This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Commentary

The use of eye movements in the study of multimedia learning

Jukka Hyöna

Department of Psychology, University of Turku, FI-20014 Turku, Finland

Abstract

This commentary focuses on the use of the eye-tracking methodology to study cognitive processes during multimedia learning. First, some general remarks are made about how the method is applied to investigate visual information processing, followed by a reflection on the eye movement measures employed in the studies published in this special issue. It is argued that global eye movement measures indexing attentional and encoding processes during the entire learning period should preferably be complemented with more fine-grained analyses that are either time-locked to important events taking place in an animation or that by other means provide information about the time course of learning. As nicely documented in the present set of studies, it is also of importance to complement the eye-tracking data with offline measures indexing the end product of learning. Such a complementary approach is likely to yield important new insights into the process of multimedia learning.

Keywords: Eye movements; Multimedia learning; Visual attention; Comprehension

1. Introduction

In this commentary, my focus is on the use of the eye-tracking methodology to study cognitive processes during visual learning tasks, especially during learning from written texts, graphics and animations. I start out by saying that I very much welcome the application of the method to multimedia learning. As the papers of this special issue witness, eye-tracking can reveal important insights into the ongoing learning process. To date, the method has been successfully applied, for example, to the study of cognitive processes in scene perception (for a review, see Henderson, 2003) and in reading (for a review, see Rayner, 1998), but eye-tracking studies of the processing of multimedia materials are still relatively sparse.

In studies of scene perception, one key question has been to determine the relative contributions of low-level visual features versus higher-level cognitive factors on human gaze behavior. In reading research, the focus has been on uncovering the mental processes that contribute to successful recognition of words and to successful parsing of the sentence structure. However, relatively little effort has been devoted to the processing of expository texts mimicking textbook materials (for exceptions, see e.g., Hyöna & Lorch, 2004; Hyöna, Lorch, & Kaakin, 2002; Hyöna & Nurminen, 2006; for possible eye movement measures to be used in such studies, see Hyöna, Lorch, & Rinck, 2003). Although I think it is a pity that eye-tracking has not been used much to study the processing of and learning from expository texts, I am also aware of some of the reasons for this shortage. One has to do with the fact that the method provides a very rich data set that may be highly challenging to analyze and make sense of. A similar challenge is faced by researchers applying the method to multimedia learning. However, as it becomes evident from this special issue, the challenge can be successfully met.

My commentary is structured as follows. First, I present some background to eye movement research for readers not familiar with the method, followed by a brief discussion of the limitations of the method. The primary focus, however, is on the choice of eye movement measures employed in the present set of studies. In this section I make suggestions for how such
data may be analyzed in further detail to tap into the time course of the learning process.

2. Some general notes on eye movement research

The present third era of eye movement research (see Rayner, 1978, 1998) is characterized by the use of human eye movements to index mental processes that are ongoing when people interact with different types of visual environments (e.g., written texts, illustrations, human facial expressions, or traffic scenes). In doing so, researchers have subscribed to the so-called eye-mind hypothesis (Just & Carpenter, 1980), according to which there is a close link between the direction of human gaze and the focus of attention. In other words, it is assumed that people attend to and process the visual information that is currently looked at. Naturally, in order to this assumption to hold, the available visual environment needs to be relevant to the task at hand. While I am writing this text, I find myself looking out of my office window at the library facilities located opposite to my office. However, my mind is busy finding a good formulation to my thoughts, so my gaze behavior is not reflecting what I am attending to at the moment. With this example I hope to illustrate the point that gaze behavior can serve as an index of current attentional processes only as long as the available visual environment in front of our eyes is pertinent to the task we would like to study.

The third era of eye movement research began soon after the cognitive revolution in psychology. The research was further boosted by the availability of microcomputers and commercially available recording apparatuses (see McConkie, 1997, for a reminiscence of the early days of eye movement research). Recent technological development has made the devices increasingly user-friendly: eye-trackers are relatively easy to operate, they are often unobtrusive to the participant, and ready-made analysis software packages greatly help to make sense of rich data sets. In the earlier days, researchers needed to write their own software to collect and analyze data; thus, the methodology was available only for the most devoted ones.

