Important New Enhancements to Inclusive Learning using Recorded Lectures

Mike Wald

ECS, University of Southampton, Southampton, UK

m.wald@soton.ac.uk

Abstract. This paper explains three new important enhancements to Synote, the freely available, award winning, open source, web based application that makes web hosted recordings easier to access, search, manage, and exploit for learners, teachers and other users. The facility to convert and import narrated PowerPoint PPTX files means that teachers can capture and caption their lectures without requiring institution-wide expensive lecture capture or captioning systems. Crowdsourcing correction of speech recognition errors allows for sustainable captioning of any originally uncaptioned lecture while the development of an integrated mobile speech recognition application enables synchronized live verbal contributions from the class to also be captured through captions.

Keywords: speech recognition, recorded lectures, learning.

1 Introduction

This paper explains three new important enhancements to Synote1 [1], the freely available, award winning, open source, web based application that can make any public web hosted recording easier to access, search, manage, and exploit for learners, teachers and other users. Commercial lecture capture systems (e.g. Panopto[2], Echo360[3], Tegrity[4], Camtasia[5]) can be expensive and do not easily facilitate educational student interactions. Synote overcomes the problem that while users can easily bookmark, search, link to, or tag the WHOLE of a recording available on the web they cannot easily find, or associate their notes or resources with, PART of that recording [2]. As an analogy, users would clearly find a text book difficult to use if it

[1] www.synote.org
[2] http://www.panopto.com/
[3] http://echo360.com/
[4] http://www.tegrity.com/
[5] http://techsmith.com/Camtasia

had no contents page, index or page numbers. Synote can use speech recognition to synchronise audio or video recordings of lectures or pre-recorded teaching material with a transcript, slides and images and student or teacher created notes. Synote won the 2009 EUNIS International E-learning Award[6] [7] and 2011 Times Higher Education Outstanding ICT Initiative of the Year award[8]. The system is unique as it is free to use, automatically or manually creates and synchronises transcriptions, allows teachers and students to create real time synchronised notes or tags and facilitates the capture and replay of recordings stored anywhere on the web in a wide range of media formats and browsers. Synote has been developed and evaluated with the involvement of users and with the support of JISC[9] and Net4Voice[10]. Fig. 1 shows the Synote player interface. The technical aspects of the system, including the Grails Framework and the Hypermedia Model used, have been explained in detail elsewhere [3]. The synchronised bookmarks, containing notes, tags and links are called Synmarks (see Fig. 2). When the recording is replayed the currently spoken words are shown highlighted in the transcript. Selecting a Synmark, transcript word or Slide/Image moves the recording to the corresponding synchronised time. The provision of text captions and images synchronized with audio and video enables all their communication qualities and strengths to be available as appropriate for different contexts, content, tasks, learning styles, learning preferences and learning differences. Text can reduce the memory demands of spoken language; speech can better express subtle emotions; while images can communicate moods, relationships and complex information holistically. Synote's synchronised transcripts enable the recordings to be searched while also helping support non native speakers (e.g. international students) and deaf and hearing impaired students understand the spoken text. The use of text descriptions and annotations of video or images help blind or visually impaired students understand the visually presented information. So that students do not need to retype handwritten notes they had taken in class into Synote after the recording had been uploaded notes taken live in class on mobile phones or laptops using Twitter[11] [12] can be automatically uploaded into Synote. Until Microsoft Office 2010 was published with its undocumented changes to its saved .PPT format Synote could successfully create synchronized and searchable audio, transcripts and slides (including titles, text and notes) from narrated PowerPoint slides. The ability to import narrated PowerPoint files means that anybody can capture their lectures without requiring institution wide expensive lecture capture systems and section 2 of this paper reports on the development of a new PPTX to Synote xml format converter so that PowerPoint 2010 can be used successfully for recording lectures for Synote. Synote builds on 12 years work on the use of speech recognition for learning in collaboration with IBM, and the international

[6] http://www.ecs.soton.ac.uk/about/news/2598
[7] http://www.eunis.org/activities/tasks/doerup.html
[8] http://www.ecs.soton.ac.uk/about/news/3874
[9] http://www.jisc.ac.uk
[10] http://spazivirtuali.unibo.it/net4voice/default.aspx
[11] http://twitter.com/synote
[12] http://www.ecs.soton.ac.uk/about/news/2812

