AUTOMATED VIDEO GENERATION OF LECTURE VIDEOS FOR EMERGENCY REMOTE TEACHING DURING THE COVID-19 PANDEMIC

Christopher Cheong
Department of Information Systems and Business Analytics School of Accounting, Information Systems, and Supply Chain RMIT University, christopher.cheong@rmit.edu.au

Follow this and additional works at: https://aisel.aisnet.org/siged2020

Recommended Citation
https://aisel.aisnet.org/siged2020/24

This material is brought to you by the SIGED: IAIM Conference at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2020 AIS SIGED International Conference on Information Systems Education and Research by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
AUTOMATED VIDEO GENERATION OF LECTURE VIDEOS FOR EMERGENCY REMOTE TEACHING DURING THE COVID-19 PANDEMIC

Christopher Cheong
Department of Information Systems and Business Analytics
School of Accounting, Information Systems, and Supply Chain
RMIT University
christopher.cheong@rmit.edu.au

Abstract:
The COVID-19 pandemic forced a sudden shift to fully online delivery of learning in many higher education institutions across the world in 2020. Many academics used pre-recorded videos as a replacement for live in-person lectures. It can be time consuming to create these lecture videos and to update them for subsequent course offerings.

In this practice paper, we present an approach which automates the creation of lecture videos using presentation slides with speaker note text. We implement an Automated Video Presenter (AVP) that uses images of the presentation slides as the visual element of the video and uses text-to-speech (TTS) technology to synthesise speech from the speaker note text.

Having used this approach and the AVP-generated videos for two semesters, we discuss the benefits and the issues associated with them, as well as the lessons learnt. More importantly, we see wide ranging implications for both pedagogy and research and future work.

Keywords: lecture videos, automated generation, emergency remote teaching, COVID-19

I. INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) pandemic has accelerated the adoption of technology in many fields so that these industries can continue to operate while abiding by social distancing measures and various other restrictions to limit the spread of the virus. Education is one such area in which online and digital technologies are leveraged to ensure the continuation of education for many students.

Many universities shifted from face-to-face or blended learning to a fully online learning environment. Although the online medium is used, recent publications which investigate this approach to learning and teaching, termed “Emergency Remote Teaching” (ERT), state that, it is clearly something different (Bozkurt and Sharma 2020; Hodges et al. 2020). A significant difference is that it is a forced, sudden, and mass migration to online learning. The sudden shift of environment meant that academics did not have time to intentionally design their courses for a purely online environment. Instead, they had to unexpectedly adapt their existing materials for a new mode of delivery in a very short time.

A key issue with ERT during COVID-19 is that it is a change in teaching delivery. While it makes sense for some types of classes to be delivered synchronously online using video conferencing platforms such as Microsoft Teams¹, Zoom², and Collaborate Ultra³, other types of classes, such as lectures may benefit from asynchronous delivery as pre-recorded videos that can be accessed on demand and as many times as students wish. Over the years, there has been a myriad of studies on lectures (Bati et al. 2013; Dolnicar et al. 2009; Yeung, Raju, and Sharma 2016),

² https://zoom.us/
³ https://help.blackboard.com/Collaborate/Ultra
however, one clear trend is that lecture attendance has been decreasing for a long time (Rodgers 2001). Furthermore, we have anecdotally noticed a trend in lecturers either considering or replacing their live lectures with video recordings before the pandemic as a response to the decrease lecture attendance. Combining this trend with the need to shift to online delivery during the pandemic, it seems that COVID-19 may be a catalyst for the widespread use of video lectures.

The initial recording of video lectures can be time consuming and cumbersome. For example, it takes a lot of practice and time to present lectures with almost perfect speech, ambient environmental sounds can affect the quality of the audio recording (this is especially problematic as many academics are working from home during COVID-19), and video editing can be very time consuming. Furthermore, updating these videos is also time consuming as they are likely to require re-recording of certain sections and re-editing of the videos.

