Augmented Learning on Anticipating Textbooks with Eye Tracking

Shoya Ishimaru, Syed Saqib Bukhari, Carina Heisel, Nicolas Großmann, Pascal Klein, Jochen Kuhn, and Andreas Dengel

Abstract

This paper demonstrates how eye tracking technologies can understand providers to realize a personalized learning. Although curiosity is an important factor for learning, textbooks have been static and constant among various learners. The motivation of our work is to develop a digital textbook which displays contents dynamically based on students' interests. As interest is a positive predictor of learning, we hypothesize that students' learning and understanding will improve when they are presented information which is in line with their current cognitive state. As the first step, we investigate students' reading behaviors with an eye tracker, and propose attention and comprehension prediction approaches. These methods were evaluated on a dataset including eight participants' readings on a learning material in Physics. We classified participants' comprehension levels into three classes, novice, intermediate, and expert, indicating significant differences in reading behavior and solving tasks.

Keywords

Attention; Augmented Text; Comprehension; Didactics; Eye Tracking; Learning; Reading Behavior; Physics.

O. Zlatkin-Troitschanskaia et al. (Eds.), *Positive Learning in the Age of Information*, https://doi.org/10.1007/978-3-658-19567-0_23

1 Objective

Curiosity is an important factor for learning. Every human has a different way of learning based on individual speed and preferences. However, teaching activity has been, traditionally, static and consistent among various learners. We assume that the system which provides individualized information for each learner based on their interests can foster positive learning (cf. Zlatkin-Troitschanskaia et al. 2017). This paper demonstrates how technologies can provide such kind of personalized learning. Since textbook has played an important role in learning and education, we propose the concept of "Anticipating Textbook," which displays the information need based on gaze, i.e. using eye tracking devices to measure visual attention and employ them for vivid interaction with textual information. In order to develop the system, it is necessary to predict the timing when learners have or lose their interest on the content in real time. If, for instance, learners are overwhelmed by difficult learning content misconceptions may occur, and the system intervenes (e.g., showing illustrative videos, switching to less complex representations, etc.); if readers need additional data, instructional support or more advanced information, the anticipating textbook reacts accordingly by presenting this kind of information. It is estimated that about 80% of all knowledge stored in memory is captured via the eyes (Murphy 2016). Gaze can be interpreted as a proxy for the user's attention, and eye movements are known to be usually tightly coupled with cognitive processes in the brain, so that a great deal about those processes can be observed using eye tracking (Dengel 2016). We propose attention and comprehension prediction approaches by measuring students' reading behaviours. The objectives of this paper are to present 1) the concept of the anticipating textbook and 2) attention and comprehension prediction methods while reading.

2 Theoretical framework

Tracking eye movements on text has a long history. In first experiments conducted during the 19th century, subjects reading text were monitored with the simplest means and the findings were basically of descriptive nature. Javal (1878), Landolt (1881) and Lamare (1892) were among the first to conduct eye tracking studies on text (Wade and Tatler 2009). While early experiments were of rather descriptive nature and provided early evidence that the eye moves in a series of jerks (i.e., saccades) while reading, the second half of the 20th century started to focus on cognitive aspects. Especially during the last thirty years the available tracking methods improved dramatically and with the availability of remote eye tracking devices and a computer-based evaluation of eye movements there was a remarkable increase in insights into the human perception and reading process (Rayner 1998, p. 372). Sophisticated experiments could be performed with gaze-contingent stimuli, based on the subject's eye movements and behavior. Furthermore, the first truly interactive eye tracking applications were implemented (e.g., Bolt 1990) in which eye tracking was used for entertainment applications. However, the real-time usage of gaze on text, for the sake of education and training or information provision, has not explicitly been considered for a long time. The first application focusing on that aspect was iDict by Hyrskykari et al. (2000). The system was implemented to provide translations on comprehension problems detected in the reader's gaze patterns. In several papers, we presented an algorithm for online reading detection based on eye tracking data (Biedert et al. 2012) and introduce an application for assisted and augmented reading called the eyeBook (Biedert et al. 2010). The idea behind the eyeBook is to create an interactive and entertaining reading experience and to help the reader to better understand the text and what is behind. Eye tracking systems observe which text parts are currently being read by the user not only on the screen but also on paper (Kunze et al. 2013, Ishimaru et al. 2016, and Toyama et al. 2013).

