# Towards an Intelligent Textbook: Eye Gaze Based Attention Extraction on Materials for Learning and Instruction in Physics

**Shoya Ishimaru**
German Research Center for Artificial Intelligence
Kaiserslautern, Germany
Shoya.Ishiamru@dfki.de

**Jochen Kuhn**
University of Kaiserslautern
Kaiserslautern, Germany
kuhn@physik.uni-kl.de

**Syed Saqib Bukhari**
German Research Center for Artificial Intelligence
Kaiserslautern, Germany
Saqib.Bukhari@dfki.de

**Andreas Dengel**
German Research Center for Artificial Intelligence
Kaiserslautern, Germany
Andreas.Dengel@dfki.de

**Carina Heisel**
University of Kaiserslautern
Kaiserslautern, Germany
heisel@physik.uni-kl.de


*Ubicomp/ISWC'16 Adjunct*, September 12–16, 2016, Heidelberg, Germany.

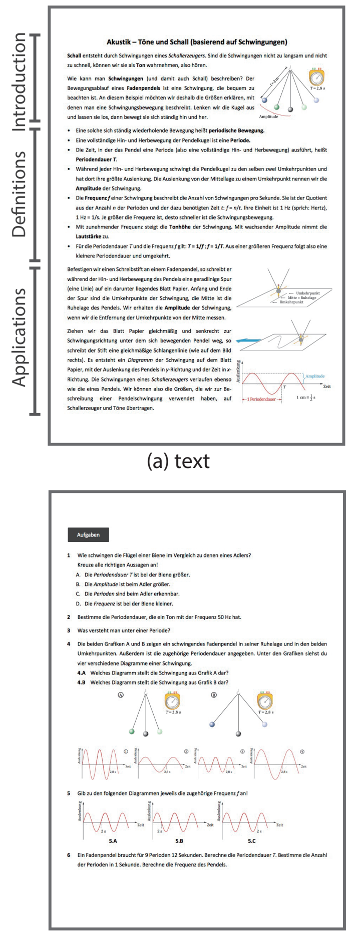
ACM 978-1-4503-4462-3/16/09 $15.00.
DOI: http://dx.doi.org/10.1145/2968219.2968566

## Abstract
In this paper, we propose the attention extraction method on a textbook of physics by using eye tracking glasses. We prepare a document including text and tasks in physics, and record reading behavior on the document from 6-grade students. The result confirms that students pay attention to a different region depending on situations (using the text as a material to learn the content, or using it as hints to solve tasks) and their comprehension.

## Author Keywords
Eye tracking; attention; education; document analysis

## ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

## Introduction
Reading a textbook is an important way to obtain new knowledge. However, students often avert their eyes from reading a textbook because it is static and boring. One of the solutions to this problem is to develop a digital textbook which can make the materials for learning and instruction adaptive and dynamic on a mobile tablet computer. Especially in physics education, it is highly efficient to show phenomenon, experiments, representations (such as text, diagram, formulas, etc.), and 3D models as dynamic con-



(a) text

(b) tasks

**Figure 1:** 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.

tents. Furthermore, the combination of a digital textbook and activity recognition (e.g. recognizing students' interest, workload, and trouble with understanding) makes it possible to provide dynamically-generated content individually optimized for each student and context. Our final goal is to create such an "anticipating" textbook and enhance students' learning performances.

As an initial analysis, we investigate students' reading behavior on a textbook in order to find specific patterns which are related to the context including situations and comprehensions. For this purpose, we prepared a document on "Basic Phenomena in Acoustics" (cf. Figure1) including both, text and related exercises. In addition, we recorded their eye gaze during the period when only the text page is shown and subsequently when both, texts and exercises, are shown on a display.

The contributions of this paper are two-fold. 1) We find some rough relations between levels of students' expertise and their reading behavior on a textbook. 2) We recorded the data from 6-grade students at school. It realized some limitations to using eye-tracking devices in a realistic educational scenario.

## Related work

There is a whole bunch of research investigating efficient visualizations and representations to improve students' skills of understanding and solving in Physics Education Research [7]. However, most of them obtain students' insights from only answering sheets afterwards and do not care about their learning process while reading texts and solving tasks.

One of the interesting approaches to get insights into cognitive behavior is employing gaze data. The strong relationship between cognitive states and eye movements is well explored in the field of activity recognition and psychology [3]. Concerning reading behavior, there is a research focusing on English skill level estimation from eye gaze [10]. As the authors employ general features from eye (e.g. the number of blinks, average distance of each eye movements), their approach can be used on a physics textbook as well.

Therefore we can use preliminary research from Mozaffari et al. [8], where students' eye gaze is recorded while solving tasks of physics by using a remote eye tracker mounted to a tablet computer. The authors have revealed that students prefer different representations (vector, table and diagram) depending on their skill level. We follow their basic idea. But compared to their work, we are interested in students' natural reading behavior on a textbook. We do not optimize the text and tasks so much for the recording.

