Gaze-based Notetaking for Learning from Lecture Videos

Cuong Nguyen
Portland State University
Portland, OR 97207-0751
cuong3@pdx.edu

Feng Liu
Portland State University
Portland, OR 97207-0751
fliu@cs.pdx.edu

ABSTRACT
Taking notes has been shown helpful for learning. This activity, however, is not well supported when learning from watching lecture videos. The conventional video interface does not allow users to quickly locate and annotate important content in the video as notes. Moreover, users sometimes need to manually pause the video while taking notes, which is often distracting. In this paper, we develop a gaze-based system to assist a user in notetaking while watching lecture videos. Our system has two features to support notetaking. First, our system integrates offline video analysis and online gaze analysis to automatically detect and highlight key content from the lecture video for notetaking. Second, our system provides adaptive video control that automatically reduces the video playback speed or pauses it while a user is taking notes to minimize the user's effort in controlling video. Our study shows that our system enables users to take notes more easily and with better quality than the traditional video interface.

Author Keywords
Lecture video; Notetaking; Eye tracking; Video interface

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces

INTRODUCTION
Video lectures are now popular for learning. Notetaking that has been shown helpful for learning [21], however, is challenging for students who learn from watching lecture videos. Taking notes while watching a video typically requires a lot of user effort [22]. To take notes, a user needs to identify important information like facts, numbers, and formulas while comprehending the constantly streaming information from video. Moreover, when writing notes, the user needs to deal with extraneous video control tasks such as pausing the video to avoid missing video content or rewinding to catch up with the video progress. These tedious video interactions can often discourage students from taking notes and keeping them from many educational benefits of notetaking.

This paper aims to develop a system to assist notetaking. We focus on slides-based lecture videos that show PowerPoint slides in the video with the voice from the instructor [11], as shown in Figure 1B. Education research shows that when watching a lecture video, a user's attention is often guided toward important points in the slides by the instructor's teaching (raise voice, write on slide) or by the content presentation (well-structured headings and lists) [21]. Thus, we can infer important points in the video by analyzing the user's gaze activities and highlight them as note candidates to users to ease the tasks of locating, annotating, and writing notes.

CHI 2016, May 07-12, 2016, San Jose, CA, USA

DOI: http://dx.doi.org/10.1145/2858036.2858137

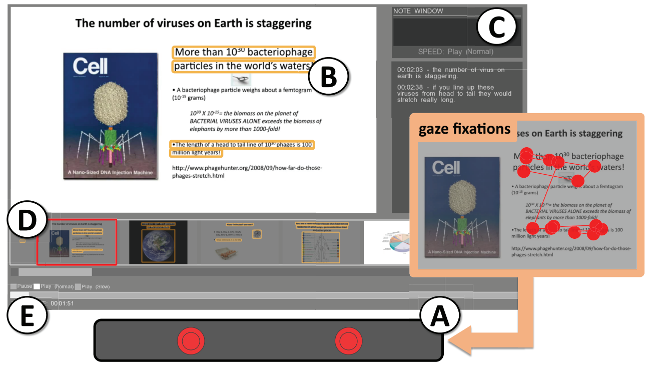

Figure 1. System overview. *GazeNoter* uses an eye tracker (A) to capture the user gaze activities to infer potentially useful lecture contents, visualizes and recommends them to users for notetaking (B). Users can also type texts as notes in the *Note Window* (C). To support manual note typing, *GazeNoter* can automatically control the video playback to free users from these tasks. *GazeNoter* also provides interfaces for users to browse the notes.

This paper presents *GazeNoter*, a gaze-based system to support notetaking. First, *GazeNoter* combines offline video analysis and online user gaze monitoring to identify potentially useful lecture content, visualize and highlight them to users for notetaking (Figure 1B). Users can quickly dismiss or accept these recommendations with little effort. Besides this advanced notetaking mode, *GazeNoter* also supports regular notetaking by allowing users to type in a *Note Window* (Figure 1C). When users type notes, they often need to pause or rewind the video to keep up with the video progress, which is often inconvenient. To free users from these tedious video control tasks, *GazeNoter* analyzes when and where the user focuses on the video during typing and adaptively reduces the video playback speed or pauses the video so that users can focus more on writing notes and less on controlling video. Finally, *GazeNoter* supports users to review notes through interacting with the video or the slide thumbnails (Figure 1D).