Considering the above work around eye tracking, we apply the approach of augmented text to educational textbook. Figure 1 shows a concept sketch of the anticipating textbook. The system recognizes a student's cognitive state (e.g., attention, interest, comprehension) using several sensors including an eye tracker. Then the system changes the content or the layout dynamically to improve a student's motivation and understanding. For example, playing a video instead of showing a static picture should attract students' interest. Since students prefer different representations depending on their skill level (cf. Klein et al. in this volume), the system displays the adapted representation based on cognitive state analysis. If the system tracks the level of understanding while reading, it can pick up or generate tasks a student should solve to correct his/her misunderstanding.

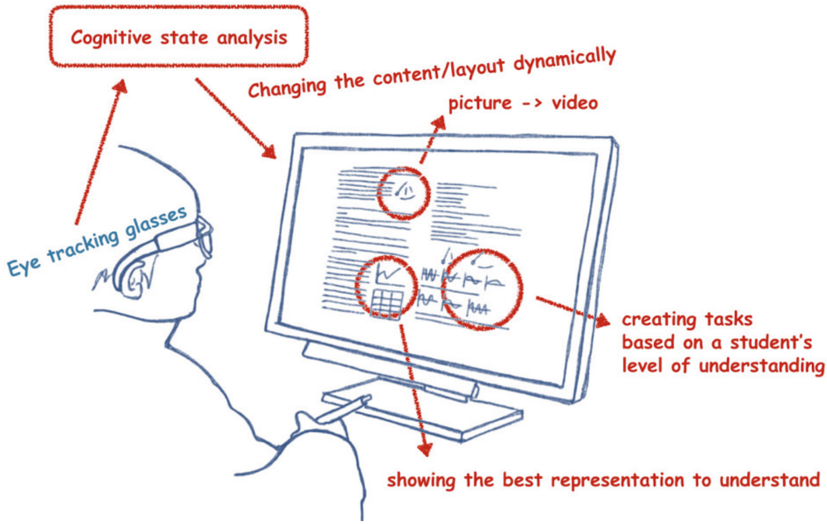


Fig. 1 The concept sketch of the anticipating textbook

3 Methods

In order to implement the anticipating textbook, we start from investigating students' cognitive states while reading a textbook. In following, this paper presents our attention and comprehension extraction methods. As preprocessing, raw data from eye tracking glasses are converted to gaze points on a document with a projection function based on SIFT features (Lowe 1999) and classified into fixations and saccades (Buscher et al. 2008).

3.1 The AOI based attention extraction

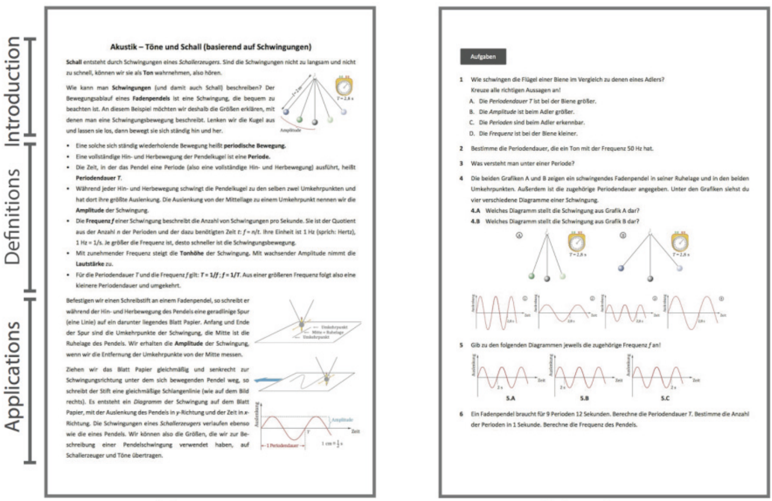
We divide a text beforehand based on the roll (e.g., the introduction, definitions, applications on the document shown in Figure 2) then focus on the period of time needed to read the content to obtain knowledge. Thus, for each area a sum of fixation durations is calculated, which is divided by the size of area to be normalized.

3.2 The AOI based comprehension prediction

We apply a support vector machine (SVM) to predict students' comprehension. On the basis of AOI based fixation duration described as above, each duration in AOIs are calculated as features. From the document in Figure 2, for example, three features (durations on the introduction, definition, and application) are used. Since this method requires a student's reading behavior from the beginning to the end of a document, it can only be applied as an offline analysis.

3.3 The subsequence based comprehension prediction

On the other hands, an online analysis is required in order to change the content dynamically while reading. Therefore, we also investigate whether a subsequence (e.g., 1 minute of reading) is enough useful to predict students' comprehension. In this approach, we calculate four features (mean and standard deviation of fixation durations and saccade lengths) in a subsequence and apply SVM based classification.



(a) text (b) tasks

Fig. 2 A document with text and tasks in physics

These two figures are in one page on a display (text on the left and tasks on the right) during the experiment.

