do not look at the right parts of the picture. So eye-tracking methods do not only test theoretical assumptions in multimedia learning, but can also provide alternative explanations for the effects that are found.

However, as far as we know, none of the studies inspired by Mayer's theory or by cognitive load theory has taken a closer look at the process of integrating text and picture by measuring the eye movements of the learners. One of the reasons is that the multimedia learning materials used in this research area are often presented as interactive web pages or animations. In these dynamic

environments the analysis of eye movement data is a tough job, because the eye position is usually calibrated in relation to a static image. That is why tools are needed that integrate the eye movement data with the user's interactions with the computer and simplify the subsequent analyses.

9.3 The *GazeTracker*TM Software

*GazeTracker*TM is a tool for analyzing eye movement data in dynamic multimedia environments, and resulted from the work on the Eye-gaze Response Interface Computer Aid (*ERICA*) at the University of Virginia (Lankford, 2000a). The *ERICA* system helps individuals with disabilities communicate via the computer, and takes the eye movements of the user as input to operate mouse and keyboard functions in software applications. To facilitate the analysis of eye-movement data, the *GazeTracker*TM software was developed (Lankford, 2000b).

The program combines the input from eye-tracking systems like *ERICA*, *ASL* or *SMI* with information about the activities of the user of a computer application, like keystrokes and mouse clicks. It receives the eye-tracking data through a serial port and uses a global timer to synchronize the data it reads from the serial port with the mouse and keyboard data it intercepts from the operating system. *GazeTracker*TM accomplishes this by integrating itself into the low-level functions of the Windows operating system. The integration with Windows also allows the program to track the web pages that each test subject visits in the Internet Explorer, and to compensate the recorded eye-gaze and mouse data with the current scroll bar position. This ensures that all captured data is associated with the proper content shown on the screen during the experiment. Moreover, the program can parse the HTML-code of web pages and automatically create areas of interest (LookZones) for each hyperlink and image (based on information in tags like `<a>` and ``). These LookZones can also be manually defined by the user and can take any size or shape. After recording, the data including the interactions of the user with the applications can be replayed, and can be displayed as a gaze trail, which depicts the scan path of a test subject superimposed on an application window (see Fig. 9.1).

So *GazeTracker*TM relates all activities on the screen to gaze position data, and has the opportunity to track eye movements in several applications simultaneously and even control for scrolling behavior. That way it becomes much easier to conduct eye movement research with dynamic interfaces like web browsers, and to study the way people integrate textual and pictorial information in these environments. Moreover, with LookZones information can be gathered on how long and how often a test subject observed different areas of interest like text boxes and pictures. For further analysis, the program provides several graphical methods, such as bar charts in Excel based on the LookZone data, or three-dimensional views of the application window with the time

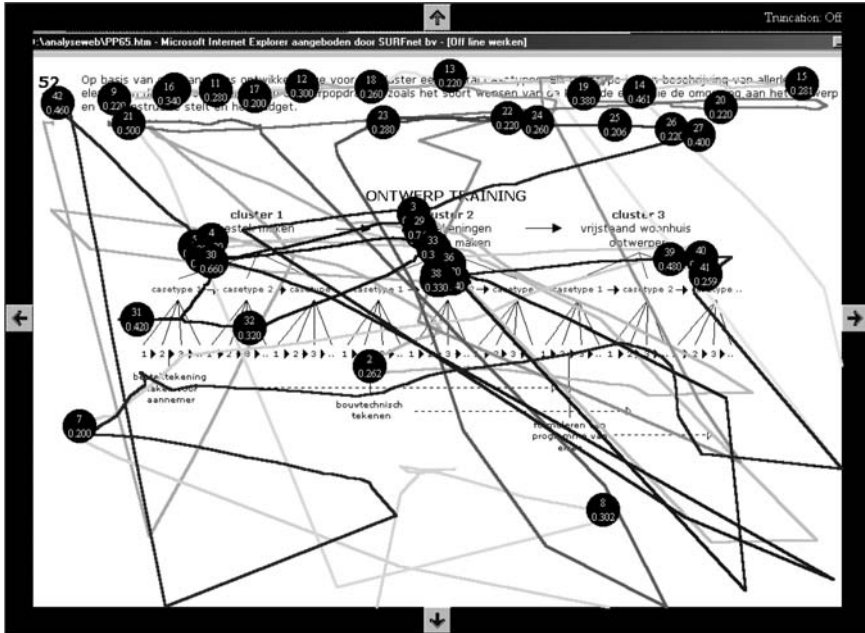


Fig. 9.1 Screen example of the gaze trail superimposed on a fragment of the multimedia materials used in the experiment described in this article. The gaze trail is depicted as a (multicolored) line, and the fixations are depicted as numbered black circles with the fixation duration printed inside

duration of the fixations in different regions depicted in the z-dimension. *GazeTracker*TM also allows experimenters to export the data to text files or Microsoft Excel for further statistical analysis in other statistical software packages.

