Implementing Eye-Based User-Aware E-Learning

Marco Porta
Dip. di Informatica e Sistemistica
Università di Pavia
Via Ferrata, 1 – 27100 – Pavia – Italy
marco.porta@unipv.it

Abstract
We propose an e-learning scenario where eye tracking is exploited to get valuable data about user behavior. What we look at — as well as *how* we do that — can in fact be used to improve the learning process, revealing information which would otherwise remain hidden. The prototype system we are developing at the University of Pavia takes into account both the user's "emotional status" and the way learning activities are carried out, employing these data to adapt content presentation in real-time.

Keywords
E-learning, eye tracking, eye communication, emotion recognition, multimodal interaction, perceptive interfaces.

ACM Classification Keywords
H.1.2 [Models and Principles]: User/Machine Systems – human factors; H.5.2 [Information Interfaces and Presentation]: User Interfaces – input devices and strategies, interaction styles; K.3.1 [Computers and Education]: Computer Uses in Education – distance learning.

Introduction

The offer of e-learning courses is increasing at an unrestrainable pace. However, the learning experience is often perceived by the user as a one-way highly-constrained communication process, where the computer is only the mechanical device that conveys the content. What is actually absent is the "emotional" part of the interaction, which generally characterizes the relationship between teacher and learner [2,9]. We strongly think that "naturalness" is a key factor in e-learning [11,1]: a computer-mediated tutoring system provided with sensing capabilities could (at least partially) make up for the missing teacher in the flesh. In particular, eye monitoring can disclose important information concerning *what* the user is doing, as well as interesting data about *how* and *when* certain actions are being (or have been) performed.

Eye movements occur as sudden (almost instantaneous) *saccades*, followed by *fixation* periods of about 200-600 milliseconds, during which eyes are still. Eye tracking technology has evolved very rapidly in the last years. Current commercially-available devices look almost like ordinary LCD screens, and allow a relatively high freedom of movements, thus not constraining users too much in their activities. Apart from costs, which remain very high, eye trackers are now a technology at hand, that can significantly improve the way we interact with the computer.

Related work

To date, only few projects have explicitly considered eye tracking for e-learning. Among these, AdeLE (Adaptive e-Learning with Eye tracking) is probably the first, started with the main goal to dynamically capture user behavior based on real-time eye tracking [10]. Our work has several points in common with AdeLE, at least for what concerns purposes. Another interesting system is an empathic software agent interface developed to facilitate empathy-relevant reasoning and behavior, in which eye movements are used to get indications of learner interest and to provide feedback to character agents [12]. Other projects are more specifically focused on building models of user features, to properly shape the interaction. For instance, the system presented in [4] uses real time eye tracking information to try to assess student meta-cognitive behavior during the interaction with an "intelligent learning environment". Also noteworthy is *iDict* [3], a translation aid designed for language courses which exploits the user's eye movements to understand whether and when the reader needs help while reading a document written in a foreign language.

The *e5Learning* Project

e5Learning (from *enhanced exploitation of eyes for effective eLearning*) is the project we are carrying out at the University of Pavia. Its aim is to practically implement and test an e-learning environment characterized by four main components:

- A *monitor of accessed screen areas*, which keeps track of the parts of the e-learning content that have or have not been looked at for sufficiently long times on the screen.
- A *contextual content generator*, that can be used to contextualize the presentation of additional information according to the main content being accessed by the user.
- A *history recorder*, which saves the "history" of eye-related user activities and uses it to suggest

proper learning paths to the user or better organize the learning material.

- An *emotion recognizer*, i.e. a module able to assess the user's "emotional state".

Monitor of accessed screen areas

We want the designer of the e-learning course to be able to decide "how much attention" the user must pay to certain portions of content. A subject matter is usually composed of different kinds of media, such as text, images, animations, etc. Each part of the content is useful for the comprehension of the topic being described, but some elements, more than others, may be essential for the theme to be correctly understood. In general, we want the course creator to be able to specify, for rectangular areas on the screen, "how much" and/or "how long" the user should look at them.