Liberated Learning Consortium [4] [5]. The integration of the speaker independent IBM Hosted Transcription System with Synote has simplified the process of transcription giving word error rates of between 15% - 30% for UK speakers using headset microphones. This compares well with the National Institutes of Standards (NIST) Speech Group reported WER of 28% for individual head mounted microphones in lectures [6]. Commercial rates for manually transcribing and synchronising a lecture recording are typically around £2/minute[13] (rates vary dependent on quality and quantity) and for automatic speech recognition to be used sustainably it must therefore cost less than this; including the manual correction of speech recognition transcription errors. A possible sustainable approach to obtaining accurate transcriptions is described in section 3 and involves students in the classes themselves correcting errors they find in the transcript, either voluntarily or through being paid or through being given academic credit. The requirement of using headset microphones to obtain good speech recognition transcription accuracy means that contributions from students in the class are not easily recorded or transcribed. To address this problem Syntalk, a mobile transcription server, has been developed and is described in section 4. Section 5 summarises some evaluations that have been undertaken.

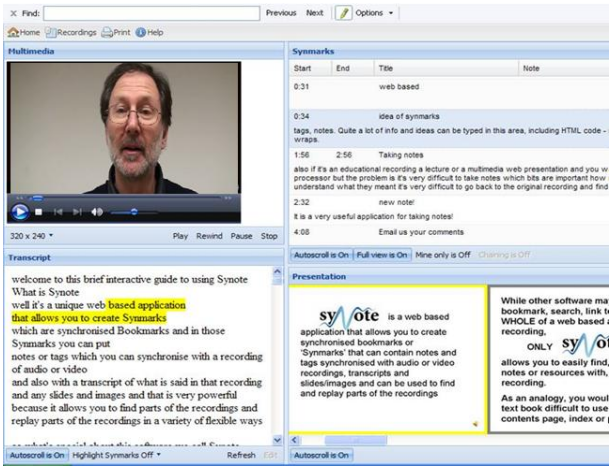

Fig. 1. Synote Player Interface

[13] http://www.automaticsync.com/caption/

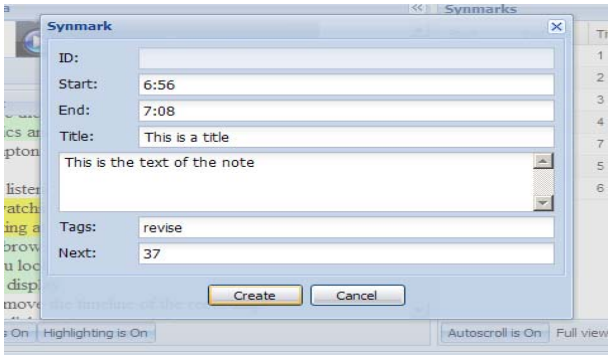

Fig. 2. Synote Synmark Creation

2 PowerPoint PPTX Converter

A narrated PowerPoint file can be used by a teacher to capture their lecture but each student would have to download the PowerPoint file to replay it on their own system. Office 2010 allows the narrated PowerPoint to be converted into a video that could be replayed by students but the video would not be captioned and the slide text would not be readable by a screen reader. The Synote PowerPoint converter enables the user to simply caption the recording using the slide notes and creates screen reader accessible text annotations for the slide images. The original Synote PPT converter was written in Java using the Apache POI9 library[14] (POI-HSLF) which only provides an API for data extraction for Microsoft PowerPoint's original PPT format. There is no Java library available that supports Microsoft's 2010 PPT format file. The new Synote