4 Data sources and the experimental design

We asked 8 participants to wear eye-tracking glasses, to read a physics textbook and to solve respective exercises as shown in Figure 3. The participants were 6-grade students at a German high school (11 or 12 years old). The document we prepared is “Basic Phenomena in Acoustics” (cf. Figure 2). It consists of four parts: the introduction, itemized definitions, applications, and related tasks. Only an explanation of about the content (the left page in Figure 2) was displayed at first. After participants understood the content, they could make tasks appear by pressing a key. They could go back to read the content to help them in their solving tasks. In this paper, we define these two steps as “reading” and “solving.”

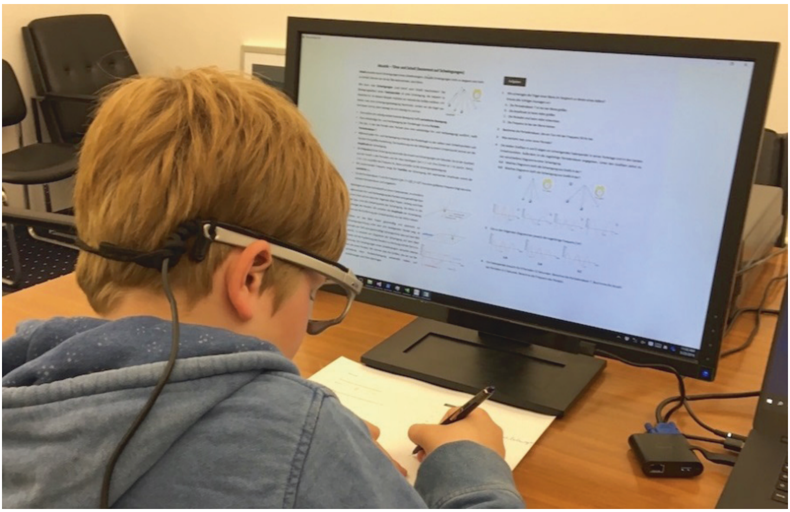


Fig. 3 An overview of the experiment
A participant is solving questions on a display with wearing SMI Mobile Eye Tracking Glasses 2.

To evaluate whether our proposed method works with different eye tracking devices, two types of eye tracking glasses were used during the experiment. We used *Tobii Pro Glasses 2* with five participants (*a, b, d, e, f*). The glasses record eye gaze at a sampling frequency 100 Hz and a scene video at 25 Hz. We applied one-point calibration with a marker before starting each recording. The data of the other three participants (*c, g, h*) were recorded with *SMI Eye Tracking Glasses 2*. The glasses record eye gaze at a sampling frequency 60 Hz and a scene video at 30 Hz. We apply three-point calibration with this device.

For evaluations of the comprehension prediction methods, training and testing dataset was created by leave-one-subject-out. All data from one participant are used for testing and data from other participants are used for training.

5 Results and discussion

5.1 The attention extraction

Table 1 shows percentages of time participants paid attentions for the introduction, definitions, and the applications on the document. We calculated the percentage depending on each situation while reading a text and solving tasks. The data in Table 1 is sorted by the number of correct answers. We categorized 8 participants to 3 comprehension levels based on their scores: novice (the score is 4 or less), intermediate (the score is 5), and expert (the score is 6 or more).

Tab. 1 Percentages of time participants paid attentions

Participant	Score (out of 14)	Expertise	Attentions while reading [%]			Attentions while solving [%]		
			Intro.	Def.	Appl.	Intro.	Def.	Appl.
a	3	Novice	14	49	37	13	59	28
b	4	Novice	17	43	40	17	48	35
c	5	Intermediate	7	51	42	4	44	52
d	5	Intermediate	31	41	28	21	49	30
e	5	Intermediate	23	37	40	27	40	33
f	6	Expert	16	47	37	12	60	28
g	7	Expert	34	50	16	25	56	19
h	7	Expert	28	64	8	22	70	8

By calculating mean values for each comprehension level, it has become obvious that students with high-level comprehension do not pay attention to the applications part while both reading and solving tasks compared to other levels (cf. Figure 4 and Figure 5). They understand that the applications part is useful for understanding the content, yet there is not much information that can be used as hints for solving tasks. They preferred to read definitions part because there are direct hints (principles, formulas, etc.). Intermediates and novices spend much time to paying attention to the application part while both reading and solving.