When people interact with visual environments, they make a sequence of fixations separated by fast eye movements (so-called saccades that are the fastest motoric movements human beings can make). Intake of visual information takes place during fixations, while saccades bring the center of the eyes (fovea) to new locations in the visual scene. Depending on the visual task and the momentary processing difficulty, individual fixations typically last about 200–500 ms. There exists now ample evidence demonstrating that increased processing difficulty is capable of inflation the duration of individual fixations. Moreover, fixation density may also be affected (i.e., the number of fixations is greater on a stimulus that is difficult to encode and/or comprehend). As evidenced by the present set of studies, also relevance assignment influences eye behavior. More fixation time is devoted to task-relevant stimulus features than task-irrelevant features (see, e.g., Kaakinen, Hyöniä, & Keenan, 2002, for a similar finding in the processing of expository texts). Finally, as rightfully pointed out in several papers, the eyes are also drawn to visually salient features in our environment. For example, abrupt onset of stimulus, motion, and stimulus brightness are features capable of attracting the eyes. In sum, the eyes are guided both endogenously (i.e., to meet the learner’s task-relevant goals) and exogenously (i.e., by perceptually salient stimulus characteristics). The present set of papers is to be praised for taking seriously into consideration both types of factors (see also, e.g., Lowe, 2003).

3. Limitations of the eye-tracking method

It is important to note the limitations of the eye-tracking method. Even though it provides highly valuable information (i.e., about what is perceived as task-relevant) it does not as such tell the researcher anything about the success or failure of comprehending the relevant piece of information. The learner may spend a lot of time attending to a relevant stimulus feature without adequately comprehending its relevance or the underlying principle it denotes (e.g., the learner may be looking at visually cued features in an animation of the workings of the human cardiovascular system without necessarily comprehending the operation of the cued subsystem; see De Koning, Tabbers, Rikers, & Paas, 2010). Thus, the eye-tracking data must be complemented with other performance measures, such as retrospective comprehension tests or think-aloud protocols. The studies included in the special issue are excellent examples of this complementary approach (see also Kaakinen & Hyöniä, 2005, for an example of such an approach applied to the study of text comprehension).

4. Complementing eye-tracking with offline measures

An interesting and innovative approach in combining eye-tracking with a retrospective report on the comprehension process is used by (Jarodzka, Scheiter, Gerjets, and Van Gog, 2010; see also De Koning et al., 2010). They showed the learners their eye movement patterns registered when they viewed videos of different types of swimming fish in order to determine the fish’s locomotion pattern. Using the eye movement pattern as a memory cue learners were asked to verbally report their thought contents while they were viewing the animation. The idea is that the played-back eye movement pattern may cue the learners to recover how they encoded and interpreted the various stimulus features. This sounds an intriguing approach definitely worth trying out in future studies. To further improve the effectiveness of the cue, perhaps one could use this “cued retrospective reporting” after each stimulus presentation, rather than at the end of the study. The downside of this could be that learners become increasingly aware of their eye movements, which may influence the viewing of subsequent animations. On the positive side, a frequent exposure to learners’ own eye movement patterns may increase their intrinsic value as an effective memory cue.

Playing back eye movement data to the participants may also be an informative and useful tool to teach efficient
scanning strategies. For example, when teaching airport security personnel how to detect dangerous items in X-ray images of air passenger luggage (Liu, Gale, & Song, 2007), playing back the eye movement patterns registered during image scanning may be highly revealing.

5. Eye movement measures used in the present studies to investigate multimedia learning

As is apparent from Table 1 of Mayer’s (2010) commentary, the primary eye movement measure employed in all the studies of the special issue is total fixation time on pre-specified areas of interest. As the name indicates, total fixation time reveals how much time is spent fixating, for example, on visualization versus written text when learning the principles related to the formation of lightning (see Schmidt-Weigand, Kohnert, & Glowalla, 2010). It is a useful measure, as it indicates how learners allocate their visual attention during the entire learning trial. However, it does not tell us how learning proceeds on a moment-to-moment basis. For example, learners may first read the entire text before studying the visualization (this is in fact how elementary school students study biology textbooks, see Hannus & Hyönä, 1999). Thus, other, more detailed eye movement measures are needed to complement the global picture derived from total fixation time. In the present set of studies, this does to some extent, but I think additional measures could be added to the researcher’s arsenal. It is often the case that methodological innovations pave the way for subject-matter innovations. Thus, in order to increase the usefulness and popularity of the eye-tracking method in the field of multimedia learning, additional useful and informative measures are called for. In the following, I first go over measures suggested in the present set of studies, followed by my own reflection on other possible measures.