A good feature of *GazeNoter* is that it provides a hybrid format of notes that combines regular texts summarized by users and the content directly annotated from the lecture videos that are particularly useful for information like tables and pictures. Its tools can free users from tedious video control and save efforts in identifying and annotating important video content. Our preliminary user study shows that *GazeNoter* enables users to take high-quality notes easily.

RELATED WORK

Notetaking has been shown important for learning [21]; however, it is cognitively demanding [22], and particularly for video lectures, it incurs additional challenges such as finding key content and controlling the video playback [1, 19]. Accordingly, several technologies used head tracking to detect when a user is looking away from the screen in order to slow down video playback to support writing notes on paper [20, 30]. Our work extends this idea to the scenario of taking digital notes and combines both offline video analysis and online gaze monitoring to better control the video playback. Several systems have also been developed to support collaborative digital note taking and sharing [2, 17, 19, 26]. Our work focuses on assisting individual notetaking and can be well integrated into these systems.

Eye tracking is now widely used to enhance learning [7, 10, 16, 28]. Of particular relevance to our work is the use of eye tracking to locate and annotate content to support reading [5, 13] and e-learning [24]. Our work further extends this idea to support notetaking in lecture videos. By integrating video and gaze analysis, our system can infer and annotate important points on video to highlight them as note candidates, allowing users to quickly take notes.

GAZENOTER PROTOTYPE

As shown in Figure 1, *GazeNoter* is built upon a classic video player with a video window (B) and a timeline (E). It also has a slide thumbnail window (D). It is equipped with an eye tracker (A). It supports two notetaking modes: 1) it allows users to type and edit texts in a *Note Window* (C) and 2) it employs gaze analysis and video analysis to highlight video content to users as candidate notes. It also automatically adjusts video playback speed to minimize the user effort in video control when taking notes. *GazeNoter* stores notes and annotations in an XML database with video time stamps, and supports users to manually add, edit, and remove them in video. Finally, *GazeNoter* supports users to browse and review the text-based notes in the *Note Window* (C) and the annotation-based notes via the slide thumbnails (D).

Lecture Video Analysis

We first pre-process a lecture video and partition it into slide segments such that each segment only contains one slide. Briefly, we detect the slide transition by computing the differences between neighboring video frames using an algorithm from Gigonzac *et al.* [9]. We then follow the method from Denoue *et al.* [6] to detect areas of interest (AOIs) from each slide. Specifically, we first estimate an edge map from the video frame using the Canny edge detection algorithm and then find the bounding boxes of connected components as AOIs, as illustrated with yellow rectangles in Figure 2.

Candidate Notes Detection

When a user is watching a lecture video, our system employs gaze activity analysis to select the AOIs that attract significant fixations as candidate notes. Selected AOIs can serve as visual notes just like users highlighting texts in a book [18].

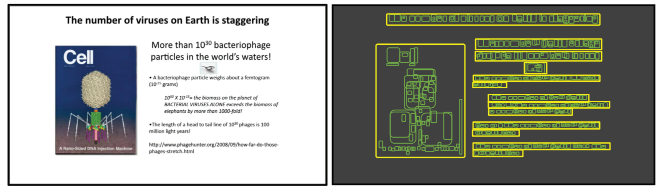

Figure 2. Lecture video analysis. Left: a lecture video slide; Right: Detected areas of interests (yellow) and their enclosing blocks (green) which represent finer details such as letters or numbers.

They can complement text-based notes, especially to note complex illustrations that are difficult to describe verbally.

It has been reported that important content often attracts more fixation visits and longer gaze duration than those less important ones [29]. Therefore, when a user is watching video, our system records her gaze activities, namely re-visit count and gaze duration on each AOI. Gaze activities are accumulated when a user fixates on contents inside an AOI. We use *Tobii EyeX* eye tracker and implement Salvucci's i-DT algorithm to detect fixations [27]. A fixation is defined as a collection of gaze points within a 35-pixel diameter and lasts for a minimum of 100 ms. To filter non-meaningful fixations that hover on empty regions within an AOI, we further decompose the AOI into finer blocks such as letters or numbers, as indicated by the green rectangles in Figure 2, and record fixations within 35 pixels of any of these blocks. To remove spurious glances, gaze re-visit count and duration are updated only if the user has fixated on an AOI for more than a second.

