"What did I just miss?!" Presenting ClassTranscribe, an Automated Live-captioning and Text-searchable Lecture Video System, and Related Pedagogical Best Practices

Mr. Chirantan Mahipal, University of Illinois, Urbana-Champaign

I'm a Computer Science grad student at University of Illinois, Urbana-Champaign, working under the mentorship of Prof. Lawrence Angrave. Prior to this, I was working as a Research Fellow at Microsoft Research in the Technology for Emerging Markets (TEM) group.

Prof. Lawrence Angrave, University of Illinois at Urbana-Champaign

Lawrence Angrave is an award winning Teaching Professor at the department of computer science at University of Illinois at Urbana-Champaign (UIUC). His interests include (but are not limited to) joyful teaching, empirically-sound educational research, campus and online courses, computer science, engaging underrepresented students, improving accessibility and creating novel methods to create, adapt and enhance learning opportunities and learning communities.

Mr. Yuren Xie, University of Illinois at Urbana-Champaign

I am a senior Computer Science and Statistics and Mathematics student studying at the University of Illinois at Urbana-Champaign. I really appreciate that I can live in a world where science and technology have been being used to make the world better, and it is my honor that I can get involved in this huge revolution. My belief is to make the world better with the combination of education and computer science.

Mr. Biswadeep Chatterjee, University of Illinois, Urbana-Champaign

Currently a undergraduate computer science student at the University of Illinois Urbana-Champaign

Mr. Hongyu Wang, University of Illinois, Urbana-Champaign

Hongyu Wang is currently a CS undergraduate student at the University of Illinois, Urbana-Champaign.

Mr. Zhengru Qian, University of Illinois, Urbana-Champaign

Computer Science Undergraduate

"What did I just miss?!" Presenting ClassTranscribe, an automated live-captioning and text-searchable lecture video system, and related pedagogical best-practices.

Abstract

Captioning of recorded video content has been available for over four decades, but its importance and pedagogical use has seen a recent revival due to the legal accessibility compliance requirements of MOOC and online content, and secondly, due to efforts by universities to bring digital accessibility to their on-campus learning environments.

We present an open-source web-platform, ClassTranscribe, and related pedagogical practices, to enhance student learning both during lectures, and during later searching and viewing of recorded lectures. Using recent advances in automated Speech-To-Text (STT) technologies, it is possible to provide low-cost, accurate, and timely, text-searchable recorded lecture videos, and in-class live-captioning that is informed by domain-specific words. Using a crowd-sourced approach, captions of lecture recordings can be incrementally improved above the limitations of automated STT approaches.

Secondly, we report on suggested technology and practices to support live class captioning for instructors who use PowerPoint, a document camera, live coding, and other presentation styles. Using Universal Design for Learning, we discuss lecture captions and recorded lectures as an assistive technology that can also assist all students. As an example of this, we describe how distractions in a lecture and a student's limited attention span may cause students to lose focus and context ("What did I just miss?!"), and fail to grasp the topic. By providing a live caption history on student devices, students can refocus, reconnect, and thus have an opportunity to learn the current lecture topic being presented.

The design of the ClassTranscribe platform is extensible and scalable. We demonstrate captioning of content by integrating with two websites used to host lecture videos, youtube.com and echo360.com.

Introduction

Today, undergraduate and graduate engineering students enroll in courses that employ live lectures, which may or may not include exposition, active learning and student-centered techniques (e.g., POGIL [1]), online video content, and a blend of multiple presentation formats. However, student and technology issues such as non-disclosed hearing impairments and

technology distractions, affect the learning potential of these experiences. We therefore ask, how can we improve the utility and digital accessibility of both live presentations and recorded video material? For brevity we will limit discussion to lectures and recorded videos, though the results, technology and practices presented have applicability to other pedagogical practices that employ social "live" human experiences and online audio-visual content, such as office hours, mentoring programs, video elaboration feedback embedded in quizzes [2], technology-enhanced remote study groups, and audio-video content designed for online learning platforms.