In this “practice paper” on “COVID-19, Learning, Pedagogy, and Educational Systems”, we present a technology-based approach to address a number of these key issues in the production of lecture videos. Our approach, named Automated Video Presenter (AVP), uses a combination of technologies to automate the generation of such videos using presentation slides with speaker notes. The presentation slides are displayed as the visual component of the video and text-to-speech (TTS) technology is used to synthesise the text from the speaker notes to the audio component.

We discuss the design and implementation of AVP in Section II and one of its applications in an Information Systems course in Section III. We then provide a discussion of our experience using AVP and the lessons learnt in Section IV.

II. DESIGN, IMPLEMENTATION, AND ISSUES

In this section, we present the technical details on how AVP automates the generation of videos using a combination of existing technologies. We also discuss our design and implementation choices.

The main inputs into AVP’s video generation process are the presentation slides as shown in Figure 1. As we wanted to support a wide range of academics who have varied technical skills, we opted to use PowerPoint (PPT) slides as the input. PowerPoint is commonly used to create presentation slides and has a much lower learning curve than other alternatives, such as LaTeX\textsuperscript{4}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Automated Video Generation Process}
\end{figure}

\textsuperscript{4} https://www.latex-project.org/
In our particular implementation, we used Python on a Mac to generate the various outputs from a PowerPoint file saved in the Microsoft PowerPoint Open XML (PPTX) format. As the PPTX Python libraries on a Mac cannot generate images of the slides, this is done manually by exporting the PPTX file in PowerPoint to an image format (PNG in our case). Thus, the two inputs to the automated video generation process are the PPTX file (with speaker notes) and the images of the slides.

The text in the speaker notes section is converted to audio files using text-to-speech (TTS) technology. While there are a number of TTS engines available, we used Google Cloud Text-to-Speech’s WaveNet engine\(^5\) as it generates natural-sounding speech in a variety of voices and multiple languages. The video is generated by combining the images of the slides (as the visual element) and the audio files of the speaker notes. In our case, we generated the video in MP4 format.

Generating transcript files is simply a matter of creating a file that contains all the text from the speaker notes of each PowerPoint slide. One issue with lecture videos is that it is time consuming to search for a particular segment as users need to (semi-randomly) click on different time points in the video to find the segment. This is somewhat easier with screencasts as the slides are presented as the visual element of the video and users can use the PPT slides as a guide for the search. However, this can be done more easily by searching for keywords. Thus, we created the transcript files with references to slide numbers and timing information (i.e., when the slide starts and ends in the video). This allows users to search the transcripts for keywords and then select the appropriate time in the video to view that particular slide.

Additionally, we used the Google Translate API\(^6\) to translate the speaker notes text. This allowed us to not only generate transcripts but to also generate audio files in a variety of languages.

A significant and practical challenge to using AVP, especially considering the limited time academics had, to develop lecture videos, is writing the speaker notes text. In many cases, academics do not have these on hand as they \textit{ad lib} their presentations based on their experiences and the slides during a live lecture. Thus, writing the speaker notes can be a time-consuming and laborious task.

There are two alternative solutions to typing the text in the speaker notes. First, academics can use dictation software to create the speaker notes text. These are existing features in some freely available software, such as Google Docs\(^7\). Second, if academics have existing live lecture capture videos Speech Recognition (SR) can be used to extract the text from the audio. The accuracy of the text extracted largely depends on the quality and clarity of the audio in the video. In many cases, the text needs to be significantly edited. Although this is time-consuming, it is usually faster that writing the speaker notes from scratch.

### III. APPLICATION OF AUTOMATED VIDEO GENERATION

In mid-March 2020, when our institution shifted its teaching delivery to a fully online mode, many course instructors had the option to deliver lectures in a variety of formats online. This included synchronous mode, such as live online lectures, and asynchronous mode, e.g., pre-recorded lecture videos.

As we already had the development of AVP in train, we opted to complete its development and create pre-recorded videos to replace the live in-person lectures. The creation of pre-recorded (non-AVP) video lectures can be time consuming, especially if they are created error-free (e.g., no misspeaking). However, the benefit is that they can be re-used for future semesters. With

---

\(^5\) [https://cloud.google.com/text-to-speech](https://cloud.google.com/text-to-speech)

\(^6\) [https://translate.google.com/](https://translate.google.com/)

\(^7\) [https://www.google.com.au/docs/about/](https://www.google.com.au/docs/about/)
AVP, we envisaged that it would decrease the recording time of creating the pre-recorded videos, one of the most time-consuming aspects of video creation.