An AOI is selected as a candidate note if its re-visit count reaches three (3) or the gaze duration is beyond a pre-defined threshold. The threshold values are defined empirically to make the system conservative in selecting candidate AOIs while still capturing most of the important notes. Determining a fixed time threshold for the gaze duration can be tricky because the gaze duration can vary depending on many factors such as the user's perceptual capability or the image content [25]. To accommodate each individual user's need and preference [13], our system sets the default time-threshold to be two (2) seconds, and allows each user to adjust this threshold to modify the proactivity level of the system, which allows novice users to select an appropriate threshold. Specifically, a user can select three modes: high, normal, and low, which correspond to coefficients 0.5, 1, and 2, respectively. These coefficients are multiplied to the default time threshold to calculate the final threshold that determines how easily an AOI can be selected as a candidate note.

Once an AOI is identified as a candidate note, our system highlights it using an orange-colored rectangular box such as those shown in Figure 1B. The box slowly fades in to minimize the visual distraction. All boxes are kept by default to capture most of the important notes, but users can easily dismiss irrelevant ones by pressing the *Esc* key on the keyboard, which in turn will erase the last annotated AOI and reset its corresponding gaze record. Finally, a user can always manually annotate the video to create the AOI based visual notes by marking them if either the video analysis or gaze analysis fails to highlight an AOI that a user considers important.

Adaptive Video Playback Control

GazeNoter also supports users to type notes in the *Note Window*. Writing notes, while being useful for learning [21], is cognitively demanding as users have to simultaneously gather information from the streaming video to recall or summarize thoughts and type them. Previous studies found that users tend to frequently pause the video before they take notes [3], which is distracting. For a typical lecture video that lasts more than 45 minutes, the tedious video playback control can discourage users from writing notes. To make matters worse, if a user forgets to pause the video, the streaming video can either overwhelm the user's limited working memory or make her miss some important content. When the latter happens, the user needs to rewind the video to catch up.

To help a user write notes more effectively, our system employs gaze analysis and video analysis to adaptively control the video playback when the user is typing to help her focus more on notetaking and less on extraneous video control. In our system, a user enters the *writing mode* when she starts typing and she stays in this mode until she submits the note by pressing Enter. During the *writing mode*, if the user fixates on the video content for more than one second, our system considers her in the *information gathering mode* and the *writing mode* simultaneously and automatically reduces the video playback speed to 0.85x to give the user more time to process the coming information, as shown in Figure 3. 0.85x is determined in part empirically and also based on the recommendation from previous work [20, 30] although they used 0.80x, slightly slower than our choice. We used 0.85x to avoid distorting the audio. Moreover, when the video is about to switch to a different slide and the user has not finished taking the note, our system pauses to the video. Similarly to the Pause-and-Play system [23], our system resumes the normal video playback after the user submits the note.

In addition, while a user is typing a note, she often needs to frequently switch her focus between the *Note Window* and the video content. This frequent attention switch often poses a cognition burden during notetaking. Thus, beside the candidate AOI highlights for notetaking, *GazeNoter* supports another type of AOI highlights to ease the attention switch. As inspired by the "visual placeholders" [14], our system highlights the latest AOIs in the video that the user attends to during the *writing mode* by enclosing them with blue rectangles (Figure 3). Note, we use the blue color here to differentiate from the default orange-colored AOI highlights for notetaking. These blue rectangles are visible only during *writing mode*. Looking at them can connect the user's writing to the lecture content that she wants to refer to.

Notes Browsing and Reviewing

GazeNoter lets users review notes in the context of the video by providing interfaces for browsing the text-based notes that are typed by the user and the highlighted AOI notes. Inspired by existing notetaking tools [2, 12, 19], the user can review and edit the text notes in the *Note Window* or click on a note to seek to its time stamp (Figure 1C). For the AOI notes, we first enhance the slide thumbnails with AOIs notes by overlaying the AOI windows on top. The user can also click the thumbnail to seek to the corresponding video segment.