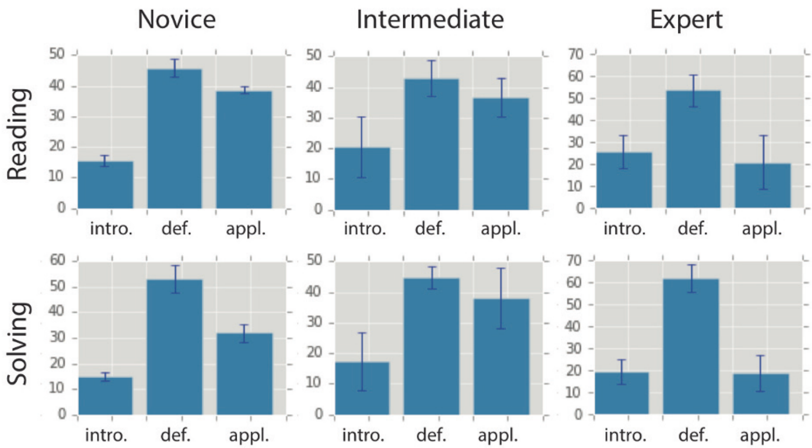


Fig. 4 Histograms of the time students paid attentions [%]
Error bars represent standard deviations.

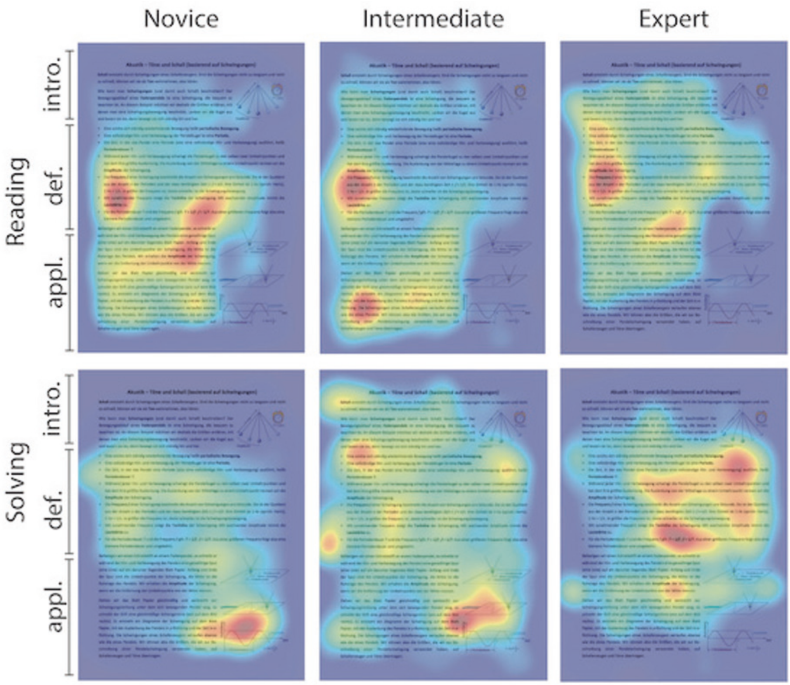


Fig. 5 Fixation duration based heat maps while a student is reading the text and solving tasks

5.2 The comprehension predictions

By using the categories (novice, intermediate, and expert) as ground truth, we estimated participants' completions. Figure 6 and Figure 7 represent confusion matrices of the estimation results. The AOI based approach succeeded to estimate all completions of participants. The estimation accuracy of the subsequences based approach was 70%.

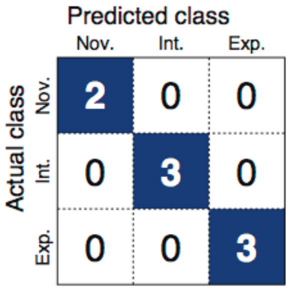


Fig. 6 The AOI based prediction result

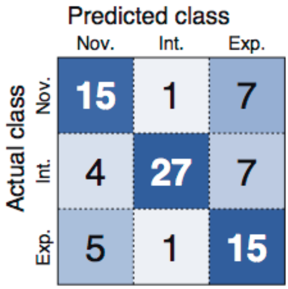


Fig. 7 The subsequence based prediction result

Figure 8 shows all participants' feature plot in subsequences based approach. The higher the participant's completion is, the larger mean saccade length is measured. Novice students read a textbook with large fixation duration and small saccade length, and intermediate students read with small fixation duration and large saccade length. It cleared that novice students read a textbook slowly with small steps. The distribution of data plot from experts is larger than others. Expert students sometimes skip their eyes on the text, focus on the content they are interested in.

