Teach Me Fast: How to Optimize Online Lecture Video Speeding for Learning in Less Time?

Devangini Patel
NUS Graduate School for Integrative Sciences and Engineering
National University of Singapore
Singapore
devangini@u.nus.edu

Shengdong Zhao
NUS Graduate School for Integrative Sciences and Engineering
National University of Singapore
Singapore
zhaosd@comp.nus.edu.sg

Debjyoti Ghosh
NUS Graduate School for Integrative Sciences and Engineering
National University of Singapore
Singapore
debjyoti@u.nus.edu

ChineseCHI '18, April 21–22, 2018, Montreal, QC, Canada
ACM 978-1-4503-6508-6/18/04.
https://doi.org/10.1145/3202667.3202696

Abstract
Online video-based learning is popular among the global student community. We investigate ways of leveraging Artificial Intelligence (AI) to reduce the video watching time without losing comprehension of the video content. Using an in-house designed video player prototype, we conducted an observation study to understand user behavior in adjusting the video playback rate. We present preliminary results from this study and discuss implications of our observations in designing potential AI-based solutions to semi-automatically adjust the video playing speed to reduce the watching time without affecting the users' comprehension.

Author Keywords
Online Video Learning; Video Player Interface; Video Speed; Video Speed Modelling; Speaker's Speech Rate; Video Speed Prediction.

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

Introduction
Online video-based learning has become more popular with MOOCs and YouTube. However, viewers might not have the time, patience or intention to watch through the whole video

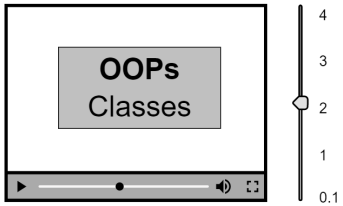

Figure 1: The proposed interface.

Video speed factors

Speaker's speech rate (SSR): SSR is the number of spoken speech units per time. Listening comprehension is inversely proportional to SSR [3]. It is the most common problem faced by EFL(English as a Foreign Language) learners [13]. SSR (syllables/second) is calculated using Praat Library [2] and Speech Rate script [5].

English proficiency: The limit of SSR a person can understand is lower for beginner language learners than that of advanced learners [11].

Prior knowledge in the topic of the video: Other barriers in listening comprehension include background knowledge and new vocabulary [13]. Thus, having more expertise in the topic might make it easier to comprehend.

at its normal playback rate. They might prefer to skip over parts of the video to search for specific content or watch at an increased playback rate to skim the content [1]. First time viewers usually consume the content linearly (without scrubbing) [8]. To preclude random viewing behavior while scrubbing, we scope our work to linear video consumption.

Research related to optimizing the playback speed of videos can be grouped into three categories: (1) interaction peak (play, pause, navigate, replay and quit) detection, (2) navigation interfaces and (3) viewing time reduction heuristics. Interaction peaks can be detected using changes in media type, topic, speech features such as speaker's speech rate (SSR) and help decide when to reduce video speed [8] but not increase it. Video navigation interfaces [7, 9] provide features like slide view, search, user notes for faster navigation to a video segment of interest but do not provide faster linear viewing. Pause removal heuristic [9] removes silent segments to shorten the video duration. However, silence in lecture videos might coincide with visual information key to understanding content and pauses might assist in speech comprehension. SmartPlayer [4] adapts the video speed according to visual scene changes but lecture videos usually have rich audio content and less visual changes.

Standard video player interfaces for online video lectures provide a drop down selection of discrete speeds to change the video speed which involves multiple user interactions and user decision making for selecting a preferable speed. Also, the limited discrete speeds does not guarantee an optimal speed at which the viewer watches a video without compromising the comprehension of its content. While much previous research and current video players have focused on designing better manual controlled video navigation techniques, with the advance of AI, we see an opportunity to combine AI and HCI to design a semi-automatic control for optimizing video playing speed for users.

Video Speed Control Interface

Online lecture platforms like YouTube[1], Udacity[2], edX[3] and Coursera[4] use YouTube[5] and HTML5[6] video players which provide a drop down of discrete speeds to change the video speed. This video speed control interface (VSCI) requires the viewer to divert focus to click a button that displays the drop down, view the options, determine an apt speed and select it. This might hamper video content understanding. Flexible design for VSCI is inspired from volume control slider found in operating systems and video players: a slider over many discrete values separated by a small steps can give the *"feeling"* of continuous control. Hence, we have proposed a VSCI slider as shown in Fig. 1 to control video speed from 0.1x to 4x in steps of 0.1x using the mouse scroll wheel. SmartPlayer [4] uses a continuous dial for video speed control. Both SmartPlayer and popular media player, VLC, provide speed steps of 0.1x. By reducing the cognitive load and effort, viewers can easily explore many speeds to reach the optimal video speed using this VSCI.