When examining the allocation of visual attention between text and visualization, Schmidt-Weigand et al. (2010) computed the number of transitions between text and visualization. This measure provides an index of the frequency with which learners shift their visual attention between the two information media. Thus, if the learners would behave similarly to the elementary-school students studied by Hannus and Hyönä (1999), there would often be only one transition from text to visualization at the end of text reading. This did not turn out to be the case, however. Depending on the stimulus presentation time, learners made about two to four transitions (Experiment 1). Given that the texts appeared to be pretty short, one may argue that four transitions imply a reasonable degree of interplay between the two information media. It should be noted that the number of transitions measure does not indicate the exact timing of the attentional shifts. To remedy this, Schmidt-Weigand et al. (2010) analyzed the locations of the first five fixations made on the stimulus slide and showed that most of them fell on the text, which indicates that learners prioritize text over visualization (see also Hannus & Hyönä, 1999; Hegarty & Just, 1993; Rayner, Rotello, Stewart, Kéir, & Duffy, 2001). Thus, this aspect of the Schmidt-Weigand et al., this issue study is a good example of how a more fine-grained analysis procedure adds “more flesh to the bones”.

To study the similarity of scanpaths among experts and novices viewing fish locomotion animations, Jarodzka et al. (2010) applied the so-called Levenshtein’s distance measure to their eye movement data. It is a count of “the minimal number of edit operations needed to transfer one string into another” (p. 150). The smaller the number, the greater the similarity is between two scanpaths. The analysis revealed the scanpaths to be more heterogeneous between experts than novices. A similar algorithm was used by Brändt and Stark (1997) who compared the eye movement scanpaths made during picture viewing and when creating a visual image of a previously presented picture (see also Johansson, Holsanova, & Holmqvist, 2006, who analyzed relative rather than absolute similarity between viewing and mental imagery). Apart from comparing different learner groups, these similarity measures may be used in assessing effects of instruction (Canham & Hegarty, 2010), animation speed (Meyer, Rasch, & Schnitz, 2010), or visual cues (Boucheix & Lowe, 2010; De Koning et al., 2010). All in all, these similarity measures are likely to be very useful.

6. Other possibilities for data analysis

An attractive feature of the eye-tracking method is that it provides an online protocol of the encoding and attentional processes carried out during the learning phase (i.e., what is attended first and for how long, what is attended next and for how long, how much switching of attention is done between different components of the learning materials, what components are linked together during attentional switching, etc.). As mentioned above, total fixation time does not capture in minute detail the learning process, so other eye movement measures are needed to complement the picture derived from the global measure. To illustrate how eye-tracking data can be used to measure the time course of processing, I give an example from reading research (for further details see, e.g., Rayner, 1998). When studying word identification during reading via eye movement recordings, the researchers have delineated three stages in the processing time course: processing taking place prior to fixating the critical word, processing done during its fixation, and processing taking place after the immediate encounter with the word. For each stage, an eye movement measure is defined that captures the processing done during that point in time. As these measures are not directly applicable to multimedia learning, I do not go over them here. Rather, I try to sketch in the following some possibilities of tapping the time line of multimedia learning.

---

1 Schmidt-Weigand et al., (2010); see also De Koning et al., (2010) refer to attentional switching between text and illustration as splitting of visual attention. I understand this term is in frequent use in the field. I prefer the use of attentional shifting or switching to that of attentional splitting, as the former terms clearly imply the serial nature of processing (one can only attend either to text or illustration at one time), whereas the latter term may unintentionally and wrongly imply parallel processing (i.e., splitting the attention between text and illustration).
In what follows, my goal is to point out some possible ways researchers interested in applying eye-tracking to study multimedia learning may gain additional insights into the learning process as it evolves over time. I am not in a position to provide any detailed list of useful measures. Instead, I invite researchers interested in the domain of multimedia learning to come up with sensitive and innovative eye movement measures that are tailor-made to suit the specific needs in the field. As I argued above, it is important to keep in mind that methodological innovations contribute in a very significant way to the advancement of science.

