Automatically Adjusting the Speed of E-Learning Videos

Sunghyun Song
Dept. of Human ICT
Sungkyunkwan University.
Suwon, Republic of Korea
boss.ssong@gmail.com

Jeong-ki Hong
Dept. of Human ICT
Sungkyunkwan University.
Suwon, Republic of Korea
burnness1@gmail.com

Ian Oakley
School of Design and Human Engineering. UNIST.
Ulsan, Republic of Korea
ianoakley@unist.ac.kr

Jun-dong Cho
Dept. of Human ICT
Sungkyunkwan University.
Suwon, Republic of Korea
jdcho@skku.edu

Andrea Bianchi
Dept. of Computer Science
Sungkyunkwan University.
Suwon, Republic of Korea
andrea.whites@gmail.com

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).
CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.
http://dx.doi.org/10.1145/2702613.2732711

Abstract
Videos are becoming a commonplace way for students to view instructional material. Although current technology allows customization of default playback speeds to cater to individual students' desired pace, we highlight a need for more dynamic or reactive control systems capable of varying playback in response to viewer needs or activities (e.g. slowing down during note-taking). This article instantiates this idea by describing a system that tracks a user's head position in order to infer and respond to their activities whilst watching an educational video. We describe the design and implementation of the system and a user study that highlights usage patterns and shows automatic tracking and playback speed adjustment can effectively lower users' workload.

Author Keywords
Video; e-learning; playback speed; head controller;

ACM Classification Keywords
H.5.2. [User Interfaces]: Input devices and strategies.

Introduction
e-Learning, facilitated by the widespread availability and distribution of rich audio-visual media, is a rapidly growing educational sector. Prominent universities,

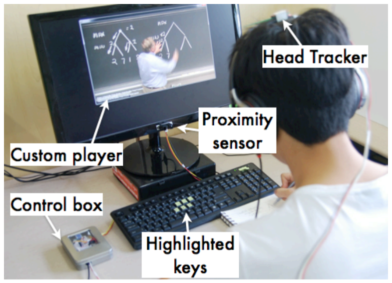

Figure 1. Overview of the video control head-tracking.

private academies and public networks are all rushing to provide online video courses across a broad range of topics in a bid to capture the enticing global market in the area - one that is predicted to grow by as much as 8% a year through 2016 [12]. These approaches also provide many benefits to students, such as convenient access and the ability to consume material at a self-defined pace. Indeed, one of the main advantages of audio-video class material is that students can flexibly play, pause, rewind or skip around material in response to their learning needs [2].

Reflecting these observations, researchers have suggested that e-Learning students would benefit from systems capable of tracking their attention and activity and adjusting video playback properties automatically [9, 12] by, for example, reducing playback speed if a student is taking notes [12]. The importance of this kind of dynamic control of content speed is also reflected in commercially available players, such as the Apple iTunes University [1], and research projects [3] that, respectively, enable users to adjust playback speed manually or depending on specific qualities of the displayed content. However, to the best of our knowledge, there is no current video-playback system that tracks user activities to dynamically adjust video playback in order to support e-Learning scenarios.

Accordingly, this article presents an implementation of such a system. We developed a novel natural user interface that tracks a user's head movements to infer, without explicit user input, whether a video should be paused, played at a normal speed, sped up or slowed down. Using this system we then conducted a user study to understand how it affects student behavior during watching a video lecture. The contributions of this article are 1) the design and implementation of a head-tracking system for adjusting the speed of video playback, and 2) a usability study to understand how this system affects the way students watch e-Learning videos and take notes.

Related Work

The idea of tracking head position to control aspects of computer use is well explored. Harrison and Dey [7], for example, present Lean and Zoom, a system that based on a camera facing the user calculates the distance between head and screen in order to adjust the magnification of displayed contents for optimal viewing. Yamaguchi et al. [15] extend this idea in the context of a video-chat system and use head movement to adjust not only the zoom but also the location and orientation of on screen content. Head-trackers have also been used to replace or extend the capabilities of pointer input [10]. Finally accelerometer sensed head pose [5] and eye-tracking technologies [14] have also been used to enable new interaction possibilities.