In this paper we suggest captions and transcriptions are an invaluable tool for learning in both live and recorded settings, from five varied perspectives: First, a caption history enables lost students to re-engage during a live presentation. Second, this technology offers an opportunity to address non-disclosed hearing impairments. Third, we introduce the availability of a new free tool, ClassTranscribe, that generates automatic transcriptions, and provides text-searchable video content. Fourth, we report on the successful use of ClassTranscribe in a limited number of large engineering classes at University of Illinois, a land-grant research university. Fifth, our results show College of Engineering undergraduate students valued captions for learning in the supported courses and advocated for the expansion of captions and transcriptions in other classes.

Previous studies in live lectures transcriptions include [3], [4]. Our new contributions for live-transcriptions are the use of modern Speech-To-Text systems that support greater accuracy and use of domain-specific words directly from presentation text, the ability to stream transcriptions using a web interface, and a report on using this system within the context of U.S. Engineering education.

Students' ability to absorb, learn, and use new content is dependent on the individual student [5], presenter variations [6] and the lecture format ([7], [8]). Active learning and student-centered methodologies, an engaging presentation style, interactive experiences, and other pedagogical best practices attempt to keep all students engaged and connected, however, these are not deployed in all engineering classes. Students may suffer from mental fatigue, they may multitask, or become distracted during the 60-90 minute duration of a typical lecture slot. Scholastic attempts to quantify learning and recall suggest that student retention is limited to 20-50% of the material immediately after a traditional lecture ([9],[10]). Attempts to explain difficulties in student learning are usually wrapped in repeated mantras such as "Attention in lecture is limited to the first 10-20 minutes." However the scientific rigor of studies that support these claims has been repeatedly questioned [5],[6]. Despite the lack of a robust, empirically-sound model of attention and learning in a lecture setting we can still ask: "How can we enable all students to learn more from a live lecture experience?" "Is it possible to address equality and make lectures more accessible and useful for learning, particularly for students with

non-disclosed impairments who may be unserved by centralized disabilities accommodations and services?" We discuss this opportunity next.

In compliance with Section 504 of the Rehabilitation Act of 1973 and Title II of the Americans with Disabilities Act of 1990, which prohibit discrimination on the basis of disability, U.S. universities provide centralized disability resource support and services for students with registered impairments and disabilities. For example a hearing-impaired student, after notifying the university and having their disability evaluated, may request that all videos be professionally captioned. Unfortunately, faculty receive these accommodation requests and assume that is the *only* form of impairment and accessibility issue that needs to be accommodated. Precise research in this field is difficult because we are attempting to quantify and circumscribe the effects on learning and career outcomes from non-disclosed impairments within the university environment. However, it is possible to quantify the prevalence of disabilities and impairments in the U.S. population and consider the fraction of students who request personal accommodations. The prevalence of significant hearing impairment at speech frequencies in the U.S. for adults aged 20-29 is 2.2% [11]. In a recent study [12] at the University of Wisconsin, only 26.2% of students who reported a disability in the survey disclosed their disability to the University. Recognizing the limitation that the reported fraction of unreported disabilities is for all disabilities, out of a typical class of 200 students, there will be an expected average of three students with non-disclosed hearing impairments.

Impairments and challenges to learning arise from disabilities and aging but also include temporary and contextual factors, that are often overlooked. For example, a burst eardrum, temporary hearing loss from loud music or sports injury, inner ear infection, accent-related language comprehension difficulties arising from non-native speaker, non-native language understanding by the student or projector fan noise may negatively impact the learning potential from the auditory information channel. Providing equivalent information in non-auditory form (e.g. text) may ameliorate these difficulties.

A key insight of the above examples is that all students may benefit at some time during their degree from access to multi-modal (audio, visual) forms of an information stream, and some students will greatly benefit from this support. Similarly, while the prevalence of students with significant physical hand-coordination difficulties may be a rare concern for instructors, a web application or phone app that supports voice commands not only supports students with dyslexia and limited physical motion, but benefits others as well: physically active students reviewing materials on a treadmill, or a parent multitasking with a carried baby sending a text message, for example. These are examples of the application of principles of Universal Design for Learning (UDL), which due to space limitations we do not discuss further. See [13] for an introduction and

[14] for a more recent accessibility-oriented discussion. Voice control of ClassTranscribe, discussed below, is planned for later 2019.

