

# Using Eye Tracking as Human Computer Interaction Interface

Holger Schmidt<sup>(✉)</sup> and Gottfried Zimmermann

Stuttgart Media University, Stuttgart, Germany  
schmidtho@hdm-stuttgart.de, gzimmermann@acm.org

**Abstract.** In the project *AAMS*, we have developed the e-learning platform *ALM for Ilias* as a technical basis for research in education. ALM uses eye tracking data to analyze a learner’s gaze movement at runtime in order to adapt the learning content. As an extension to the actual capabilities of the platform, we plan to implement and evaluate a framework for advanced eye tracking analysis techniques. This framework will focus on two main concepts. The first concept allows for real-time analysis of a user’s *text reading status* by artificial intelligence techniques, at any point in the learning process. This extends and enriches the adaptive behavior of our platform. The second concept is an interface framework for multimedia applications to connect to any eye tracking hardware that is available at runtime to be used as a user interaction input device. Since accuracy can be an issue for low-cost eye trackers, we use an object-specific *relevance factor* for the detection of selectable or related content.

**Keywords:** Eye tracking · e-learning platform · Adaptivity · Artificial intelligence · High-level gaze events · Real-time analysis · Human computer interaction interface · Relevance factor

## 1 Introduction

Over the last years, eye tracking has become a valid tool to conduct research studies of user interaction behavior. Usually, the eye tracking data is analyzed after runtime (not during runtime). As part of our research in the *Adaptable and Adaptive Multimedia Systems (AAMS)* project [1], we have developed the web-based e-learning platform *Adaptive Learning Module (ALM for Ilias)* [2] with real-time eye tracking support. The project evaluates methods of using eye tracking for adaptive system behavior [3] by analyzing gaze-in and gaze-out events on selected learning content [4].

With *ALM*, psychologists in our project conduct studies on educational research in online learning. They focus on research questions like “how can we adapt learning content to improve a learner’s performance via gaze tracking” [5]. Currently, *ALM*’s gaze data analysis is limited to static fixations on learning content (areas of interest) and transitions between fixations of the user’s gaze movement. In order to enhance the platform’s capabilities, we are developing a framework for the analysis of high-level gaze events. For example, the reading status of a learner reading a text passage could be “intensive reading”, “text skimming” or “text scanning”.

Also, we are looking into eye tracking as a user input device, for improving the user experience. The development of these features is done in an iterative manner, including proof-of-concept studies (e.g., usability tests, stability tests and performance tests).

This poster gives a brief overview of the relevant research plans in the *AAMS* project.

## 2 Advanced Eye Tracking Techniques

Our current e-learning platform *ALM* is limited to static gaze-in and gaze-out analysis on predefined areas of interest. Our new framework will implement advanced eye tracking techniques, with artificial intelligence components. We have developed two separate concepts to extend the platform's analysis capabilities. The first concept focuses on the automated analysis of text components. The second concept focuses on the analysis of visual multimedia content, i.e. pictures, graphics, diagrams and complex 3D models.

### 2.1 Text Analysis: Average Reading Speed Theory

The "intensive reading" status is our definition of a concentrated learning situation. The learning platform should support the learner in keeping this status as long as possible. The platform has to be able to detect if the reading status is changing. Thus the platform can adapt the learning content in order to help the reader to re-enter the "intensive reading" status.

The concept of the framework component for text analysis is based on the average reading speed of a user constantly reading a text. This indicator determines the theoretical status "intensive reading". Thus the framework can detect when the user's reading status changes from "intensive reading" status to a different status, i.e. "skimming" or "scanning".

In a first attempt to determine the average reading speed, we follow a pre-processing approach. The user has to read several types of text (altering text difficulty) in a continuous manner. For each text passage, the framework measures the reading time and calculates the deviation from the average reading speed as a factor (which is both specific for the text type, and for the learner).

### 2.2 Context Adaptation

Content adaptation is the ability of the e-learning platform to modify the presented content in dependency of the user's gaze and input actions. One research theory of the *AAMS* project is about the optimization of a user's learning process by comparing their gaze paths with pre-analyzed gaze paths of dedicated good e-learners. The aim is to maximize the learner's knowledge gain within an e-learning session [5]. For this purpose, when the learner reads text, the platform tracks their fixations on predefined areas of interest and counts them. This value is compared to threshold limits which are gleaned from reference studies to differentiate between "good" and "bad" learners (e.g. number of fixations on a

specific area of interest). If the learner moves their gaze to another learning content or clicks the continue button before the threshold limit is reached, the platform will block the presentation of new text, and hold on to the current text for a predefined timeframe to force the learner to read the text again. The goals is that the learner will adjust their personal gaze behavior to that of “the optimal learner”.

### 2.3 Text Analysis: Detection of Reading Status Change

In addition to the actual context adaptation and based on the average reading speed, the framework can detect gaze jumps (transitions) to uncommon positions on the screen in relation to the context. This detection gains its improvement from two techniques. First, the actual text needs to be classified in reading difficulties, along the lines of the reference texts for the average reading speed detection (see 2.1). This allows for a more accurate analysis of the gaze data since it takes the difficulty level of the text into account.