A broad range of video browsing techniques have been proposed and studied, including a number that focus on control of playback speed. Dragicevic *et al.* [6], for example, created a system that lets users navigate the timeline of a video by directly dragging objects in the displayed scene along their past or future visual path. Video control through gestures has been explored by Chu et al. [4] in an interface powered by the popular Microsoft Kinect and Ryokai et al. [13] describe a playful physical interface that lets children control video content through squeezing and stretching bubble-like objects. Kim et al. [9] introduce a set of techniques that augment existing video graphical user interfaces,

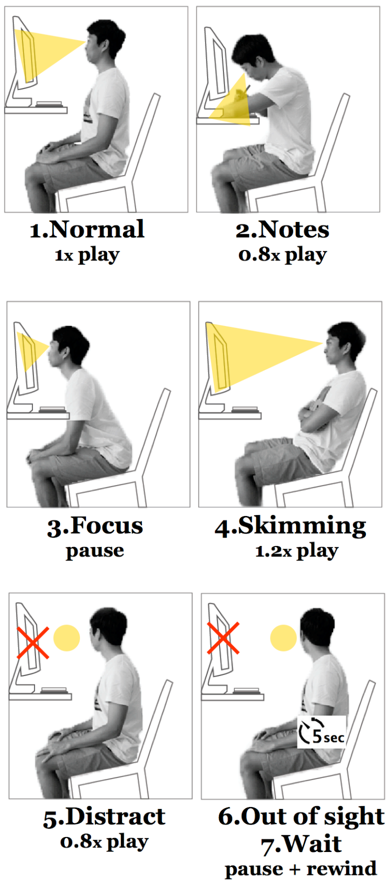

Figure 2. The seven different postures that the head tracker can distinguish. States are sensed by combined angular data from a head-mounted six-axis IMU and distance data from an infrared proximity sensor placed at the bottom of the screen and facing the user.

including dynamic and non-linear timelines. Finally, the SmartPlayer [3] semi-automatically adjusts the video playback speed depending on the complexity of the scene presented or user-defined events.

Hardware and Software Prototype

The prototype (Figure 1) is composed of a video-playback PC and screen, a forward-facing infrared proximity sensor (Sharp GP2Y0A02YK0F) attached to the base of the screen and a six-axis Inertial Motion Unit (an MPU-6050 IMU) mounted on top of a pair of headphones. All electronics are physically wired to a control box (115 x 85 x 25 mm) that contains an Arduino Leonardo micro-controller (connected to the playback PC through USB) and a switch that is used to toggle a calibration mode. In its head-mounted configuration, the IMU sensor can accurately report even very rapid head rotations (+-250° per second) and maintains an accurate representation of current angular position (e.g. pitch, roll and yaw) while the proximity sensor returns data describing the user's distance from the screen. All sensor data are sampled, filtered, and poses are calculated every 6ms. Excluding headphones and PC, the prototype cost about 40 USD.

The system is operated as follows. A user wears the headphones and calibrates the system by facing the monitor in a comfortable viewing position and pressing the control box button. Home states for the sensors are then recorded. As the user watches the video, data from the IMU and proximity sensor are compared to the home states to determine if the user is looking at the monitor (pitch and yaw angles ±15° of home state), looking sideways (yaw between 15° and 60° of home state), or even further away (yaw >= 60°). The system also records if the user is looking downwards (pitch >= 15°) and, using the proximity sensor, whether the user's has moved nearer (-10 cm) or further (+10 cm) from the screen.