During a live lecture, there is no option to replay recently missed information. A disengaged student, upon hearing an interesting phrase or completing a technology-based distraction or mental break, will attempt to re-engage and reconnect with the current lecture content and context. The attempt to re-engage can fail due to an incomplete mental model and the inherent limitations of the student's immediate audio-visual working memory used to mentally replay recent exposition. This failure mode is the inspiration for this paper's title, "What did I just miss!?" With insufficient context, the learning opportunity of the current moment is lost, which may further cascade i.e. limit learning opportunities for the remainder of the lecture. A transcription history of recent presenter utterances provides the student with an invaluable time machine. By appending transcriptions in a scrollable window, a student is able to quickly review recent utterances. Thus, the transcript supplements the student's own working auditory-visual memory to allow the student to sufficiently rebuild the context of the current topic. As an example, consider an engineering lecture slide that presents a detailed equation and a graph. It is the supplemental and contextual information spoken by the presenter - and absent from the slide - that enables the student to follow the exposition, line of reasoning, and ultimately gain the intended insight and understanding. Empirical evaluation of if, and how, live captioning systems with transcription history can ameliorate student re-engagement failures in technical lectures is an interesting open research question.

We next present ClassTranscribe, a web application which automatically transcribes video content and enables students to review, search and discuss recorded video material in a course.

ClassTranscribe Architecture

ClassTranscribe is an open-source web application [15]. The original web-based experiment used a novel crowdsourcing technique where transcriptions were entered by all students in class [16]. The new version presented here has been re-written with web accessibility standards [17] and scalability in mind. It also supports automated transcription creation and, as inspired by the original crowdsourced version, allows students to fix transcription errors.

ClassTranscribe is an HTML5 web-platform built using javascript libraries (node.js, express), a SQL database and front-end web components that support digital accessibility. Docker, a cross-platform container environment, allows developers to work effectively on Windows, Linux and OSX platforms. Docker is also used to deploy new builds to the production server. Users login to ClassTranscribe with Instructor or Student permissions. Instructors can create or modify a course. Students may view courses. An instructor can add content to their course from one of

the three sources: A YouTube PlayList, Echo360 course public URL, or upload from their local computer. For the first two choices, ClassTranscribe automatically and routinely checks for new additions and automatically downloads and transcribes the latest recorded lectures. An overview of the data schema is shown in Figure 1.
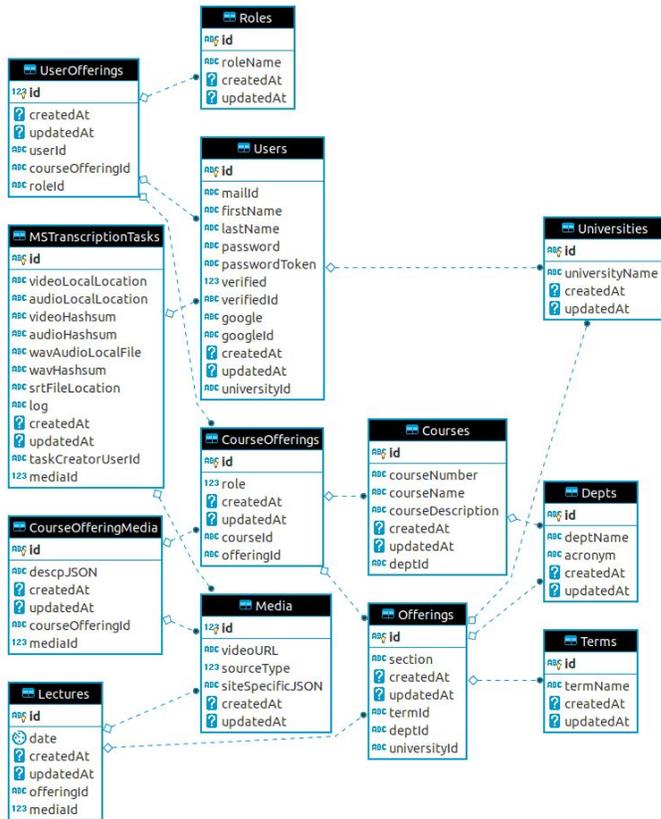

Figure 1. An overview of the main ClassTranscribe database schema.