9.4 Experiment

9.4.1 Objectives

To illustrate the usefulness of a tool like *GazeTracker*TM for research on multimedia learning, we set up a small experiment that builds on our previous work on the modality effect in multimedia learning (Tabbers, 2002; Tabbers, Martens, & Van Merriënboer, 2001, 2004). In these studies, we used a multimedia lesson that consisted of a series of diagrams accompanied with an explanatory narration. Not only did we vary the modality of the accompanying text (spoken text versus on-screen text), but we also varied the pacing of the instructions. Earlier research by Mayer and others had shown that giving

learners control over the presentation rate might have a positive effect on multimedia learning in terms of higher transfer scores (Mayer & Chandler, 2001; see also Mayer, Dow, & Mayer, 2003). In Tabbers et al. (2001), we compared multimedia instructions with a pacing based on the speech rate of the narration with learner-controlled instructions. We found that with system-paced instructions, spoken text yielded superior learning results as would be predicted by the modality effect, whereas with learner-paced instructions, hardly any difference in effectiveness was found between spoken text and visual text. In Tabbers et al. (2004), we even found a reverse modality effect (superior learning with visual text) when the learners controlled the pace of the instruction. So based on these results we concluded that the modality effect does not apply when learners control the pace of the instructions.

However, the question is how to explain these findings. The general assumption behind the modality effect in multimedia learning is that the integration of spoken text and pictures is mentally less demanding than the integration of visual text and pictures. Sweller (1999) points out that the split format of visual text and picture requires holding components of the picture or the text in working memory while searching for the relevant referents in the text or picture. Furthermore, once the right section of the text or picture has been found, both information sources have to be mentally integrated. These processes of visual search and mental integration take up a good deal of working memory capacity, but are not essential to learning, according to Sweller. Preventing this unnecessary cognitive load, for example by physically integrating text and picture, will make extra working memory resources available for the learning process.

Another way of increasing the available working memory resources is by presenting text as spoken word. Both Mayer (2001) and Sweller (1999) base their explanation of this modality effect on the working memory model of Baddeley (1992). According to his model, working memory consists of separate processors for auditory and visual information. When text and picture are presented in visual form, they will both be processed in the visual channel (at least initially), so they have to compete for the same limited resources. Presenting the text in auditory form will take off the load of the visual subsystem. Moreover, the auditory subsystem will be used more optimally, so that the available working memory resources for learning will increase. Thus, the explanation of the modality effect is mainly in terms of cognitive processes (increasing working memory resources).

However, this cognitive explanation alone does not suffice to explain the disappearance or reversal of the modality effect with the introduction of learner control. Therefore, a closer look is warranted at what goes on when learners are watching a multimedia instruction. Apart from a cognitive advantage in terms of an increase in working memory resources, learners listening to a narration and watching a picture can immediately integrate text and picture, provided they are watching the right parts of the picture. Learners with visual-only instruction have to split their attention between visual text and picture and

cannot process them simultaneously. That implies that if the pacing of the instruction is based on the narration, learners in the visual-only condition have less time available to study both text and picture. This might not be such a big problem. After all, as long as learners are reading faster than the pace of the narration, they will have enough time left to look at the picture as well. However, one could argue that giving the learners control over the pacing of the instructions will make it a lot easier for them to integrate visual text and picture, because more time can be spent on both text and picture. In fact, in one of our studies we did find that students in the visual-text condition spent 25% more time on the instructions when the instructions were learner-paced (Tabbers et al., 2001). That way, the cognitive load of the visual-only instructions may have been decreased, undoing the advantage of dual-mode instructions and making the modality effect disappear (or even reverse).

This hypothetical explanation for the disappearance of the modality effect with learner controlled multimedia instructions cannot be studied by looking at outcome measures alone. Process-based information is needed that reveals how much time is spent on either reading a text or looking at a picture. Measuring eye movements and looking at the different fixation patterns might provide exactly this.