Leveraging AI to Control Video Speed

AI should adapt the video speed to the viewer's needs and abilities. Modelling the user, task, interaction and their relationship is the key to provide adaptive user interface. We will consider video's SSR, viewer's English proficiency and prior knowledge in the video topic to model video speed.

Modelling Temporal Changes in Speed

For the observation study, 10 participants were asked to watch (without seeking) 5 lecture videos using our interface, shown in Fig. 1, as fast as possible such that the video content is understandable. They were informed beforehand

[1] https://www.youtube.com/
[2] https://www.udacity.com/
[3] https://www.edx.org/
[4] https://www.coursera.org/
[5] https://developers.google.com/youtube/
[6] https://www.w3.org/standards/webdesign/audiovideo

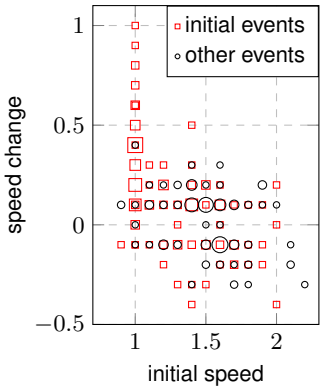

Figure 2: The relationship between the speed changed and initial of speed change events. The size represents number of occurrences. Initial events occur during the first 50s of the video and other events occur during the remaining duration.

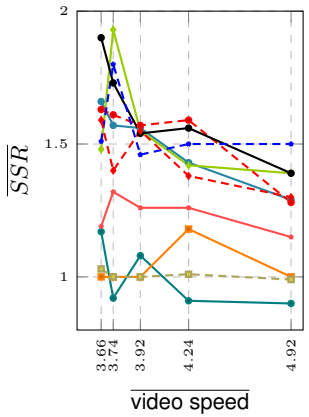

Figure 3: The relationship between the $\overline{SSR}$ and video speed

about follow-up questions to test their comprehension of the video content. The participants were 20 to 30 years old and included 7 females. They were professionals or students from various fields including computer science, engineering, science and arts. The videos were roughly 5 minutes long and the topics were: (1) bitcoins, (2) time management and computer scheduling, (3) agricultural revolution, (4) paleo diet, (5) LiDAR. Prior to watching the videos, they were asked to rate their English proficiency and topic expertise (both out of 10). After watching each video, 5 (2 direct and 3 derived) multiple answer questions were asked.They selected 66.4±15.2% of 42 correct options. SSR and participants' speed patterns for video 1 are shown in Fig. 4.

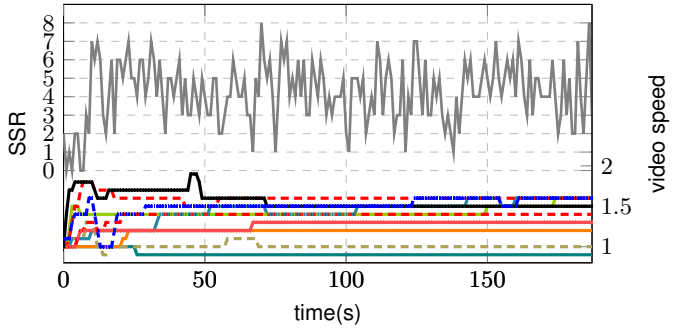

Figure 4: SSR (grey) and video speeds for video 1

Our findings from analyzing these speed patterns for various participants and videos in our observation study reveal:

- Participants change their speed significantly in the initial 50s of the video as seen in Fig. 2. A speed change event contains several speed changes such that two successive changes are less than 3s (seconds) apart and the speed changed in this event is the difference between the final and initial speeds. Average SSR is 2.53 words/s [12] and a spoken sentence contains an average of 6.22 words [10], so one sentence would last 2.45s.
- After 50s, participants maintain a constant speed and rarely change the speed by a small amount ($\leq$ 0.3x).

Based on these findings, AI should predict: (1) initial speed and (2) speed changes after 50s. The viewer has no idea of how SSR will change when watching the video. The viewer might not take the decision of changing the speed or make it slightly later because of more interest in watching the video than changing the video speed. Hence, AI can be used to compute when to change speed and by how much, based on previous viewer behavior, for optimal viewing.

Predicting Initial Viewer Speed

We need to model the relationship of video speed with the aforementioned factors. Since SSR and video speed are changing through the video, we will use their means $\overline{SSR}$ and $\overline{\textit{video speed}}$. Fig. 3 suggests that $\overline{\text{video speed}} \propto \frac{1}{\overline{SSR}}$. When the video speed is changed, the *Effective Speaker's Speech Rate (ESSR)* is *video speed* $\times\ SSR$. In Fig. 5, we can observe that each participant has their own comfortable range of $\overline{ESSR}$ ($= \overline{SSR} \times \overline{\textit{video speed}}$). Fig. 6 shows a correlation of 0.43 between grand mean of SSR ($\overline{\overline{ESSR}}$) and English proficiency. The participants were asked to self-rate their English proficiency which is a common practice. But the participants were not able to gauge their proficiency properly as noted in [6]. The correlation between $\overline{\overline{ESSR}}$ and expertise is -0.23. It seems difficult for the viewers to predict the video contents based on its title and rate their knowledge. Based on Fig. 5, Eqn. 1 can predict the initial speed. As shown in Fig. 5, there is a range of comfortable speed and the initial speed might not be optimal. To explore this range, speed is increased until the user interacts and reduces the speed. For this, the relationship of speed increase with these factors needs to be studied.