Seven states (Figure 2) were derived from this data, named for the actions users would typically be performing and linked to appropriate changes in video playback behavior. States and relative playback speed were designed closely following results in prior work. Specifically, playback speeds were informed by the work of Park et al. [12] and Kurihara [11]. States are: *normal*, if the user is simply looking at the screen (video playback at regular speed); *notes,* if the user looks forward and down (e.g. note-taking, playback reduced to 0.8 speed); *focused,* if the user moves closer to the screen (pause); *skimming,* if the user moves further from the screen (e.g. reclines, fast forwarding the video at 1.2 speed); *distracted,* if the user looks sideways (playback reduced to 0.8 speed); and *out-of-sight,* if the user is looking entirely away from the screen (pause). If the state remains *out-of-sight* for 5 seconds or more, the video automatically rewinds 15 seconds (*wait* state). All states were calculated based on sensor data in a 0.5 - 2 seconds window (depending on the current state) to reduce noise and transmitted to a PC that displays the video content accordingly.

Evaluation

Twelve participants (seven male and five female) completed the study. All were students engaged in undergraduate or postgraduate studies at Sungkyunkwan University and they had a mean age of 24 (SD 2.8) years. They were compensated with a gift card worth 10 USD. The study took place in a quiet room with participants seated comfortably in front of a

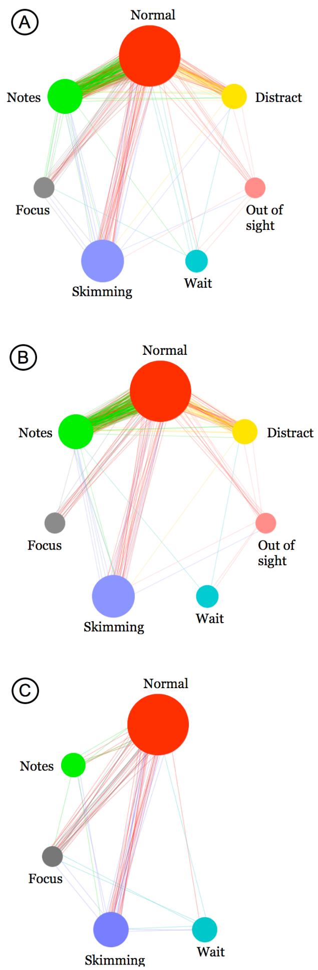

Figure 3. Graphs representing the usage pattern across different state-postures: (A) Tracker & keyboard input; (B) Tracker with keyboard input removed; (C) control condition (keyboard only).

desk, 24-inch monitor and keyboard. To ensure there was sufficient space for note-taking, the keyboard was positioned 30 cm (and the monitor approximately 50 cm) from the edge of the desk. Videos were presented on the monitor and participants used headphones to listen to the audio.

All participants completed both a tracked and a control condition in a fully balanced repeated measures design. In the control condition all input was through the keyboard (the keys mapped to video-playback actions were highlighted as in Figure 1), while in the tracked condition participants could use both the keyboard and the tracking system proposed in this paper. Two videos were used. Each video was a 8 minute long excerpt of a lecture about high-school geology ("cold and warm fronts" and "the solar system") broadcast online by a local educational television network and with all speech content in the participants native language. The topics were chosen to be relatively unfamiliar to the users (e.g. away from their majors), but at the same time not too challenging considering the diversity of their backgrounds. The presentation of the videos was also fully balanced: within the two main condition orders, half of the participants watched the videos in one order and the other half in the reverse order.

The study procedure was as follows: on arrival at the experimental room, participants were briefed about the study structure and purpose and given instructions on how to use both the tracked and control interfaces. They were then allowed to freely operate the systems for a maximum of two minutes to familiarize with the gestures/postures in the tracked condition. Next, the first condition was presented: the participant watched a video using the relevant interface. During this process, we logged all state changes to the video playback and the time they occurred. At the end of the condition, participants answered a short exam on the video contents and completed a NASA TLX workload questionnaire [8]. After a short break, they completed the second condition in the same way. The study closed with the participants responding to a request for subjective comments and opinions. All data captured in the experiment was automatically logged by software running on the PC.