Therefore we set up a small-scale experiment in which we studied eye movements using the same multimedia materials as in our previous studies (Tabbers et al., 2001, 2004). We compared three different presentation formats: system-paced instructions (in which the pacing was based on the narration) with either spoken text or visual text, and learner-paced instructions with visual text. Tabbers et al. (2001) showed that system-paced visual-text instructions resulted in the worst transfer performance, and explained this effect by stating that the students in this condition might lack the time to inspect the diagram after reading the text. Translated to eye movement data that results in the following hypothesis: Total time fixated in the diagrams will be shorter in the system-paced visual-text condition than in the audio and learner-paced visual-text condition.

Secondly, we wanted to check for the explanation of the modality effect in terms of differences in working memory load, and tried to see if eye movement data could provide additional support for this explanation. Therefore, we looked at some possible indicators of mental workload that are related to eye movements like fixation frequency (number of fixations per second) and average fixation duration (Van Orden, Limbert, Makeig, & Jung, 2001) and compared these to a more commonly used self-report measure of mental effort (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). We expected memory load to be lowest in the audio condition, resulting in the lowest mental effort scores, the lowest fixation frequency and longest average fixation duration, and to be highest in the system-paced visual text condition, with the highest effort scores, highest fixation frequency and shortest average fixation duration.

9.4.2 Method

9.4.2.1 Participants and Design

The participants were 12 students from a Teacher Training College for Primary Education (age between 17 and 23; 1 male and 11 females). They had applied on a voluntary base and were paid 10 euros for their participation. Because of the large individual differences in looking patterns, we used a within-subjects design in this small-scale study. Each participant studied the multimedia instructions in three parts and each part was presented in a different presentation format (system-paced audio, system-paced visual text, learner-paced visual text). To prevent any sequencing effects, the order of presentation formats was counterbalanced between the participants.

9.4.2.2 Apparatus

The eye movements were recorded with a 50 Hz video-based remote eye-tracking device from SensoMotoric Instruments (*SMI*). The infrared camera was placed under the 21-inch display screen of the stimulus PC on which the multimedia instructions were presented. Special *SMI*-software to operate the camera and the calibration process ran on a separate PC that was connected to the stimulus PC. On the stimulus PC, the *GazeTracker*TM program combined the input of eye movement data from the *SMI*-PC with data of the user interactions with the web browser. A chin and forehead rest was placed in front of the screen in such a way that the subject's eye was 70 centimeters from the computer screen and level with its center. To calculate fixations (the relatively stable moments in the gaze trail during which information is most likely to be processed), *GazeTracker*TM uses a dispersion-threshold identification algorithm with a moving window (see Salvucci & Goldberg, 2000). The dispersion threshold was set at 25 pixels, which corresponds to approximately three or four letter spaces in the instructional material or 1 degree of visual angle, and the duration threshold was set at 100 milliseconds.

9.4.2.3 Materials

- Multimedia instructions.

The instructions used in the experiment discussed the four-component instructional design model (4C/ID model) of Van Merriënboer (1997) and were developed with Microsoft FrontPage as a linear sequence of web pages. Each page consisted of a diagram representing a skills hierarchy or an elaborated sequence of learning tasks and a textual explanation accompanying the diagram. The textual explanation that accompanied the eight diagrams was presented in smaller fragments of only one or two sentences long, that were presented one at a time. Together, the (eight) diagrams formed three

worked-out examples showing how the 4C/ID model was applied in designing a blueprint for a training program.

Each of the three worked-out examples was presented in a different format: a system-paced audio format, a system-paced visual-text format or a learner-paced visual-text format (see Fig. 9.2 for screen examples of each presentation format). In the system-paced audio format, students could listen to the text fragments that accompanied a diagram, whereas in the system-paced visual-text format, students could read these text fragments from screen right above the