Another time-consuming aspect of creating pre-recorded videos is writing the scripts for the lecture. As we had Echo360 recordings (live lecture capture) from the previous semester, we used Speech Recognition (SR) to convert the audio of the recording to text, which we then edited.

In semester 1 of 2020, with the sudden and unanticipated transition to fully online teaching that occurred in Week 3, we created pre-recorded lecture videos using AVP 1 – 2 weeks ahead of the course schedule. Once that was completed, we also created the lecture videos for Weeks 1 and 2 for the sake of completeness.

In the usual blended learning offering of the course (i.e., pre-pandemic), students were presented with a summary of the entire cohort’s performance for each of the 3 assignment milestones. These are typically presented as part of the lecture (using a separate set of slides from the formal lecture) and students can compare their individual performance to the cohort performance to gain some context about their results. We also used AVP to generate cohort-level feedback videos using these feedback slides.

In addition to making the AVP videos available, students are also provided with the lecture slides, the transcripts of the video files, and the Echo360 live lecture capture videos of the previous semester. We also used this approach in semester 2 of 2020.

IV. DISCUSSION

As is evident from Section III, the implementation of AVP is not technically challenging. Much of the AVP script involves manipulating data, using existing programming libraries, or using web Application Programming Interfaces (APIs). However, its applications and implications, which we discuss in this section, are wide-ranging.

Benefits and Issues

The benefits of AVP compared to manually pre-recording video lectures are clear and obvious. There is a reduction effort and time in both the initial creation of the videos and in subsequent updates (more significantly so for the latter). Additionally, the generation of video transcripts with time stamps facilitates keyword searches for learners.

From an instructor’s perspective, AVP greatly facilitated the sudden and almost immediate transition to fully online learning. By reducing the amount of effort and time required to create the initial videos for semester 1 and revising the videos for semester 2, this liberated the instructor to focus on addressing other pedagogical issues that arose due to the online delivery of the course.

Anecdotal evidence indicates that students generally do not have major issues with AVP videos. In the two semesters that AVP videos were used, two students (one from more than 100 students in each semester) opposed the synthesised voice. However, it seemed that their criticisms were related to more than just the synthesised voice, e.g., issues in adapting to COVID-19 enforced changes and fully online learning, and not understanding that an academic crafted the speech in the videos. Other students stated that they did not mind the synthesised voice as it meant that they did not have to read the lecture notes themselves. Some also stated that they sometimes watched the videos at a faster speed.

Another issue with using the synthesised voices in the videos is that it further reduces the “teacher presence” online. Some students found that particularly difficult as they are more dependent on the teacher-student rapport to support their learning.

8 https://echo360.com/
Implications for Pedagogy and Research

Using the translation capabilities of existing web APIs, it is possible to easily generate transcripts and, perhaps more interestingly, speech in a variety of languages. The implication is that it is possible to personalise these videos for learners based on their mother tongue, dominant or preferred language for learning.

An interesting aspect of using TTS is that it is easy (and cheap) to change the voice of the “speaker” (cf. re-recording the lecture videos with another speaker/voice actor). Not only is it possible to change the gender and accent of the voices, it is also possible to change other characteristics such as pitch and speech rate.

Studies have shown that computers (and by extension, synthesised voices) can be seen as social actors and that young children have social responses to computers. These responses can be used to exploit increases in recall and recognition in young children (Bracken and Lombard 2004). Furthermore, young children pay more attention to female voices compared to a male voices (Wright and Huston-Stein 1979). While other studies show that voice characteristics such as pitch and speech rate, can be used to promote cooperation (Oksenberg and Cannell 1988; Van der Vaart et al. 2006) and, thus, alluding to being able to use voice characteristics to positively persuade students to learn more effectively.

Studies on “foreign” accents show mixed results for learning. Some studies show that “foreign” accents negatively impact listening comprehension (Major et al. 2002) and some students have a preference for specific accents and recalled more information from these accents (Gill 1994). However, other studies show that accents do not necessarily impact learning, but they do negatively impact the perception of learning as students tend rate instructors with accents less favourably (Sanchez and Khan 2016).