To make my suggestions more concrete I present them in reference to the studies published in this special issue. Thus, my aim is not by any means to criticize these studies, but simply sketch potential additional analysis procedures. I start out with the two studies examining effects of visual cueing on attentional guidance during learning. Boucheix and Lowe (2010) studied learning of a piano mechanism illustrated by an animation where the different components of the piano were visually signaled by spreading color cues. During learning, participants spent more time and made more fixations on cued than non-cued components. It would be interesting to find out how fast and faithfully learners “obey” the presented cues. This could be done by analyzing the fixation frequencies and durations time-locked to the appearance of each new cue. The speed with which a cue is “ obeyed” may be assessed by measuring the time it takes from the cue onset until the first fixation falls on the cued location (provided that the cued location is not fixated at the cue onset). The effectiveness of the cue may be estimated by the percentage of time spent fixating on the cued versus the non-cued location from the cue onset until the cue offset (when one cue is replaced by another cue). For the cue to be effective (i.e., that it maximally engages attention), it needs to grab and keep attention in the presence of other visually highly salient stimulus features that also possess the capacity to automatically capture attention (and the eyes).

A similar set of analyses time-locked on the cue onset and offset may be conducted on the data of De Koning et al. (2010). Moreover, as some of the cued components of the cardiovascular system are spatially not tightly unified and some components appear superimposed on each other, additional measures may also be needed. One option could be to compute how widely or closely fixations are spread from each other. When scanning a cued component, the consecutive fixations may be positioned relatively close to each other (or the traversed distance between consecutive fixations is rather small) whereas, when no visual cue is provided, the pattern of fixations may be more wide spread.

In the study of Jarodzka et al. (2010), experts in marine biology and novices (biology students) viewed videos of different types of swimming fish in order to identify the specific locomotion pattern. The eye movement analyses were computed for the entire animation (they lasted 8–9 s on average) as well as for the first 4 s. As learners are likely to make three to five fixations per second, even a more detailed analysis may be in place here. It would be interesting to know what is guiding the eyes during the early stages of video (or animation) viewing. For example, perhaps all experts first pay attention to the features relevant to species classification, after which their gaze behavior becomes more idiosyncratic, whereas novices may first pay attention to irrelevant features but only in the last stages to features relevant to locomotion classification. This pattern of results is mere speculation, the important point here is that a more detailed time-locked analysis carries the potential of providing additional relevant information about how expertise influences stimulus encoding. To illustrate this point with real data, Liu et al. (2007) found experts to fixate faster on target areas (dangerous items in an X-ray image of air passenger luggage) than naïve observers.

A similar suggestion may be made with respect to the study of Canham and Hegarty (2010), who examined the effects of instruction of two meteorological principles on estimating the direction of wind on meteorological maps. The study showed that newly acquired knowledge helped learners to devote more attention to task-relevant and less attention to task-irrelevant information. The degree of expertise (i.e., carrying out the task with or without instruction) may also determine how quickly learners find the relevant information in a meteorological map (see above the description of Liu et al., 2007). Or, put it another way, with no prior knowledge map features that are visually salient but non-informative (e.g., the color-coded temperature information) may be attended to during the earliest processing stage (on average the maps were viewed for about 5–6 s) but not during later stages — Canham and Hegarty (2010) found no overall difference in the allocation of attention to irrelevant information between the map versions that contained or did not contain the irrelevant temperature information. These kinds of questions may be answered by examining eye movement measures that are sensitive to the time course of learning.

Finally, Meyer et al. (2010) investigated the effectiveness of dynamic animations on learning the workings of a four-stroke engine. They were particularly interested in the effects of the animation’s presentation speed. They reasoned that a slow versus a fast rate may differentially highlight the micro- versus macro-events. In Experiment 2, they observed that, irrespective of the animation’s presentation speed, learners focused more on micro-events during the first animation, while the viewing of the second animation was affected by the playing speed. The two animations were played for 2 min each, and the above summary is based on the overall viewing data. Again, it would be of interest to examine in more detail the viewing as a function of time. For example, is the scanpath observed, say during the first 20 s, qualitatively different from that observed for the last 20 s? Does the presentation speed of animations modulate the potential difference between the early versus late viewing?

7. Concluding remarks

The present special issue eloquently illustrates the usefulness of the eye-tracking method in the study of multimedia learning. As pointed out above, eye-tracking provides an
online protocol of attentional allocation during the learning phase. Eye movements are necessary and essential during learning of visually presented materials; thus eye movement registration can be employed without introducing any additional task extraneous to the actual learning itself. Although highly useful, the method must be complemented with offline measures tapping into the end product of learning. By combining online (eye-tracking) and offline (comprehension tests, verbal protocols) measures the researcher is in a position to tease apart, for example, the extent to which a learning failure is a result of inadequate intake and encoding of relevant features of the learning materials. The present studies provide excellent examples of such a complementary approach.

References