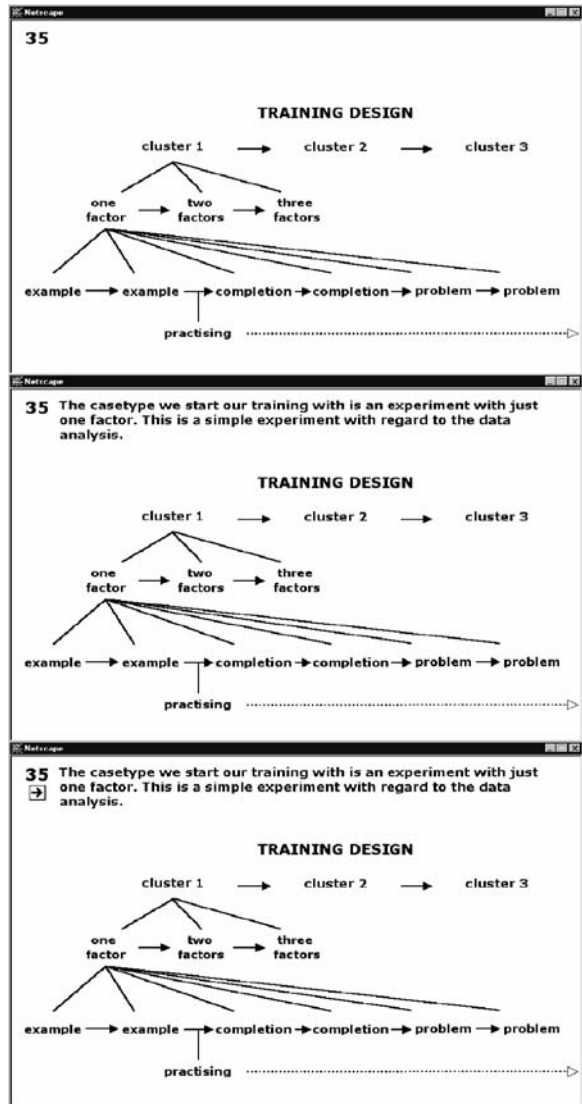


Fig. 9.2 Screen examples of the three different presentation formats (translated from Dutch). From top to bottom: the system-paced audio format, the system-paced visual-text format and the learner-paced visual-text format

diagram, with the same pacing as the audio fragments. With the learner-paced visual-text format students could reread each text fragment as many times as they wanted to before continuing with the next piece of text by clicking on a forward button. The presentation time of each worked example was about 6 minutes, except of course in the learner-paced visual-text format where the total time to study a worked-out example was variable.

- Mental effort scale

To measure mental effort a 9-point scale was used on which the students could rate the mental effort they had spent ranging from *very, very low mental effort* to *very, very high mental effort*. The scale was developed by Paas (1992), based on a measure of perceived task difficulty of Borg, Bratfish, and Dornic (1971). The scale's reliability and sensitivity (Paas, Van Merriënboer, & Adam, 1994) and its non-intrusive nature make this scale a useful measure of perceived working memory load, and it has been used extensively in studies of multimedia learning (for an overview, see Paas et al., 2003).

- Evaluation questionnaire

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

9.4.2.4 Procedure

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

9.4.3 Results and Discussion

The main dependent variables in the experiment were total time fixated and number of fixations (overall, in the text and in the diagrams), and average fixation duration, fixation frequency and perceived mental effort. We conducted a repeated measures MANOVA, with presentation format as the within-subjects factor. For any post-hoc analyses we used paired t-tests. For all statistical tests, a significance level of 0.05 was applied. Table 9.1 shows the means and standard deviations for all dependent measures.

For the overall eye movement results, we found a significant effect of presentation format on total time fixated and number of fixations (Wilks' lambda = 0.24, $F(4, 42) = 10.88, p < 0.01$), but no specific differences in the post-hoc tests. Looking at the division of attention over text and diagram, the results showed that students in the audio condition spent more than 98% of their total fixation time in the diagrams, versus 44% in the system-paced visual-text condition and 38% in the learner-paced visual-text condition. When analyzing the fixations in the diagrams separately, again a significant effect of presentation format was found on total time fixated and number of fixations (Wilks' lambda = 0.61, $F(4, 42) = 2.93, p < 0.05$). Post-hoc comparisons showed

Table 9.1 Means and standard deviations of dependent measures

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

that in the audio condition, students' total fixation time was significantly longer and number of fixations was higher than in the system-paced visual-text condition ($t=2.62$, $p < 0.05$ and $t=2.46$, $p < 0.05$, respectively), and than in the learner-paced visual-text condition ($t=2.71$, $p < 0.05$ and $t=2.47$, $p < 0.05$, respectively). However, no significant differences were found between the visual-text conditions. When the visual-text conditions were compared on total time fixated and number of fixations in the text only, again no significant differences were found (Wilks' lambda = 0.87, $F(2, 10) = 0.77$, $p > 0.05$).