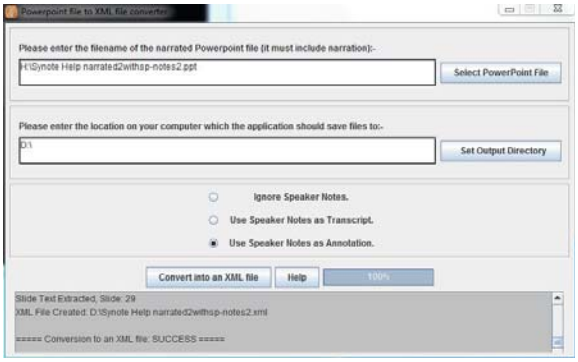

Fig. 3. PowerPoint PPTX to Synote Converter

[14] http://poi.apache.org/

PPTX converter (Fig. 3) changes the extension of the .PPTX file to .zip and extracts the slide text data and timing information and concatenates the audio .wav files that PowerPoint saves for each slide. The user needs to manually save the slide images as .png files. As PowerPoint does not record audio during slide transitions either the lecturer should not use any slide transitions or they should not speak during these transitions. The converter can automatically synchronise any text that the lecturer types into the slide notes with the narration either as a transcript or as an annotation.

3 Crowdsourcing Speech Recognition Transcription Correction

Universal Subtitles[15] is a recent Mozilla Drumbeat project designed to allow users to manually caption web based videos but only allows one person at a time to create or edit the captions. YouTube and Synote both enable automatic speech recognition captioning of videos but also allowed only one person at a time to correct its captions. Since there is no correct version of the transcript in existence there is no way of knowing whether the person creating or correcting the captions is making errors or not. The approach that we have adopted therefore is to allow many people to edit the captions and then compare their edits. The newly developed crowdsourcing correction tool shown in Figure 4 stores all the edits of all the users and uses a configurable matching algorithm to compare users' edits to check if they are in agreement. The tool

Fig. 4. Crowdsourcing Correction Tool

[15] http://www.universalsubtitles.org

allows utterances from specified sections of the transcript to be presented for editing to particular users or for users to be given the freedom to correct any utterance. Administrator settings allow for different matching algorithms based on the closeness of a match and the number of users whose corrections must agree before the system accepts the edit as 'correct'. The red bar on the left of the utterance indicates to a user that they are not allowed to edit the utterance and the white on green tick on the right denotes that a successful match has already been achieved and so no further editing of the utterance is required. The green bar on the left of the utterance denotes that the required match for this utterance has yet to be achieved. Users can be awarded points for a matching edit and it is also possible to remove points for corrections that do not match other users' corrections.

4 Captioning Contributions From Students Using Syntalk

Syntalk (Fig. 5) consists of two applications: an Android application which is used by students to capture and transcribe and if required also correct their utterances and a web application (Fig. 6) which is used by lecturers for managing the system. Users can choose to use any of three different free server based speech recognition systems, Google, EML[16] or iSpeech[17] . At the start of a lecture the lecturer makes their lecture 'live' using the Syntalk web application control panel. Users can then select this live lecture on their Syntalk mobile application. When the user talks into their mobile's microphone the Syntalk mobile application sends the speech to the speech recognition server and when the transcribed text is returned by this server to the Syntalk application it is then sent to the Syntalk web server as well as being displayed on the mobile's screen for editing. If the user chooses to edit any speech recognition errors the corrected text is then also sent to the Syntalk server which creates an XML file containing the text captions and timings which can be uploaded into Synote as

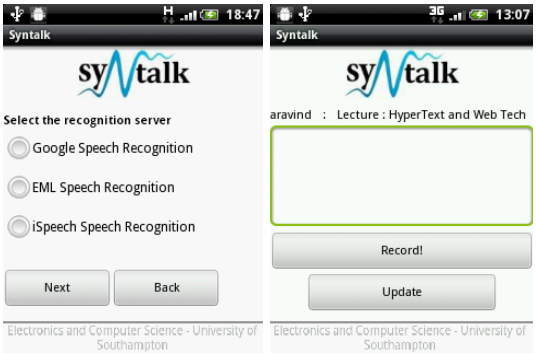

Fig. 5. Syntalk Android Application

[16] http://www.eml-development.de
[17] http://www.ispeech.org

Fig. 6. Syntalk Web Application

synchronized annotations. If everybody in a class used the Syntalk application on their personal mobile phone it would be possible to transcribe all spoken interactions. The current Syntalk application does not capture the spoken audio for Synote to replay.