Second, the framework has to modify a virtual map that classifies each sector of the screen in relation to the actual fixation. I.e., from the viewpoint of the actual fixation every text passages before this fixation becomes classified as red. Only a small passage of text behind the fixation is classified as green (green indicates the passage of continuous reading). A buffer zone around the green zone is classified as orange due to possible spikes. This has to be calculated in real-time to benefit from it. With this virtual map, the framework can predict whether the upcoming reading position (next fixation) is a result of “intensive reading” or a jump or status change. Thus we are able to detect interruptions of continuous (intensive) reading. In addition, the framework is able to detect different reading statuses; hence, the framework needs to be extended to support several reading patterns like, “text skimming” [6] and “text scanning”.

### 2.4 Relevance Factor Adaptation

The reference application for the “relevance factor” concept is a 3D multimedia application which shows a car engine as full dismountable 3D model. The aim of the 3D model is to make the learner understand the individual components of an engine and how they work together. An engine has to be dismounted in a certain order, e.g., the cylinder head cannot be dismounted before the valve cover is removed. The framework needs to know the dismounting order of the components at every point in the process; we call this “the plot”. With this knowledge, we can improve the user’s experience of using eye tracking as an input mechanism by having their gaze “snap in” to those multimedia objects that are most relevant at the current time. This concept can be applied to arbitrary multimedia applications.

Therefore, every part of the dismountable engine has a “relevance factor” attached to it, which indicates the probability of this particular part to be dismounted next. Globally viewed, the relevance factor for every object is null. As the application drives through the plot, the framework decides about the relevance factor of every object in real-time. The more relevant an object is in relation to the plot, the higher the relevance

factor of this object becomes. The eye tracking framework analyses the list of relevance factors and selects the highest values as “active objects”. With an eye tracker connected, the framework can link these selected objects to the user’s gaze. This results in an automated selection of highly relevant objects as soon as the user’s gaze is within the objects nearby region.

### 2.5 Using the Framework as Input Device

In 3D environments, it is difficult to exactly determine small areas of interest for gaze tracking, especially in case of the z-axis (depth of the screen) without the information of the user’s depth focus. The “relevance factor” can be used to compensate for this problem. With the “relevance factor” concept, we can use eye tracking as an input device for selecting explicit content objects in a 3D environment.

## 3 Conclusion

We have introduced two new concepts as an extension to our *Adaptive Learning Module (ALM)* platform. The first concept is about the analysis of text reading statuses. We will employ artificial intelligence techniques as an extension to the currently implemented static local real-time analysis of gaze fixation and transition counting. The second concept uses relevance factors for objects in a story-driven multimedia environment. A prototype will allow eye tracking to be used as an additional user interface input device, with explicit content dependency, especially for 3D environments. We will implement these concepts as a prototype to conduct usability studies with students, stability tests and performance tests to proof both concepts.

As a second use case, the “text reading status” concept can be applied in certain multimedia applications with text objects. We could analyze text based tutorials (which are essential for the usability of the related multimedia application) in the same way we analyze the text components in the e-learning environment. Also, the “relevance factor” concept can be used in an e-learning environment to analyze picture content or diagrams in relation to text components. For this purpose, the text component stands for the plot, thus driving the relevance factors of the individual graphic objects to guide the user’s gaze to the most relevant graphical components.

With both concepts, we aim to improve the user’s experience in a wide range of learning and multimedia applications, and develop and validate new functionalities for our e-learning platform.

## References

1. Science Campus Tübingen, campus website, April 2015. <http://www.wissenschaftscampus-tuebingen.de/-www/-en/index.html>
2. Ilias 4 Open-Source Framework, project website, April 2015. [http://www.ilias.de/docu/-goto.php?target=cat\\_582&client\\_id=docu](http://www.ilias.de/docu/-goto.php?target=cat_582&client_id=docu)

3. Schmidt, H., Wassermann, B., Zimmermann, G.: An adaptive and adaptable learning platform with real-time eye-tracking support: lessons learned. In: Trahasch, S., Plötzner, R., Schneider, G., Gayer, C., Sassi, D., Wöhrle, N (eds.), Tagungsband DeLFI 2014, pp. 241–252. Köllen Druck & Verlag GmbH, Bonn (2014)
4. Wassermann, B., Hardt, A., Zimmermann, G.: Generic Gaze Interaction events for web browsers: using the eye tracker as input device. In: WWW 2012 Workshop: Emerging Web Technologies, Facing the Future of Education (2012)
5. Schubert, C., Scheiter, K., Schüler, A.: Viewing behavior during multimedia learning: can eye tracking measures predict learning success? In: 7th European Conference on Eye Movement Research. Lund, Sweden, August 2013
6. Geoffrey, B.D., Payne, S.J.: How much do we understand when skim reading? In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2006), pp. 730–735. ACM, New York (2008). doi:[10.1145/1125451.1125598](https://doi.org/10.1145/1125451.1125598)