Lectures are transcribed using Microsoft Cognitive Services, an Azure cloud-based AI web service that provides a Speech-To-Text (STT) service. Audio is extracted from the lecture video, transcoded into compatible audio format (PCM 16KHz), and sent in short segments to the STT service[18]. The results are assembled into a webvtt caption file. These transcriptions are published on the ClassTranscribe website and are available for editing by users. Each transcription is viewable as optional closed captions (embedded into the video playback area) and as searchable transcript beneath the video playback area.

Digital accessibility experts and web developers with web accessibility knowledge and experience reviewed (and continue to review) the ClassTranscribe web interface. They provided suggestions (in the form of bug reports) or HTML, style sheets or javascript patches to help ClassTranscrbe better support digital accessibility standards. An invaluable tool that helped developers discover and fix accessibility issues is the "AInspector." This tool is a free and

intuitive Firefox web browser plugin that will find, highlight, and explain Web Accessibility 2.0 compliance issues in a web page opened in the Firefox browser[19].

The "watchLectureVideos" is the most utilized web view. It comprises: (1) A view of all the course's lecture videos to allow a student to maintain a broader context of the current lecture; (2) A video playback area that includes speed controls and keyboard shortcuts to jump forward and backward; and (3) A live scrolling view of the transcriptions which allows scrolling back or forward to any point in the lecture. The current video's transcriptions or the entire course can be searched. In the screenshot (Figure 2) the user has typed "mid" and the transcriptions are interactively filtered to just the matching lines. Secondly, the user elected to fix a transcription error and is editing the topmost line. The user may also choose a different portion of the video by selecting a transcription line. The share button, provides a shareable URL link to a specific playback time of the video. The link is automatically copied to the user's clipboard and, for example, can be pasted directly into an email or the course's external discussion forum.
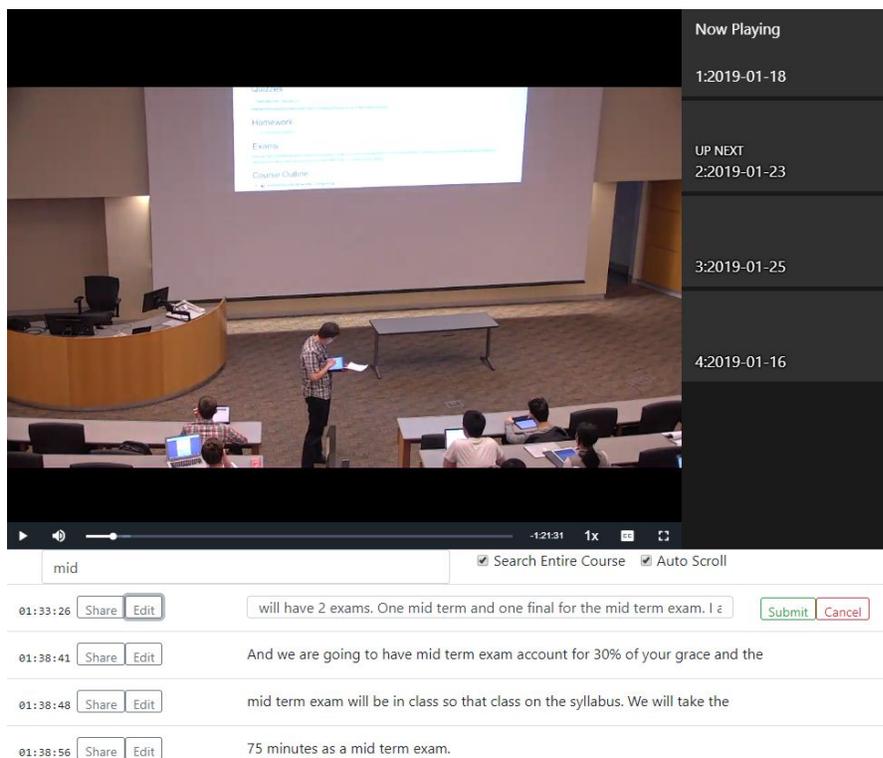