The effect of presentation format on the indicators of workload like average fixation duration, fixation frequency and mental effort was also significant, Wilks' lambda = 0.16, $F(6, 40) = 10.88$, $p < 0.01$. Post-hoc comparisons showed that the participants in the audio condition fixated less frequently than the participants in both the system-paced visual-text condition, $t = 4.85$, $p < 0.01$, and the learner-paced visual text condition, $t 8.23$, $p < 0.01$. Related to this finding, the average fixation duration was longer in the audio condition than in the system-paced visual-text condition, $t = 6.73$, $p < 0.01$, and the learner-paced visual-text condition, $t = 6.34$, $p < 0.01$. Although the participants reported a higher mental effort score in the system-paced visual-text condition than in the other two conditions, this difference did not reach statistical significance. Looking at the average fixation duration in the text, no significant difference was found between the learner-paced and system-paced visual-text condition.

Finally, the results of the evaluation questionnaire showed that two-thirds of the students had preferred the learner-paced visual-text version over the other two versions. Moreover, the students judged the part of the instructions presented in the learner-paced visual-text version as the easiest to comprehend.

So the results do show some clear differences in fixation patterns between the presentation formats, but not in the way that we hypothesized. Naturally, the looking pattern in the audio condition deviates from the patterns in the visual-text conditions, because there is no text to fixate on. However, the division of attention between diagram and text in both visual text conditions seems to be quite identical, contrary to what we expected. Moreover, no apparent differences in fixation data are found between system-paced and learner-paced instructions. A closer look at the different scan paths (how did the learner's gaze switch from text to diagram) might reveal other differences in switching behavior between the different visual-text formats, but such an analysis was beyond the scope of the current study. In their work on the integration of diagram and text, Carroll et al. (1992), Hegarty and Just (1993), Rayner et al. (2001) and Underwood et al. (2004) found that most subjects read the text first and then looked at the diagram, without much switching. As study time was not limited in these studies, the same fixation pattern could be expected in a learner-paced condition. It would be interesting to see if an identical pattern would be found in the system-paced condition, or that a different scan path would emerge.

Furthermore, looking at the possible workload indicators, it is interesting that the students fixate less frequently in the audio condition with a longer

duration, just as we hypothesized. Primarily, this difference seems to reflect the ‘calmness’ of the looking pattern in the audio condition, where students do not have to switch between text and diagram. It is unclear, however, if this is also related to less cognitive load in the audio condition, because we do not find a similar difference between audio and both visual text conditions in the mental effort scores. The relationship between mental effort and fixation duration and fixation frequency might not be as direct as supposed, so further research in this area is needed.

In sum, we hypothesized that the students in the learner-paced condition would spend extra time in the diagrams, but we do not find it in the results. So the difference in effectiveness between system-paced and learner-paced multimedia instructions found in our earlier studies (using the same materials) does not seem to derive from an overall difference in fixation pattern, at least in terms of total time fixated or number of fixations. Nevertheless, students report a relatively high mental effort in the system-paced condition, and generally prefer the learner-paced visual-text version. It might be the case that the demonstrated superiority of learner-paced over system-paced visual-text is not the result of a general difference in fixation time, but because students can control the division of attention between diagrams and text more easily and adapt it to their individual needs. To fully test this hypothesis, an approach is needed that more directly links the eye movement data to a process model of how people integrate text and picture to construct meaning, like for example the model of Narayanan and Hegarty (1998, 2002).

9.5 General Discussion and Conclusions

Our study shows that the use of a tool for analyzing eye movements like *GazeTracker*TM can produce more specific insights on processes that take place during multimedia learning. By integrating eye movement data with computer processes, interesting information can be obtained on the way that people learn with text and pictures. Despite the dynamic nature of the presented material and the large number of different web pages in our experiment, the analysis could be done relatively easy, because *GazeTracker*TM automatically loaded the areas of interest in our study (i.e. the diagrams and the text boxes) as LookZones, and simplified the subsequent data analysis by offering the opportunity to indicate which data (of different participants, web pages and LookZones) should or should not be included in the analysis.

Of course, some elements of the analysis still can be improved upon. For example, the version of *GazeTracker*TM used in our experiment did not provide any summary data on the ‘switches’ from one LookZone to the other, like from text to diagram. However, newer versions of *GazeTracker*TM do provide the opportunity to create a LookZone Order Graph that displays

the order and duration in which a subject observed different regions of interest, so that specific hypotheses on switching behavior can be studied more easily. Another drawback was that the program did not support some complex analyses, like aggregating data of multiple participants over multiple LookZones (for example all diagrams in one worked example), so we had to extract these data from the database file ourselves. Although that was not a real problem, it took us a lot more work to make these summary data available for further analyses.