Recently eye tracking technologies became more and more affordable so that it works on an unmodified tablet computer without external hardware [1], but also as an inexpensive head-mounted wearable device [9]. Therefore, in future applications using eye tracking will work on many devices and scenarios, which is not limited to reading a textbooks [4, 6].

## The Attention Extracting Method

We believe that extracting the part where students pay attention is the first step to investigate their reading behavior. We propose a method to extract attention by using an eye tracking device. The method consists of three steps.

*(1) Mapping eye gaze coordinates on a document*

We utilize mobile eye tracking glasses and record a student's reading behavior. Because the output of the device are coordinates of eye gaze on a scene camera, we need to map them to the document under consideration. As shown in Figure 2, we detect the position of the document in a

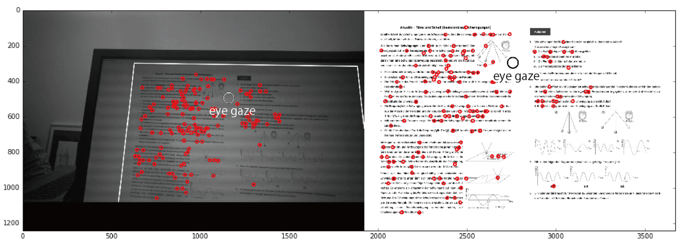
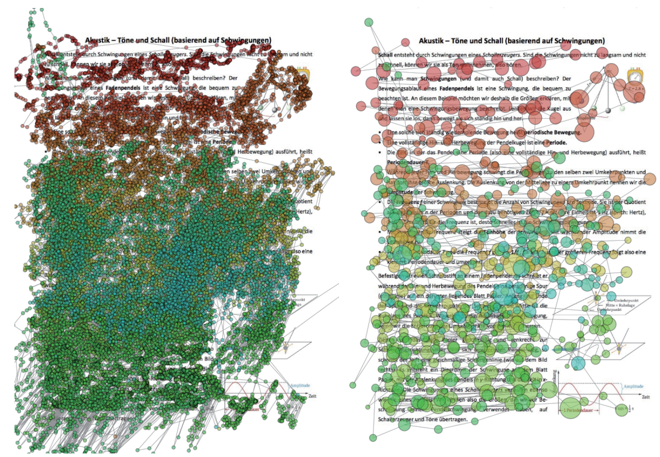



**Figure 2:** One of the results of gaze mappings. Red circles of both scene and document image represents feature points of SIFT. The documents is extracted as white rectangle on a scene image. The gaze point on document is estimated as a black circle.

frame of the camera by using SIFT features [5], and calculate a homograph to map the gaze point to the document.

*(2) Detecting fixations*
The raw gaze data on the document is classified into fixations and saccades. Fixations appear, when the gaze pause in certain position - normally lasting between 200 and 400 ms. Saccades are the jumps of the gaze between two fixations taking 10-20 ms. We apply the fixation-saccade detection algorithm proposed by Buscher et al. [2]. Figure 3 shows input and output. The radius of each circle corresponds to the fixation, and the line between two circles represents a saccade. Noises and drifts on raw data are also filtered in this step.

*(3) Calculating the features for each area of attention*
We divide a text beforehand based on the roll (e.g. the introduction, definitions, applications on the document shown in Figure 1.) We focus on the period of time a student needs to reads 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.




(a) raw gaze data (b) fixations and saccades

**Figure 3:** Eye gaze on the document while a student read a textbook. Colors represents the order of eye gaze (red to blue). (a) is raw data and (b) is an output of fixation-saccade detection

## Experimental Setup

We asked 8 participants to wear eye-tracking glasses, to read a physics textbook and to solve respective exercises. The participants were 6-grade students at a German high school (around 12 years old). The document we prepared is shown in Figure 1. It consists of three parts: the introduction, itemized definitions, and applications. To analyze the students' natural behavior while acquiring knowledge, we selected content that they had not yet learned in class.

Only an explanation of about the content (the top page in Figure 1) was displayed at first. After they understood the content, they could make tasks appear by pressing the space-key on a keyboard. 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."

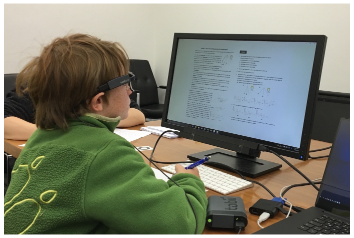


**Figure 4:** An overview of the experiment. A participant is solving questions on a display with wearing *Tobii Pro Glasses 2*.


**Figure 5:** Histograms of the time students paid attentions [%]. Error bars represent standard deviations.

Figure 4 shows the overview of the experiment. 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.

## Results and Discussion

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

By calculating mean values for each comprehension level as shown in Figure 5, 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. 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.