Figure 2. The ClassTranscribe video playback interface.

Introduction of live automated captioning technology to College of Engineering lectures in University of Illinois at Urbana-Champaign.

In the Fall 2018 semester, two 200-seat large auditorium-style lecture auditoriums were equipped with software and hardware for a semester-long trial of a live transcription service. A third, smaller 100-seat classroom was also outfitted but was not included in this study due to scheduling and IT difficulties outside the control of this study. Presenters in both large lecture rooms were actively supported via email and direct support before the start of each lecture at the start of the semester. Audio input to the transcription software on the podium PC was provided initially by a separate USB radio microphone (SAMSON XPD1 Lavalier), and later by an audio input tied to the existing radio-mic amplification system available in each room.

Engineering faculty who taught in these rooms (predominantly Computer Science faculty) were contacted by email and in-person support provided at the podium in the first week of classes immediately prior to the start of each lecture slot.

Automated transcriptions require high quality audio input. Incorrect audio settings at any stage of the audio pipeline may lead to distortion or inadequate signal. Presenters were unaware or uninformed about the cause when inadequate audio input caused transcription errors or absence of captions. There were large variations in the spoken loudness of each presenter, presenters manually tweaked the room radio-mic gain to adjust the in-room voice loudness, and the input gain of the audio volume on the podium PC was a global and persistent setting available for modification by any logged-in users, and further could be automatically modified by audio-video conferencing software that had been previously run by another user in an earlier lecture. The authors suggest a robust solution will benefit from software to automatically adjust the input gain on the transcribing machine, detect and prevent feedback loops, and provide visual alert if the input is too small, unavailable, or distorted. An enterprise-capable implementation might allow this to be centrally monitored without intervention of the presenter.

The authors asked all faculty that taught in these two rooms to use Microsoft PowerPoint together with a pre-installed add-in "Presentation Translator" to provide live captions. Presentation Translator[20] is a Microsoft-developed Windows-10 Office extension. In addition to optionally including closed captions (subtitles) on top of or below the PowerPoint slides, it streamed transcriptions via a website link. The link to the transcriptions was displayed as an additional slide that is included at the start of the presentation. The slide included a short dynamically generated URL that can be typed and an equivalent QR code. Students can open the transcription in a scrollable window on their laptop, tablet or phone. Live transcriptions were compatible with the Microsoft Translator app available on Android and iPhone platforms. The default caption position (top of the slides) visually overlapped with the title area. Most faculty preferred to move the captions to the lower edge.

Some presenters used the document camera and other (non-PowerPoint) presentation methods. To support transcriptions in these classes: (1) Presenters first logged into the podium PC and opened PowerPoint. (2) Optionally added keywords and domain words to a blank PowerPoint document to improve transcription accuracy. (3) Started captions and displayed the generated slide with the transcription link to the students; (4) After giving the students an opportunity to open the link on their own devices, switched the projector to their preferred application. The two major drawbacks of this method were that late-arriving students would need to ask their peers for the link and the significant setup burden on the presenter, especially for presenters who otherwise would not use the podium PC.

The Microsoft PowerPoint add-in internally used the same Speech-To-Text service to provide on-screen captions and an equivalent live transcription stream available through the web interface. To enable the transcriptions and captions, presenters started their presentation by using the Ribbon interface. Unfortunately, captions and transcriptions were not enabled if the presenter used the small lectern icon button to start the presentation. In-person support and conversations were beneficial in addressing this challenge. Direct emails to faculty were ineffective.