5 Classroom Use and Evaluation of Synote

Since 2008 Synote has been used by teachers in universities in the UK, Italy, Germany, Pakistan, Australia, US and Canada with over 1000 recordings publically available on Synote (most with synchronised transcripts) for students to use for their learning. Dr Wald has used Synote with over a hundred recordings of his lectures with synchronised transcripts and slides for his teaching of many hundreds of students on undergraduate and postgraduate Electronics and Computer Science (ECS) modules at The University of Southampton. The provision of a verbatim synchronised transcript enables students to concentrate on learning and take only brief synchronised notes in Synote (e.g. 'revise this section for exam', 'I don't understand this fully' etc.). This feature is of value to all students, not only deaf students who need to lipread or watch a sign language interpreter and so can't take notes or dyslexic students or non-native speakers who find it difficult to take notes. The fact that Synote is used and valued by all students means that non native speakers and disabled students feel more included and do not have to use special technology. Also the quality of recording from a teacher's wireless head worn microphone is significantly better than from small personal digital recorders placed by students at the front of the class to record lectures. Questionnaire results from hundreds of students with a wide range of abilities and disabilities confirms that Synote successfully supports most browsers, is easy to use and improves learning, attention, motivation, efficiency, enjoyment, results and notetaking. Students also want all their lectures to be presented on Synote. The PPTX converter was successfully trialed by 60 MSc students producing publically available accessible narrated captioned online presentations on Synote of their accessibility evaluations of

web and software applications. The feedback provided by the users enabled the PPTX converter to be improved with regard to usability and robustness. An evaluation of the Syntalk Android application conducted with 25 users having different levels of computer skills showed that the application was easy to use to capture and transcribe students' contributions. Users preferred Google speech recognition because of its recognition accuracy and transcription speed but users would like the latency/time delay in receiving the captions back from the Server to be reduced. The crowdsourcing editor was trialed with ten undergraduate students who understood the recorded material better as a result of the editing process suggesting that a marks incentive for editing might be justified on educational grounds. Some improvements were also suggested to the system which continues to be further developed.

6 Conclusion and Future Work

Commercial lecture capture systems are expensive and do not easily facilitate educational student interactions. Synote has been shown to provide very well received enhancements to web based teaching and learning from recordings and to integrate well with other applications including PowerPoint, Twitter and Speech Recognition Software. The PPTX to Synote XML converter enables lecturers to easily capture their lectures and replay them accessibly using Synote. Syntalk provides a simple and free way to also capture and accessibly display the rich student interactions that can occur in classrooms. There could be great educational benefits and a huge demand for speech recognition lecture transcription if our crowdsource editing tool makes it sufficiently accurate and affordable. A demonstration and further results of the use of the systems described in this paper will be presented at ICCHP 2012.

References

1. Wald, M., Wills, G., Millard, D., Gilbert, L., Khoja, S., Kajaba, J. and Li, Y. Synchronised Annotation of Multimedia. 2009 Ninth IEEE International Conference on Advanced Learning Technologies. 2009 594-596.
2. Whittaker, S., Hyland, P., Wiley, M. Filochat handwritten notes provide access to recorded conversations, Proceedings of CHI 1994 271-277.
3. Li, Y., Wald, M., Wills, G., Khoja, S., Millard, D., Kajaba, J., Singh, P. and Gilbert, L. (2011) Synote: development of a Web-based tool for synchronized annotations. New Review of Hypermedia and Multimedia. pp. 1-18.
4. Leitch, D., MacMillan, T. Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research. Saint Mary's University, Nova Scotia. 2003 http://www.liberatedlearning.com/
5. Wald, M. and Bain, K. Enhancing the Usability of Real-Time Speech Recognition Captioning through Personalised Displays and Real-Time Multiple Speaker Editing and Annotation. In Proceedings of HCI International 2007: 12th International Conference on Human-Computer Interaction, Beijing pp 446-452.
6. Fiscus, J., Radde, N., Garofolo, J., Le, A., Ajot, J., Laprun, C., (2005) The Rich Transcription 2005 Spring Meeting Recognition Evaluation, National Institute Of Standards and Technology