Results and Discussion

In each condition we considered performance in the first minute of each video as setup/practice and excluded it from all the analysis of results. Examining the remaining data, we found the time it took participants to watch the videos did not vary between conditions. The tracker trials were completed in a mean of 408 seconds (SD 154) and the control in 390 seconds (SD 120) and a matched pairs one-tailed t-test showed this difference did not attain significance ($p=0.36$). However, behavior between the two conditions varied considerably. Specifically, in the tracked condition, participants made a mean of 42.2 (SD 12.9) state changes whilst watching the videos, compared to 7.4 (SD 6.3) in the control condition, a difference that was highly significant (t-test: $p<0.001$). On the flip side, the mean time in each state in the tracked condition was significantly lower (t-test: $p=0.016$) than in the control condition - 10.6 seconds (SD 5.24) versus 132.8 (SD 176.1). These preliminary results show that participants made more frequent state changes, and stayed in individual states for shorter periods of time, in the tracked condition.

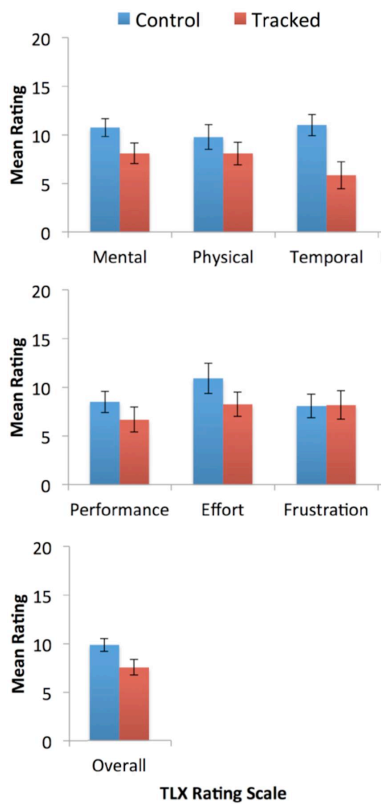

Figure 4. NASA TLX Results: lower cognitive load in the tracked condition.

To analyze this data in more detail, we plotted figures visualizing all state transitions in both conditions and for all participants (Figure 3). These figures show not only the variations in transition frequency but also illustrate which particular movements were more common with the tracker. Lines show state transitions triggered by the users and the size of the circles depicts the proportion of total time spent in each state. Beyond the increased frequency of transitions in the tracked condition (A and B) these figures illustrate that users spent most time in the *normal* state, while most common transitions took place between *normal* and *note*, *normal* and *focus* and *normal* and *skimming*. Only rarely did users move between any two other states without first returning to the *normal* state. Another interesting pattern is that in the tracked condition (A and B) the number of transitions between *normal* and *note* states is far greater than in the control condition. In the tracked condition we also found a high number of *normal* to *distracted* transitions, but experimenter observations corroborated by comments in the *post-hoc* interview suggested these were due to a misclassification: many users tilted their head sufficiently when writing notes to accidentally trigger the *distract* state. Fortunately, this was unnoticeable as playback reductions in both states were equal (0.8x).

Two independent raters were used as graders to examine the test results captured in the study. A substantial degree of reliability was found between the exam scores they produced - an Interclass Correlation Coefficient (ICC) of 0.793 (95% confidence interval between 0.522 and 0.912). This provides confidence in the validity of their grading. The overall mean grades were 86.1% using the tracker and 85.9% in the control condition. A one-tailed paired t-test showed no significant differences between these scores ($p=0.48$). The subjective measures showed greater ability to differentiate between the conditions. In the TLX data (Figure 4), an initial t-test revealed participants recorded significantly lower Overall Workload ($p=0.016$) in the tracked condition. Exploratory follow-up tests suggest this improvement was due to reductions in Mental Demand ($p=0.011$), Temporal Demand ($p<0.002$) and Effort ($p<0.04$).

Taken together these results indicate that while the tracker system did not reduce video watching time or knowledge gained (as measured by the exam scores), the ease and hands-free nature of state modifications encouraged users to take advantage of them, an extra level of control that likely contributed to the lower levels of workload users reported.