These are of course only technical drawbacks of the program that can and hopefully will be solved in the near future. Nevertheless, some more general remarks can be made on doing eye tracking research in the area of multimedia learning. First of all, the quality of the analysis with *GazeTracker*TM (or any other analysis tool) is very dependent on the quality of the eye tracking system used. For example, the system we used in our study had some drawbacks, like a relatively low resolution (50 Hz), and some difficulties in getting the participants' eyes calibrated. Care has to be taken to use optimal equipment for eye tracking research, especially when more fine-grained analyses of gaze positions are warranted. Furthermore, a more fundamental problem is that eye tracking methods produce huge amounts of process data. However, most of our current theories on multimedia learning do not provide hypotheses on the exact looking behavior of learners. This is of course complicated by the fact that large individual differences exist in the way that people process instructions. Therefore, researchers in the field of multimedia learning interested in eye tracking research should carefully consider if their hypotheses can be reformulated in such a way that they can be tested with eye movement data, and that they can indicate as precisely as possible which information they would like to extract from the data. Only then will tools like *GazeTracker*TM be of added value in simplifying the analysis of the eye movement data.

In conclusion, the use of tools like *GazeTracker*TM makes eye tracking methods available for the study of learning in dynamic multimedia environments, where different information elements are presented at different locations and at different times. With these tools, it is possible to identify where people look when they are studying multimedia materials, so that researchers can find out if learners treat their study materials as was intended in the design. Furthermore, with these tools the underlying explanations of theories of multimedia learning can be tested, at least those hypotheses that can be reformulated in terms of eye movement data. These advantages are not only interesting for the area of multimedia learning, but for any other study of human-computer interaction aimed at a further understanding of the cognitive processes that take place when people are working with a computer application.

Acknowledgments The experiment was part of a Ph.D. project that was co-funded by Aebly, a former multimedia company from 's Hertogenbosch, The Netherlands. We would like to thank Léon Sluijsmans from Hogeschool Zuyd in Heerlen for his helpful assistance in recruiting our participants.

References

- Baddeley, A. D. (1992). Working memory. *Science*, 255, 556–559.
- Borg, G., Bratfisch, O., & Dornic, S. (1971). On the problem of perceived difficulty. *Scandinavian Journal of Psychology*, 12, 249–260.
- Carroll, P. J., Young, J. R., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444–461). New York: Springer-Verlag.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233–246.
- Crowe, E. C., & Narayanan, N. H. (2000). Comparing interfaces based on what users watch and do. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 29–36). New York: ACM Press.
- d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415–427). New York: Springer-Verlag.
- Duffy, S. A. (1992). Eye movements and complex comprehension processes. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 462–471). New York: Springer-Verlag.
- Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631–645.
- Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks: Design implications. In *Proceedings of the Eye Tracking Research and Applications Symposium*. New York: ACM Press.
- Hegarty, M. (1992a). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1084–1102.
- Hegarty, M. (1992b). The mechanics of comprehension and comprehension of mechanics. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 428–443). New York: Springer-Verlag.
- 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). Amsterdam: Elsevier.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717–742.
- Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, 17, 329–343.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126–136.
- Lankford, C. (2000a). Effective eye-gaze input into Windows™. In *Proceedings of the Eye Tracking Research and Applications Symposium*. New York: ACM Press.
- Lankford, C. (2000b). GazeTracker™: Software designed to facilitate eye movement analysis. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 57–63). New York: ACM Press.
- Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instructions. *Applied Cognitive Psychology*, 17, 401–418.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93, 390–397.

- Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*, 806–813.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.
- Narayanan, N. H., & Hegarty, M. (1998). On designing comprehensible interactive hypermedia manuals. *International Journal of Human-Computer Studies, 48*, 267–301.
- Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*, 279–315.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434.
- Paas, F., Tuovinen, J. E., Tabbers, H. K., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71.
- Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419–430.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
- Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*, 219–226.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71–78). New York: ACM Press.
- Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.
- Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.
- Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 1024–1029). Mahwah, NJ: Erlbaum.
- Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257–287.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology, 57A*, 165–182.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*, 111–121.