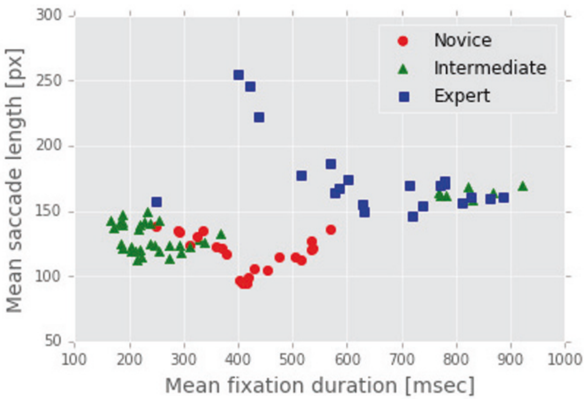


Fig. 8 Feature representation of all participants' data in subsequences based approach Each dot represents a data segment of one minute

6 Scientific significance of the study

In this paper, we present an initial method to extract students' attention by using gaze data. By applying the approach to activities including reading a text and solving tasks, it is revealed that reading behavior is related to students' comprehension. Expert students, for example, tend to pay attention on definition part to understand the content. In a next step, this information can be used to foster positive learning, for example, by giving visual cues to novice or intermediate students to identify relevant text passages for problem-solving. We also predicted students' completion (ground truth was calculated by the score of tasks) with two approaches. One is attentions on AOI based, and the other is features from gaze subsequence based prediction. The former one works better than the later one, but it requires the recording of reading from beginning to end. We found that features from a window of gaze data in one minute can enough classify students' completion into three classes with 70% accuracy. These results serves as a basis for on-line classification of learning states which can be used in a follow-up study to automatically address individual learning groups with tailored content.

Bibliography

Biedert, R., Buscher, G., & Dengel, A. (2010). The eyebook – using eye tracking to enhance the reading experience. *Informatik-Spektrum, 33*(3), 272–281.

Biedert, R., Hees, J., Dengel, A., & Buscher, G. (2012). A robust realtime reading-skimming classifier. In S. N. Spencer (Ed.), *Proceedings of the 2012 Symposium on Eye Tracking Research and Applications* (pp. 123–130). Santa Barbara, CA: ACM.

Bolt, R. A., & Starker, I. (1990). A gaze-responsive self-disclosing display. In J. C. Chew & J. Whiteside (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3–10). New York: ACM. doi:10.1145/97243.97245

Buscher, G., Dengel, A., & van Elst, L. (2008). Eye movements as implicit relevance feedback. In M. Czerwinski & A. Lund (General Chairs), *CHI' 08 Extended Abstracts on Human Factors in Computing Systems* (pp. 2991–2996). New York: ACM.

Dengel, A. (2016). Digital co-creation and augmented learning. In L. Uden, I-H. Ting & M. Santos-Trigo (Eds.), *Proceedings of the 11th International Knowledge Management in Organizations Conference on the changing face of Knowledge Management Impacting Society* (Art. No. 3). New York: ACM. doi:10.1145/2925995.2926052

Hyrskykari, A., Majaranta, P., Aaltonen, A., & Räihä, K.-J. (2000). Design issues of iDICT: a gaze-assisted translation aid. In A. T. Duchowski (Chairman), *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications* (pp. 9–14). New York: ACM.

Ishimaru, S., Kunze, K., Kise, K., & Dengel, A. (2016). The wordometer 2.0: estimating the number of words you read in real life using commercial EOG glasses. In P. Lukowicz & A. Krüger (General Chairs), *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 293–296). New York: ACM.

Kunze, K., Kawaichi, H., Yoshimura, K., & Kise, K. (2013). The wordometer – estimating the number of words read using document image retrieval and mobile eye tracking. *2013 12th International Conference on Document Analysis and Recognition,* 25–29.

Lowe D. G. (1999). Object recognition from local scale-invariant features. In IEEE (Ed.), *The proceedings of the seventh IEEE international conference on computer vision* (pp. 1150–1157). IEEE Computer Society. doi:10.1109/ICCV.1999.790410

Murphy R. (2016). Learning-related vision problems. Allaboutvision.com. http://www.allaboutvision.com/parents/learning.htm. Accessed: 12 August 2017.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin, 124*(3), 372–422.

Toyama, T., Suzuki, W., Dengel, A., & Kise, K. (2013). User attention oriented augmented reality on documents with document dependent dynamic overlay. *In IEEE (Ed.), 2013 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH 2013)* (pp. 299–300). Institute of Electrical and Electronics Engineers.

Wade, N. J., & Tatler, B. W. (2009). Did javal measure eye movements during reading? *Journal of Eye Movement Research, 2(5),* 1–7.

Zlatkin-Troitschanskaia et al. (2017). *Positive Learning in the Age of Information*. Unpublished Manuscript, Draft Proposal Cluster of Excellence, Johannes Gutenberg University Mainz.