In our prototype, a course is simply made up of web pages. We use an ad-hoc-built web browser which, together with page content, reads additional information defined by the author. Among other things, such information specifies the coordinates of screen rectangles (corresponding to relevant portions of content) and associated data about fixations and fixation times. Using a special version of the browser, the course creator easily draws (with the mouse) rectangles around those areas for which he or she wants to impose eye constraints. For instance, if the laid out conditions are not satisfied, the user may not be allowed to proceed with the next page in the course. The examples in figures 1-3 refer to a tutorial explaining how an internal combustion engine works.[1]

1 The content of the examples presented in this paper is freely adapted from *www.howstuffworks.com*.

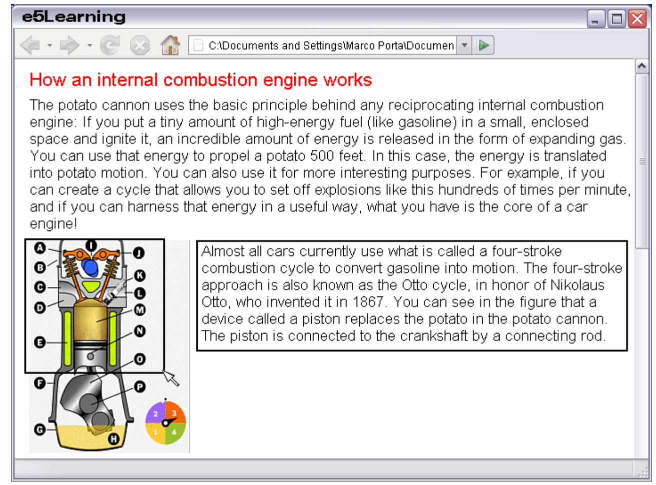

figure 1. The course author draws rectangles around those areas for which he or she wants to specify eye constraints.

We distinguish two kinds of content, namely text and non-text. For textual areas, the system tries to understand whether they have been actually read by the user. Unfortunately, due to the limited precision of the employed eye tracker — which is however not lower than that of most commercially-available devices — an accurate check of the reading process is only possible for big text. When line spacing is low, in fact, eye fixations on the different rows are hardly detectable, and we simply check that the progression of the eye scanpath within the content rectangle is roughly from left to right and from top to bottom, with right-to-left abrupt jumps. When line spacing is sufficiently high (about 1 cm), fixations on single rows can be exactly identified, and the check is much more precise. For non-textual areas, the course author specifies how many fixations there must be within the rectangular region, or their global duration.

Contextual content generator

The presentation of contextualized content depending on the main content being accessed at a certain moment by the user is another interesting feature that may improve the learning process. The creator of the course can associate new content to rectangular screen areas, and indicate the requirements for the additional information to be displayed (in the form of HTML pages appearing within a popup window; see figure 2).

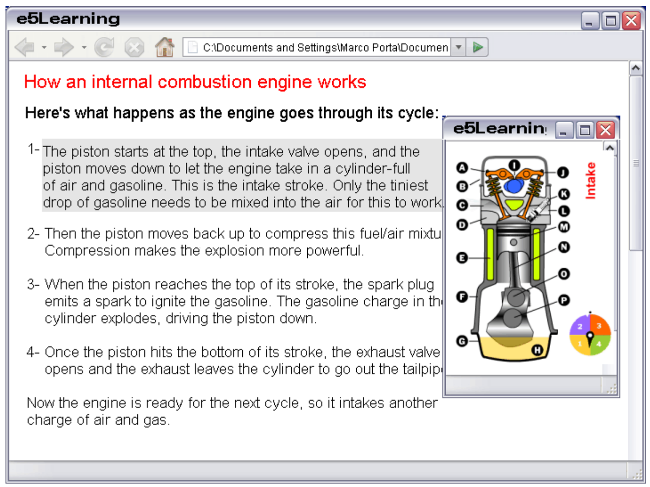

figure 2. When the user starts reading the first portion of text in the numbered list (here highlighted with a grey background), a popup window with related content appears.