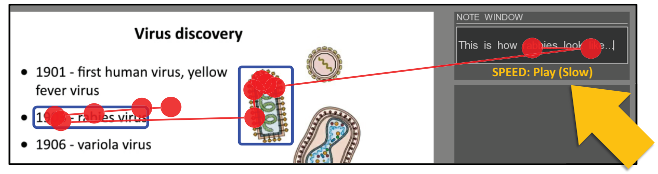

Figure 3. Adaptive playback control. Our system slows down the video playback when a user is typing notes and looking at the video simultaneously. If the video is about to change to another slide, the video is paused. Moreover, the video content which is relevant to the notetaking is highlighted with blue rectangles to ease the attention switch between the video and the *Note Window*. Note, the red dots are fixation points and are shown for illustration. They are not shown in the real system.

USER STUDY

We conducted a preliminary user study in our lab to assess how *GazeNoter* supports participants to take notes while watching a lecture video. We compared to a baseline notetaking system which was built by modifying *GazeNoter* to remove the gaze-assisted features (AOI note candidate recommendation and adaptive video control).

We chose a between-subject study design to handle the strong order effect of taking notes. Our study includes a notetaking stage where participants were instructed to watch a history video and take notes on key ideas. The video lasts 23 minutes, contains 34 slides, and is narrated by a teacher telling the chronological events of the Thirty Years' War. Similarly to previous studies [15, 26], we included an additional test stage after the notetaking stage to motivate the participants to take note but did not measure learning gains in this study. We recruited 16 participants from our university campus who had experience at taking notes during live lecture and are not familiar with the subject matter in the video. The participants' ages range from 18 to 32 years old (M = 24.8). All participants are students from a variety of departments except one software engineer who lives on campus.

Before the study, each participant was guided through a five-minute calibration step. This is followed by a practice session where participants were trained to use the notetaking features of their system. Participants in the *GazeNoter* group were also allowed to specify the desired proactivity level of the gaze analysis component. They were then asked to read a short overview of the lecture and introduced to the task. Participants were also informed that they would be tested afterward with ten multiple-choice questions on key ideas, and that both their notes and video will be available in the test. During the task, they were given as much time as needed to watch the lecture video and take notes. Upon completion, they proceeded to the test, which lasts for ten minutes.

To evaluate the participants' notetaking performance, we counted the number of notes taken by each participant. Like previous research [4, 15], we also analyzed the quality of notes by measuring the note length (in characters) and counting notes that were taken on key ideas. Key ideas are dates and events presented in the video that were hand-coded by one of the authors and another graduate student in our lab, and then compared and normalized for consistency. We also logged video interactions such as play, pause, and seek during notetaking and recorded the task completion time (in sec-

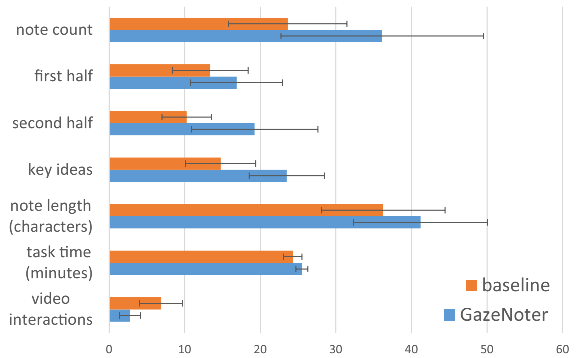

Figure 4. Study results.

onds) of each participant. Finally, the subjective feedback and comments were collected at the end of the study.

Results

The results from our study is reported in Figure 4. We first wanted to find out if the systems had an overall effect on our measurements. We conducted a one-way multivariate analysis of variance (MANOVA) on four measurements (key ideas, note length, task time, video interactions). Note, we excluded "note count" from MANOVA since it was highly correlated with "key ideas"; the Pearson correlation coefficient was 0.744. We found that notetaking systems could influence the users' notetaking behavior, as indicated by a statistical significant difference between *GazeNoter* and baseline on these measurements ($F(4, 11) = 3.85, p = 0.034$). To follow up, independent samples t-tests were performed on all measurements, and effect size is reported as Cohen's d. No adjustment has been made for multiple testing.