Speed Recommendation Model

AI should increase the speed during pauses (successive zeros in SSR) and SSR reduction. We assume that the viewer can best judge when to decrease speed. Slow speech segments are the intervals where the running averages of

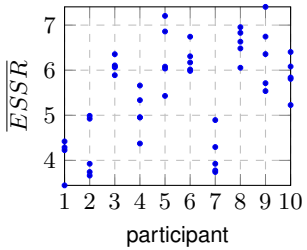

Figure 5: The plot of $\overline{\overline{ESSR}}$ for each participant and video

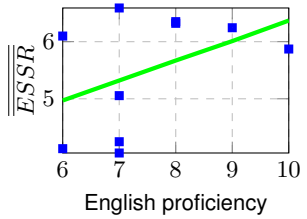

Figure 6: The relationship between $\overline{\overline{ESSR}}$ and participant's English proficiency

$$\text{new speed} = \frac{\overline{\overline{ESSR}}}{\text{new SSR}} \quad (1)$$

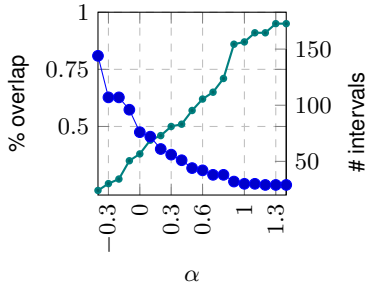

Figure 7: The % overlap (green) and count of filtered intervals (blue) for different α values

SSR (Hann window of 5) $< \tau$ and atleast 5s (2 sentences) long. τ is video dependent and is $\overline{SSR} + \alpha S_{SSR}$ where S_{SSR} is standard deviation of SSR and α is a control parameter. α is chosen based on two criteria: % overlap between filtered intervals and the start of speed change event and the number of intervals filtered. Note that the start time of speed change event is extended 5s earlier to incorporate possible viewer's delay to react to SSR change. Fig. 7 shows these two criteria versus α; there exists a trade-off between these criteria. Further study has to be done to determine optimal α. At the start and end of these intervals, the speed should be adjusted so that ESSR is maintained i.e. new speed $=$ (old speed $\times$ old SSR)/new SSR. The speed change should be capped by 0.3.

Acknowledgements

We thank the participants for participating in the observation study.

REFERENCES

1. 2013. As Data Floods In, Massive Open Online Courses Evolve. `https://www.technologyreview.com/s/515396/as-data-floods-in-massive-open-online-courses-evolve/`. (2013). Accessed: 2017-12-17.
2. Paul Boersma. 2006. Praat: doing phonetics by computer. *http://www. praat. org/* (2006).
3. Gary Buck. 1995. How to become a good listening teacher. *A guide for the teaching of second language listening* (1995), 113–131.
4. Kai-Yin Cheng, Sheng-Jie Luo, Bing-Yu Chen, and Hao-Hua Chu. 2009. SmartPlayer: user-centric video fast-forwarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 789–798.
5. Nivja H De Jong and Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods* 41, 2 (2009), 385–390.
6. David Dunning, Chip Heath, and Jerry M Suls. 2004. Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest* 5, 3 (2004), 69–106.
7. Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Daniel Li, Krzysztof Z Gajos, and Robert C Miller. 2014a. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 563–572.
8. Juho Kim, Shang-Wen Li, Carrie J Cai, Krzysztof Z Gajos, and Robert C Miller. 2014b. Leveraging video interaction data and content analysis to improve video learning. In *Proceedings of the CHI 2014 Learning Innovation at Scale workshop*.
9. Francis C Li, Anoop Gupta, Elizabeth Sanocki, Li-wei He, and Yong Rui. 2000. Browsing digital video. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 169–176.
10. Sharon Oviatt. 1996. Multimodal interfaces for dynamic interactive maps. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Acm, 95–102.
11. Willy A Renandya and Thomas SC Farrell. 2010. 'Teacher, the tape is too fast!' Extensive listening in ELT. *ELT journal* 65, 1 (2010), 52–59.
12. Jiahong Yuan, Mark Liberman, and Christopher Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. In *Ninth International Conference on Spoken Language Processing*.
13. Yajun Zeng. 2007. *Metacognitive instruction in listening: A study of Chinese non-English major undergraduates*. Ph.D. Dissertation.