**Table 1:** Percentages of time students paid attentions

| P. | Score (out of 14) | Expertise | while reading text [%] | | | while solving tasks [%] | | |
|---|---|---|---|---|---|---|---|---|
| | | | 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 |

While percentages of attention in the reading and solving phase are similar for all levels, their reading behavior is deferent for each situation. Figure 6 represents fixation-based heat maps from novice (Participant *b*), intermediate (Participant *e*), and expert (Participant *h*) for both, reading and solving. High-attention part during reading is almost same for all three skill levels, while the participants spend much time to read left side of the definitions part. Note that novices and intermediates look graphics carefully while solving, and this is one of the reasons their attention for this application is higher than for experts.

Compared to a remote eye tracker, mobile eye tracking glasses have the advantage that the device can record eye gaze not only on display but everything on a scene camera. However the big issue during this experiment was that students sometimes touch glasses unconsciously, and lose an accuracy of eye tracking adjusted during calibration. Therefore using remote eye tracker might be better than mobile glasses for young participants.

## Conclusion and Future Work

In this paper, we present an initial method to extract students' attention by using gaze data. By applying the ap-

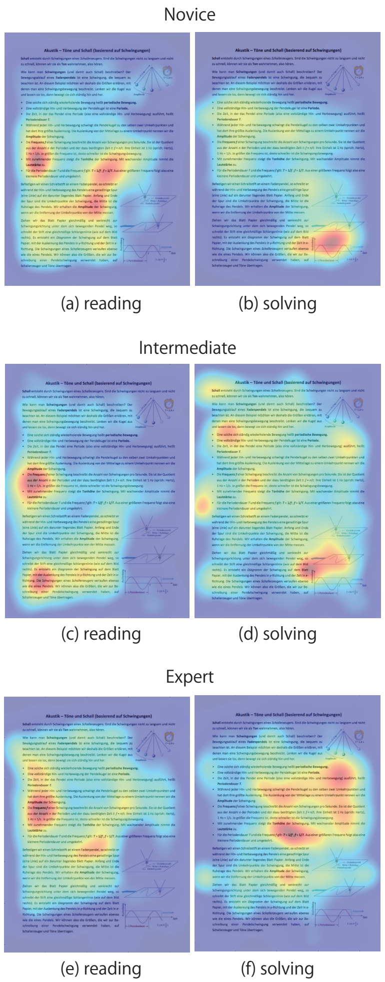



**Figure 6:** Fixation duration based heatmaps while a student is reading the text and solving tasks.

proach to activities including reading a text and solving tasks, it is revealed that reading behavior is related to students' comprehension. Especially experts do not pay attention where to find specific hints to answer questions. One of the future work is a detailed analysis of reading behavior focusing not only on the total duration on separated parts (texts and exercises) but also time-series information (e.g. re-reading contents, transition between representations).

## Acknowledgement

This work is funded by the Federal Ministry of Education and Research (BMBF) in the framework of the joint initiative "Qualitätsoffensive Lehrerbildung" of the Federal and the Federal States of Germany for the project "U.EDU: Unified Education" (support code: 01JA1616). The authors are responsible for the content of this contribution.

## REFERENCES

1. Oliver Amft, Florian Wahl, Shoya Ishimaru, and Kai Kunze. 2015. Making regular eyeglasses smart. *Pervasive Computing, IEEE* 14, 3 (2015), 32–43.
2. Georg Buscher, Andreas Dengel, and Ludger van Elst. 2008. Eye movements as implicit relevance feedback. In *CHI'08 extended abstracts on Human factors in computing systems*. ACM, 2991–2996.
3. Reinhold Kliegl, Antje Nuthmann, and Ralf Engbert. 2006. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of experimental psychology: General* 135, 1 (2006), 12.
4. Jochen Kuhn, Paul Lukowicz, Michael Hirth, Andreas Poxrucker, Jens Weppner, and Junaid Younas. 2016. gPhysics–Using Smart Glasses for Head-Centered, Context-Aware Learning in Physics Experiments. *Transactions on Learning Technologies* (2016). accepted.
5. David G Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on Computer vision*, Vol. 2. Ieee, 1150–1157.
6. Paul Lukowicz, Andreas Poxrucker, Jens Weppner, B Bischke, Jochen Kuhn, and Michael Hirth. 2015. Glass-physics: using google glass to support high school physics experiments. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 151–154.
7. David E Meltzer. 2005. Relation between students' problem-solving performance and representational format. *American Journal of Physics* 73, 5 (2005), 463–478.
8. Saleh Mozaffari, Saqib Bukhari, Andreas Dengel, Pascal Klein, and Jochen Kuhn. 2016. A Study on Representational Competence in Physics Using Mobile Remote Eye Tracking Systems, Proceedings. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. accepted.
9. Erroll Wood and Andreas Bulling. 2014. Eyetab: Model-based gaze estimation on unmodified tablet computers. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 207–210.
10. Kazuyo Yoshimura, Kai Kunze, and Koichi Kise. 2015. The Eye as the Window of the Language Ability: Estimation of English Skills by Analyzing Eye Movement While Reading Documents. In *Proceedings of the 13th International Conference on Document Analysis and Recognition*. 251–255.