Requirements are typically the minimum number and global length of eye fixations inside the areas, or, for text portions, the detection of a reading process. In a sense, we can say that the accessory content is shown when the user's focus of attention is found within the corresponding region. Also in this case, the course author creates "sensitive" rectangles by simply drawing them over the pages displayed within the special browser. Once a rectangle has been sketched, a dialog box appears that allows eye constraints to be properly specified, along with other information such as the URL of the page to be displayed and the size and position of its window.

History recorder

This module keeps track of which portions of content (i.e. rectangular screen regions) have already been accessed by the user, and "how much". Each area can be only in one of four possible states, namely *fully*, *partially*, *very little* and *not at all* accessed. For textual content, the (rough or precise) number of scanned lines is used to choose the best state, while for other elements more general criteria, based on the duration of fixations, are employed. The regions of interest are, as always, preliminarily defined by the author of the course, and may or may not coincide with areas associated with the other functionalities of the e5Learning project.

Information about the state of a portion of content can be exploited both implicitly and explicitly. For example, if in a previous session the user did not devote sufficient time to a certain area, such portion might be subsequently proposed before others, independently of its position in the logical structure of the course. Another strategy, which is the one we have already actually implemented, explicitly highlights, through colored rectangles, the state of each region, as shown in figure 3. This means that the user, if desires it, can be explicitly reminded about his or her earlier activities, which may greatly speed up and improve the learning process.

Color codes:
green=*fully* accessed,
yellow=*partially* accessed,
orange=*very little* accessed,
red=*not at all* accessed

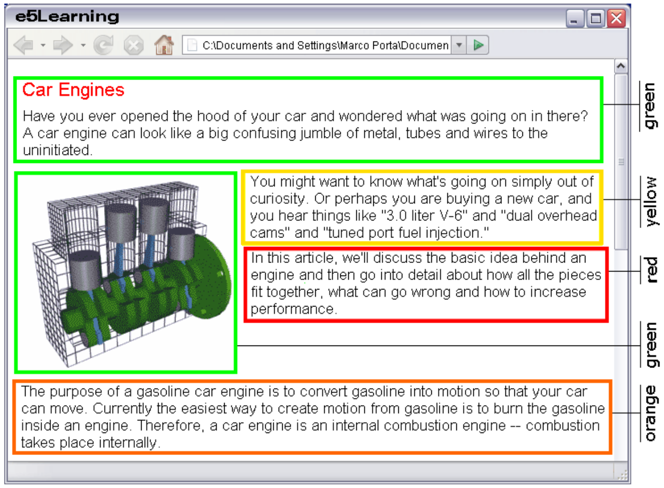

figure 3. Colored rectangles highlighting "how much" the content has been previously accessed.

Emotion recognizer
Several studies have been carried out — mostly in the Psychology and Physiology fields — to find correlations between eye behaviors and emotional states. For instance, experiments have demonstrated that pupil size is significantly larger after highly arousing stimuli than after neutral stimuli [8]. Also, pupil diameter seems to be task-dependent (being, for example, notably larger when "searching" than when simply "viewing" [6]). Other investigations (e.g. [5]) suggest that the mental workload can be assessed by analyzing the fluctuation rhythm of the pupil area. Indications about intellectual efforts can be also derived from blink rate and saccadic data. For instance, while pupil size and blink rate usually increase in response to task difficulty, both the saccade occurrence rate and saccade length typically decrease with the increased complexity of the task [7]. Moreover, saccadic and blink velocity seems to decrease with increasing tiredness [10].

In our project we are mainly concentrating on pupil size, blink rate and saccade length/occurrence rate. For example, if the average pupil size has progressively increased within the last 15 minutes (or other time interval), also user workload may have augmented. An increased blink rate and a decreased saccade rate in the same interval can further confirm such a supposition. When detected, such evidences could be used for example to dynamically modify the learning path, proposing a topic related to the main one but less complex (a sort of "break"). Or, if the user is potentially having problems in understanding something, extra information may be proposed. Since it is practically impossible to be sure that variations in eye parameters are due to changes in the user emotional state, however, we choose to undertake "helping actions" gradually — for instance, at first as links to additional material, and then, if the signs of potential non-understanding persist, as possible different presentations/explanations of the same topic.