A feature of the add-in was the ability to pre-train the speech recognizer using domain-specific words in the PowerPoint document and provide more accurate transcriptions. Only two faculty were able to arrive early and wait 5 minutes, or pre-train the model. Other faculty prepared slides up to the very last moment, or arrived on or immediately prior to the official start time of the lecture.

Three instructors were enthusiastic to include live transcriptions as a regular part of their lecture workflow. Five faculty chose not to use the live-captioning system for their lectures. The reasons included (1) Due to the sensitive nature of the class material, they did not want their voice to be transmitted and possibly archived; (2) Expressed doubts about the value of transcriptions and concern for inaccuracies; (3) Incompatibility issues with their 12-year old "ppt"-formatted Powerpoint slides; and, (4) Difficulty and mental distraction of completing the technology setup immediately prior to starting lecture due to the short time gap between each lecture.

Student Survey on Learning

At the end of the Fall 2018 semester College of Engineering students in two lecture sections representing a total of six lecture hours per week. Students were rewarded with a 1% final exam bonus for completing the survey but the responses were collected and analyzed anonymously. A total of 217 students completed the survey (78% of the total enrollment). Students were asked their opinion on the usefulness of captions in live and recorded settings. Response choices were "Extremely useful," "Very useful," "Moderately useful," "Slightly useful," "Not useful at all,"

and, "I did not use this feature." There were 70 students (28%) who reported live captions as at least moderately useful; and 102 students (47%) reported captions on recorded videos as at least moderately useful. The survey results for live and recorded captions are summarized in Figure 3 and Figure 4, respectively.
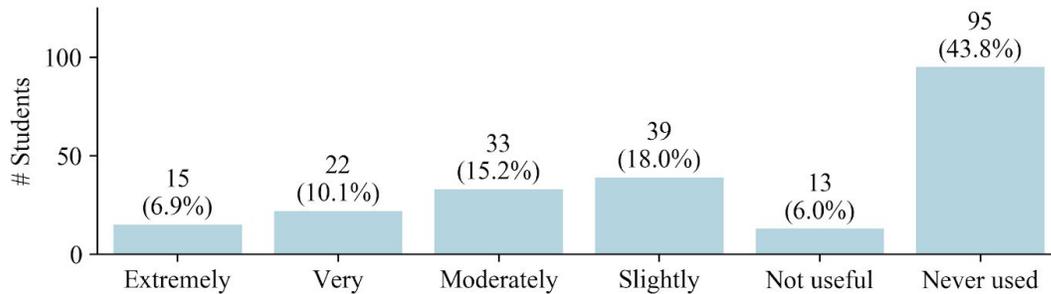

Figure 3. Utility of in-lecture live-captions for learning.
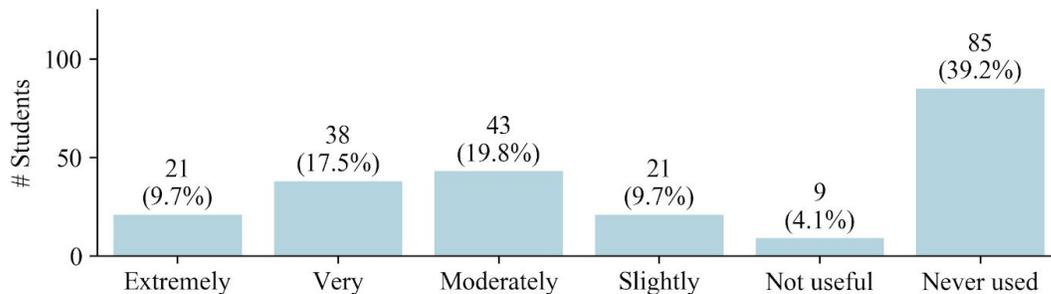

Figure 4. Utility of recorded video captions for learning.