Conclusions and Future Work

Online courses supported by video delivery of class materials are now widely deployed. But mechanisms to control playback of this detailed, complex material remain broadly the same as for much simpler standard media content: play/pause buttons and limited speed control options [1]. Accordingly, this article describes a novel system that builds on prior work on head tracking during computer use [7, 15] to give students the ability to automatically control the delivery pace of video material based on their head position. The tool frees users hands for activities such as note taking, manages playback location in cases of interruption (via an automatic rewind feature) and modulates speed according to observed activities. A user study of twelve graduate students using this system was both informative and supportive of the system's usefulness. Specifically, the study highlighted that the tracker system encouraged and supported already prominent usage patterns, such as a commonly repeated shift between note-taking and viewing activities. Furthermore, subjective data indicated that doing so lowered perceived workload levels, a reduction in cognitive effort that could be highly beneficial in learning scenarios. Future work will establish the validity of these claims by improving the tracking with additional functionalities and mapping systems (e.g., rewinding and skipping forward) and conducting a longitudinal field study to investigate the impact of automatically controlled video playback on real world e-Learners. We finally plan to test the system with longer videos in order to determine whether unintentional triggers might occur and cause frustration.

Acknowledgments

This research was supported by the Ministry of Trade, Industry and Energy (MOTIE), Korea, through the Education Support program for Creative and Industrial Convergence (Grant Number N0000717). This paper was also supported by Basic Science Research Program through NRF of Korea, funded by MOE (NRF-2010-0020210) for Andrea Bianchi.

References

[1] Apple iTune U, Website: https://www.apple.com/education/ipad/itunes-u (last access September 2014).

[2] Bassili, J. N. and Joordens, S.: Media player tool use, satisfaction with online lectures and examination performance, in J. Distance Education, 22(2), 2008.

[3] Cheng, K., Luo, S., Chen, B., Chu, H.: SmartPlayer: user-centric video fast-forwarding, in Proc of CHI '09, 789-798.

[4] Chu, T., Su, C.: A Kinect-based Handwritten Digit Recognition for TV Remote Controller, in ISPACS'12, 414-419.

[5] Crossan, A., McGill, M., Brewster, S., Murray-Smith, R.: Head tilting for interaction in mobile contexts, in MobileHCI '09, article 6.

[6] Dragicevic, P., Ramos, G., Bibliowitcz, J., Nowrouzezahrai, D., Balakrishnan, R., Singh, K.: Video browsing by direct manipulation, in Proc. of CHI '08.

[7] Harrison, C., and Dey, A.K.: Lean and zoom: proximity-aware user interface and content magnification, in Proc. of CHI '08, 507-510.

[8] Hart, S.G., Staveland, L.E.: Development of a multi-dimensional workload rating scale. In: Human Mental Workload. Elsevier, 1988, 139–183.

[9] Kim, J., Guo, P.J., Cai, C.J., Li,S., Gajos,K.Z., Miller, R.C.: Data-driven interaction techniques for improving navigation of educational videos, in Proc. UIST '14, 563-572.

[10] Kjeldsen, R.: Head Gestures for Computer Control, in IEEE ICCV Workshop on RATFG-RTS'01, 61-67.

[11] Kurihara, K.: CinemaGazer: a system for watching videos at very high speed, in AVI '12, 108-115.

[12] Park, M., Kang, J., Park, S., Cho, K.: A Natural User Interface for E-learning Learners. in I.J. Multimedia and Ubiquitous Engineering 9(7), 2014

[13] Ryokai, K., Raffle, H., Horii, H., Mann, Y.: Tangible video bubbles, in CHI EA'10, 2775-2784.

[14] Smith, J.D., Graham, T.C.: Use of eye movements for video game control, in ACE 2006.

[15] Yamaguchi, K., Komuro, T., Ishikawa, M.: Ptz control with head tracking for video chat, in CHI EA '09, 3919-3924.