So, for example, a female voice may be better to present negative feedback while a male voice is better for instructional videos, or Australian students may prefer learning by listening to an Australian accent, or a voice speaking English with a French accent may bring more authenticity to an online French class as students may subconsciously expect this.

With the range of accents and the different gendered voices available, it is also possible to diverge from contemporary teaching (e.g., instructional or explanatory videos) and more easily develop “personas” or characters in the lecture videos (Dotan et al. 2009; Jones, Floyd, and Twidale 2008). For example, a single academic could create a number of personas who, throughout a course, converse or discuss topics rather than a single academic explaining course topics to students.

As the automated generation of lecture videos is relatively new, researchers should investigate students’ perspectives on the AVP videos, and whether or not it facilitates their learning. The effect of the different synthesised voices, in terms of gender, accent, pitch, and speech rate on learning should also be investigated.

Lessons Learnt

Having used AVP for generating lecture videos for two semesters, we identified a number of important factors to facilitate their introduction and use in online courses. First, it is important to understand whether AVP videos are suitable for the particular course and how they align with the teaching philosophy and course design. As the course in which we introduced AVP videos contained self-paced lectures, these were not an issue.

Providing a range of alternatives to the AVP videos, such as the lecture slides, transcripts, and live lecture captures, allows learners flexibility in choosing how to engage with the materials. This also ensures that if particular students do not find the AVP videos suitable for their learning style, they have alternative materials to use.

As the AVP videos reduce online teacher presence, finding other ways to engage the students online is necessary. For the course in question, the tutorials were delivered online in a
synchronous manner and required students to engage with the tutor and each other in smaller “breakout groups”. This approach ensured that there was interaction and social presence online in the tutorials while the AVP lecture videos were provided as reference materials for the tutorials and assignments.

V. CONCLUSION

The technology-based approach of AVP to automate the generation of lecture videos based on presentation slides and speaker notes text accelerates and facilitates the creation and revision of such videos. It is particularly useful for cases in which lecture videos need to be produced quickly such as in ERT.

As the use of AVP videos are new, there are many future directions for future work. From a technological perspective, this includes refactoring the working AVP prototype into more stable software. From a pedagogical perspective, in addition to the suggestions made in Section IV, further investigating the use of short feedback videos may lead to improving the learning experience for students. In particular, instead of creating cohort-level feedback videos, it may be possible to create personalized feedback videos for each individual student in a course. Students may be more accepting of synthesized voices in such videos as it is more apparent that an instructor cannot personally create so many individual videos.

While it was not technically challenging to implement AVP, its implications for pedagogy is wide-ranging, can be impactful, and provides an exciting alternative to manually pre-recorded videos for both academics and learners.

VI. REFERENCES


Gill, Mary M. 1994. 'Accent and stereotypes: Their effect on perceptions of teachers and lecture comprehension'.

Hodges, Charles, Stephanie Moore, Barb Lockee, Torrey Trust, and Aaron Bond. 2020. 'The difference between emergency remote teaching and online learning', Educause Review, 27.

Jones, M Cameron, Ingbert R Floyd, and Michael B Twidale. 2008. 'Teaching design with personas'.


Wright, John C, and Aletha Huston-Stein. 1979. 'The Influences of Formal Features in Children's Television on Attention and Social Behavior'.

ACKNOWLEDGEMENTS

The author would like to acknowledge and thank Dr France Cheong and Dr Justin Filippou for feedback on drafts of this manuscript.

ABOUT THE AUTHOR

Christopher Cheong is an Associate Professor in the Department of Information Systems and Business Analytics at RMIT University, Australia. He has an academic background in Computer Science and his research focuses on the use of technology to enhance learning. He is the Editor-in-Chief for the Journal of Information Technology Education: Research and the Journal of Information Technology Education: Innovations in Practice, and a fellow of the Higher Education Academy.

Copyright © 2003, by the Association for Information Systems. Permission to make digital or hard copies of all or part 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 full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from ais@gsu.edu