Overall, we found that participants in the *GazeNoter* group took more notes with longer length and more key ideas. Statistically, the difference in note count ($p = 0.13, d = 0.79$) and note length ($p = 0.43, d = 0.4$) was not significant, but was significant for key ideas ($t(14) = 2.51, p = 0.025, d = 1.26$). To further investigate participants' performance, we divided the video into two segments to see how notes were distributed overtime. We found that participants in the baseline group took less notes both in the first and second half of the video than those in the *GazeNoter* group. The numerical difference is especially large in the second half although $p = 0.07 (d = 0.98)$. Interestingly, the declining of notes overtime in the baseline group has also been reported in previous research on notetaking in video lecture [8, 19]. One possible explanation is that participants in the baseline group might have been overwhelmed by the notetaking task toward the end of the video [22]. In contrast, gaze-based annotations in *GazeNoter* can help users quickly locate important moments in the video to take notes. Subjective feedback shows that six out of eight participants found the gaze-based annotations helpful, adding that they can use them as a reminder to write notes. Furthermore, looking at the recommended AOIs in *GazeNoter*, we found that each participant accepted 28 AOIs on average while rejection happened only once with only one participant. The low rejection rate could be possibly caused by the unwillingness of the users to reject bad highlighted AOIs; or the gaze data was very helpful in highlighting the good ones. To better understand this, we further looked into the study and found that our system was tuned to be very conservative in highlighting AOIs in order to minimize the user effort in rejecting bad ones to avoid distracting users.

Subjective feedback also shows that all participants in the *GazeNoter* group found the automatic slow and pause features helpful for them to focus and think during writing notes. As a result, they spent more time on notetaking and produced longer notes, according to note length ($p = 0.43, d = 0.4$) and task time ($p = 0.13, d = 0.8$) in Figure 4. Looking at video interactions logged during the study, we found that without the adaptive video playback control, participants in the baseline group had to spend more effort on controlling the video. With a Welch's t-test, we found that the difference in the numbers of video interactions was significant ($t(10.07) = 2.55, p = 0.029, d = 1.27$). All participants in the baseline group felt the video was distracting during notetaking. Thus, by automatically adjusting video playback during notetaking, our system could offload extraneous video control tasks and help users focus more on taking notes.

Finally, we looked into the test scores although they were used to motivate notetaking rather than evaluate learning outcomes. We found that the *GazeNoter* group only had an insignificantly better average score than the baseline group.

Discussion & Limitations

Our automatic slow and pause features are activated when users write notes and focus on video simultaneously. While our method is useful for information-dense video lecture, it will not work well for users who focus only on notetaking and ignore the video. Moreover, we currently do not differentiate whether the users are interested in some video content or are confused. Building a more sophisticated cognition model by integrating gaze analysis and face expression analysis will be helpful to enhance our system. For more complex video such as those with texts overlapping images, our current preprocessing algorithm may not work well in detecting AOIs. Future iterations can incorporate more advanced computer vision algorithms, such as using OCR to detect text elements.

Finally, the sample size in our study is small, which could limit the generalizability of our findings. Nevertheless, our study at least suggests promising performance of our system.

CONCLUSION

This paper presented a gaze-based system to support effective notetaking when learning from lecture videos. Our system integrates offline video analysis and online gaze analysis to provide automatic support for notetaking tasks such as detecting and highlighting note candidates to users and minimizing user effort in controlling video playback during notetaking. Our system also enables a hybrid format of notes that combines both text-based notes that the user typed and annotations that come directly from the video. Our preliminary results showed that our system enables users to take high-quality notes easily with minimal effort in controlling the video.

Acknowledgments. Figure 1, 2 and 3 use images from YouTube user Vincent Racaniello under a Creative Commons license. This work was supported by NSF IIS-1321119 and CNS-1218589.