Implementation issues
The system is developed in C#, within the .NET Microsoft framework. As an eye tracker, we use the Tobii 1750 (figure 4), which integrates all the components (camera, infrared lighting, etc.) into a 17'' monitor and is characterized by an accuracy of 0.5 degrees. At the moment of writing this document, the first three components of e5Learning (monitor of accessed screen areas, contextual content generator and history recorder) have been fully implemented, apart from refinements, while we are still working on the fourth module (emotion recognizer), which is also

figure 4. The Tobii 1750 eye tracker

the most "delicate" one. Of course, a very important step will be the test of the system, which we will plan very carefully. To date, the preliminary experiments we have carried out are very comforting.

Conclusions

We think that eye tracking technology is now an extraordinary opportunity for e-learning, making it possible to create new advanced computer-based teaching systems able to "understand" the user and properly adapt content presentation. Unfortunately, due to the very restricted market of eye trackers, prices are still very high (above 20,000 USD), but things may totally change when these devices will start to spread, hopefully in a not too far future. Maybe, in 15 years monitors incorporating eye tracking functionalities will be as common as LCD screens are now, and current researches in this field can certainly help such expectation to become a reality.

Acknowledgements

This work is supported by funds from the Italian FIRB project "Software and Communication Platforms for High-Performance Collaborative Grid" (grant RBIN043TKY).

Citations

[1] Cantoni, V., Cellario, M., and Porta, M. Perspectives and Challenges in E-Learning: Towards Natural Interaction Paradigms. *Journal of Visual Languages and Computing*, 15, 2004, 333-345.

[2] Hedge, N., and MacDonald, D. Affective Considerations in Distance e-Learning. In *Proc. of the IASTED Conference on Education and Technology*, 2005.

[3] Hyrskykari, A., Majaranta, P., Aaltonen, A., and Räihä, K. Design Issues of iDict: A Gaze-Assisted Translation Aid. In *Proc. of ETRA 2000 (Eye Tracking Research and Applications Symposium)*, 2000, 9-14.

[4] Merten C., and Conati, C. Eye-Tracking to Model and Adapt to User Meta-cognition in Intelligent Learning Environments. In *Proc. of IUI'06 (Conference on Intelligent User Interfaces)*, 2006, 39-46.

[5] Murata, A., and Iwase, H. Evaluation of Mental Workload by Fluctuation Analysis of Pupil Area. In *Proc. of the 20th Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Vol. 20, No. 6, 1998.

[6] Nakayama, M., and Shimizu, Y. Evaluation of a Multimedia Learning Exercise using Oculo-motors. In *Proc. of ETRA 2006 (Eye Tracking Research and Applications Symposium)*, 2006, 46.

[7] Nakayama, M., Takahashi, K, and Shimizu, Y. The Act of Task Difficulty and Eye-movement Frequency for the 'Oculo-motor' indices'. In *Proc. of ETRA 2002 (Eye Tracking Research and Applications Symposium)*, 2002, 37-42.

[8] Partala T., Jokiniemi, M., and Surakka, V. Pupillary Responses to Emotionally Provocative Stimuli. In Proc. of *ETRA 2000 (Eye Tracking Research and Applications Symposium)*, 2000.

[9] Picard, R. W., and Klein, J. Computers that Recognize and Respond to User Emotions: Theoretical and Practical Implications. *MIT Media Lab Tech Report 538*, 2002.

[10] Pivec, M., Trummer, C., and Pripfl, J. Eye-Tracking Adaptable e-Learning and Content Authoring Support, *Informatica*, 30 (2006), 83–86.

[11] Porta, M. E-learning and machine perception: in pursuit of human-like interaction in computer-based teaching systems. *Int. Journal of Knowledge and Learning*, Vol. 3., Nos. 2/3, 2007, 281-298.

[12] Wang, H., Chignell, M., and Ishizuka, M. Empathic Tutoring Software Agents Using Real-time Eye Tracking. In *Proc. of ETRA 2006 (Eye Tracking Research and Applications Symposium*, 2006, 73-78.