Note in Fall 2018, students could watch recorded lecture videos with captions but searchable transcriptions and the ability to fix transcription errors were not yet available in the new version of ClassTranscribe. To see in-lecture captions students, would need to attend lectures and open the web link. To see recorded captions, students would need to enable closed captions. Approximately 40% of the class never used captions in either setting.

From the survey results we conclude that both live-captions and recorded video captions have wide student support for learning. Both in-lecture and recorded caption technologies were reported as desirable by significant fraction of the class (47% and 28%, respectively). Despite the shortcomings of the accuracy of automated transcriptions, the surveyed students expressed interest in the availability of this technologies in their other classes. Of the respondents, 74% wanted lecture recorded videos for their other classes and 30% wanted live captions in their other classes. When asked which classes, 34 students responded, and 41% of these responses were variants of "all classes."

The large fraction of students that reported captions as useful in both live and online settings is encouraging from a pedagogical perspective, and intriguing from a research perspective. Apriori the authors expected the moderately-useful or greater responses to be a small fraction (<10%) of the respondents. This study is not without its limitations. For example, perhaps students' interest in captions was positively modulated by the novelty of live and recorded captions in their engineering classes or awareness that these features were experimental, or erroneously believed that any new technology helped make the course materials easier to learn.

Transcribed video content enables the opportunity to text-search video content. When asked about relative value of features in a video playback system that included transcribed content, students ranked the ability to text search across multiple lecture videos as the most important feature (out of 7), and in-lecture text-searching as ranked second. These were ranked more important than popular video playback features in common use today including varied playback speed and fast navigation/rewind.

Transcription Costs

The PowerPoint add-in and its live captioning functionality was free. Microsoft recently announced similar functionality would be natively available in future versions of Powerpoint. ClassTranscribe used the cognitive services Speech-To-Text Azure-based API. The cost was minimal: $1 per audio hour [21], with free options available for early development (up to 5 hours per month). Other costs included additional radio mics ($100) that were no longer required once the audio input of the podium PC was integrated with the existing radio-mic system. Indirect costs included direct support time required to show faculty how to start live transcriptions in each room.

Future opportunities

Transcription and student behavioral data can empower a data-driven foundation for new engineering educational research. One advantage of ClassTranscribe, a custom video playback system, compared to existing commercial services (e.g., echo360.com, youtube.com) is the ability to provide student-level detailed behavior data to faculty and educational researchers. Educators and researchers can explore content usage, topics searched, and identify how and when students fall behind, and better understand learning directed behaviors outside of the classroom. Secondly, researchers may use natural language processing techniques to understand how instructors structure exposition, application and review and interlink their lecture content within and across lectures. Transcription data can be used to provide a real perspective on the depth, time, and breadth of topics actually presented in a course that is more accurate and transparent than a typical syllabus website document.

Conclusions

With minimal cost, technology exists today to provide lecture captions in both live lecture and recorded forms. The prevalence of hearing impairments in the adult population suggests that in large engineering classes (N>100) there is likely to be at least one student who may choose not disclose a hearing impairment. Captions and transcriptions for both live and recorded records can help students with non-disclosed hearing impairments which may arise from permanent disabilities, temporary, and contextual causes. In our study, a significant fraction of a large class also reported that both forms were useful to help learn technical material and advocated that these features should be available in all of their classes. This is consistent with Universal Design Learning principles. Our findings, and the systems we deployed, are examples of how new technologies can reshape engineering education for all, enable digital accessibility and provide a platform for evidence-based research of engineering education.

Acknowledgements

References

[1]   R. S. Moog and J. N. Spencer, "POGIL: An overview," *Process Oriented Guided Inquiry Learning (POGIL)*, vol. 994, pp. 1–13, 2008.

[2]   A. D. Fein, "Multimedia learning: principles of learning and instructional improvement in Massive, Open, Online Courses (MOOCS)," University of Illinois at Urbana-Champaign, 2017.

[3]   K. Bain, S. H. Basson, and M. Wald, "Speech recognition in university classrooms," in *Proceedings of the fifth international ACM conference on Assistive technologies - Assets '02*, 2002.