REFERENCES

1. Aubert, O., Prié, Y., and Canellas, C. Leveraging video annotations in video-based e-learning. In *International Conference on Computer Supported Education* (2014).
2. Bargeron, D., Grudin, J., Gupta, A., Sanocki, E., Li, F., and Leetiernan, S. Asynchronous Collaboration Around Multimedia Applied to On-Demand Education. *Journal of Management Information Systems* (2002), 117–145.
3. Bassili, J., and Joordens, S. Media Player Tool Use, Satisfaction with Online Lectures and Examination Performance. *Journal of Distance Education* (2008), 93–108.
4. Bauer, A., and Koedinger, K. R. Selection-based note-taking applications. In *ACM CHI* (2007), 981.
5. Cheng, S., Sun, Z., Sun, L., Yee, K., and Dey, A. K. Gaze-Based Annotations for Reading Comprehension. In *ACM CHI* (2015), 1569–1572.
6. Denoue, L., Carter, S., Cooper, M., and Adcock, J. Real-time direct manipulation of screen-based videos. In *International conference on Intelligent user interfaces companion*, ACM Press (New York, New York, USA, 2013), 43.
7. D'Mello, S., Olney, A., Williams, C., and Hays, P. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies* (2012), 377–398.
8. Gašević, D., Mirriahi, N., and Dawson, S. Analytics of the effects of video use and instruction to support reflective learning. In *International Conference on Learning Analytics And Knowledge* (2014), 123–132.
9. Gigonzac, G., Pitie, F., and Kokaram, A. Electronic slide matching and enhancement of a lecture video. In *European Conference on Visual Media Production* (2007), 9–9.
10. Gog, T., and Scheiter, K. Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction* (2010), 95–99.
11. Guo, P., Kim, J., and Rubin, R. How video production affects student engagement: An empirical study of mooc videos. In *Learning @ scale* (2014), 41–50.
12. Hosack, B. VideoANT: Extending Online Video Annotation beyond Content Delivery. *TechTrends* (2010), 45–49.
13. Hyrskykari, A., Majaranta, P., and K.J., R. Proactive response to eye movements. In *INTERACT* (2003), 129.
14. Kern, D., Marshall, P., and Schmidt, A. Gazemarks: gaze-based visual placeholders to ease attention switching. In *ACM CHI* (2010), 2093–2102.
15. Kiewra, K., Benton, S., Kim, S.-I., Risch, N., and Christensen, M. Effects of Note-Taking Format and Study Technique on Recall and Relational Performance. *Contemporary Educational Psychology* (1995), 172–187.
16. Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., and Tsai, C.-C. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review* (2013), 90–115.
17. Lee, Y.-c., Lin, W.-c., Cherng, F.-y., Wang, H.-c., Sung, C.-y., and King, J.-t. Using Time-Anchored Peer Comments to Enhance Social Interaction in Online Educational Videos. In *ACM CHI* (2015), 689–698.
18. Marshall, C. C. Annotation: from paper books to the digital library. In *ACM ICDL* (1997), 131–140.
19. Mu, X. Towards effective video annotation: An approach to automatically link notes with video content. *Computers and Education* (2010), 1752–1763.
20. Park, M., Kang, J., Park, S., and Cho, K. A Natural User Interface for E-learning Learners : Focused on the Automatic Speed Control of Multimedia Materials. *International Journal of Multimedia & Ubiquitous Engineering* (2014), 347–358.
21. Piolat, A., and Boch, F. Note Taking and Learning : A Summary of Research. *The WAC Journal* (2005), 101–113.
22. Piolat, A., Olive, T., and Kellogg, R. T. Cognitive effort during note taking. *Applied Cognitive Psychology* (2005), 291–312.
23. Pongnumkul, S., Dontcheva, M., Li, W., and Wang, J. Pause-and-play: automatically linking screencast video tutorials with applications. In *ACM UIST* (2011).
24. Porta, M. Implementing eye-based user-aware e-learning. In *Extended Abstracts on Human Factors in Computing Systems* (2008), 3087–3092.
25. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology* (2009), 1457–1506.
26. Risko, E. F., Foulsham, T., Dawson, S., and Kingstone, A. The Collaborative Lecture Annotation System (CLAS): A New TOOL for Distributed Learning. *IEEE Transactions on Learning Technologies* (2013), 4–13.
27. Salvucci, D., and Goldberg, J. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *ACM ETRA* (2000), 71–78.
28. Sharma, K., Jermann, P., and Dillenbourg, P. How Students Learn using MOOCs: An Eye-tracking Insight. In *MOOC European Stakeholders Summit* (2014), 147–154.
29. Slykhuis, D., Wiebe, E., and Annetta, L. Eye-tracking students' attention to powerpoint photographs in a science education setting. *Journal of Science Education and Technology 14*, 5-6 (2005), 509–520.
30. Song, S., Hong, J.-k., Oakley, I., Cho, J.-d., and Bianchi, A. Automatically Adjusting the Speed of E-Learning Videos. In *Extended Abstracts on Human Factors in Computing Systems* (2015), 1451–1456.