[4]   K. Ryba, T. McIvor, M. Shakir, and D. Paez, "Liberated Learning: Analysis of University Students' Perceptions and Experiences with Continuous Automated Speech Recognition," *E-Journal of Instructional Science and Technology*, vol. 9, no. 1, p. n1, 2006.

[5]   K. Wilson and J. H. Korn, "Attention During Lectures: Beyond Ten Minutes," *Teach. Psychol.*, vol. 34, no. 2, pp. 85–89, Jun. 2007.

[6]   N. A. Bradbury, "Attention span during lectures: 8 seconds, 10 minutes, or more?," *Adv. Physiol. Educ.*, vol. 40, no. 4, pp. 509–513, 2016.

[7]   G. S. Mason, T. R. Shuman, and K. E. Cook, "Comparing the Effectiveness of an Inverted Classroom to a Traditional Classroom in an Upper-Division Engineering Course," *IEEE Trans. Educ.*, vol. 56, no. 4, pp. 430–435, Nov. 2013.

[8]   K. E. Amaral and J. D. Shank, "Enhancing student learning and retention with blended learning class guides," *Educause Quarterly*, vol. 33, no. 4, 2010.

[9]   J. McLeish, "The Lecture Method (Cambridge Monograph on Teaching Methods, No. 1),"

Cambridge, England: Cambridge Institute of Education, 1968.

[10] M. W. Scerbo, J. S. Warm, W. N. Dember, and A. F. Grasha, "The role of time and cuing in a college lecture," *Contemp. Educ. Psychol.*, vol. 17, no. 4, pp. 312–328, Oct. 1992.

[11] H. J. Hoffman, R. A. Dobie, K. G. Losonczy, C. L. Themann, and G. A. Flamme, "Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years," *JAMA Otolaryngol. Head Neck Surg.*, vol. 143, no. 3, pp. 274–285, Mar. 2017.

[12] J. M. Love, "Wording Matters: the Impact of Disability Identification in Post-Secondary Education," University of Wisconsin Milwaukee, 2017.

[13] D. H. Rose and A. Meyer, *Teaching Every Student in the Digital Age: Universal Design for Learning*. Association for Supervision and Curriculum Development, 1703 N. Beauregard St., Alexandria, VA 22311-1714.

[14] C. Rogers-Shaw, D. J. Carr-Chellman, and J. Choi, "Universal Design for Learning: Guidelines for Accessible Online Instruction," *Adult Learning*, vol. 29, no. 1, pp. 20–31, Feb. 2018.

[15] "cs-education/classTranscribe," *GitHub*. [Online]. Available: https://github.com/cs-education/classTranscribe. [Accessed: 03-Feb-2019].

[16] J. C. Ren, M. Hasegawa-Johnson, and L. Angrave, "Classtranscribe: a new tool with new educational opportunities for student crowdsourced college lecture transcription," in *SLaTE*, 2015, pp. 179–180.

[17] w3c_wai, "Web Content Accessibility Guidelines (WCAG) Overview," *Web Accessibility Initiative (WAI)*. [Online]. Available: https://www.w3.org/WAI/standards-guidelines/wcag/. [Accessed: 03-Feb-2019].

[18] fmegen, "Speech SDK audio input stream concepts - Azure Cognitive Services." [Online]. Available: https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-use-audio-input-streams. [Accessed: 03-Feb-2019].

[19] J. Gunderson, "AInspector WCAG by Jon Gunderson Firefox add on," *AInspector WCAG*, 18-Oct-2018. [Online]. Available: https://addons.mozilla.org/en-US/firefox/addon/ainspector-wcag/. [Accessed: 01-Feb-2019].

[20] "Presentation Translator - Microsoft Translator," *Microsoft Translator*. [Online]. Available: https://translator.microsoft.com/help/presentation-translator/. [Accessed: 03-Feb-2019].

[21] "Cognitive Speech Services Pricing | Microsoft Azure." [Online]. Available: https://azure.microsoft.com/en-us/pricing/details/cognitive-services/speech-services/. [Accessed: 03-Feb-2019].