Synote: Accessible and Assistive Technology Enhancing Learning for All Students

Mike Wald

School of Electronics and Computer Science,
University of Southampton, United Kingdom
M.Wald@soton.ac.uk

Abstract. Although manual transcription and captioning can increase the accessibility of multimedia for deaf students it is rarely provided in educational contexts in the UK due to the cost and shortage of highly skilled and trained stenographers. Speech recognition has the potential to reduce the cost and increase the availability of captioning if it could satisfy accuracy and readability requirements. This paper discusses how Synote, a web application for annotating and captioning multimedia, can enhance learning for all students and how, finding ways to improve the accuracy and readability of automatic captioning, can encourage its widespread adoption and so greatly benefit disabled students.

1 Introduction

Text transcriptions of the spoken word can benefit deaf people and also anyone who needs to review what has been said (e.g. at lectures, presentations, meetings etc.) The provision of synchronized text captions (subtitles) and images with audio and video enables all their different communication qualities and strengths to be available as appropriate for different contexts, content, tasks, learning styles, learning preferences and learning differences. For example, 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. Deaf learners and non-native speakers may be particularly disadvantaged if multimedia involving speech is not captioned while visually impaired learners may be disadvantaged if accessible text or spoken descriptions of visual information is not available. Multimedia has become technically easier to create (e.g. recording lectures) but while users can easily bookmark, search, link to, or tag the WHOLE of a podcast or video recording available on the web they cannot easily find, or associate their notes or resources with, PART of that recording. As an analogy, users would clearly find a text book difficult to use if it had no contents page, index or page numbers and they couldn't insert a bookmark or annotation. Therefore the growing amount of knowledge available in multimedia formats has yet to achieve the level of interconnection and manipulation achieved for text documents via the World Wide Web and so realize the exciting opportunities for learning. This paper discusses how captioning has the potential to enhance learning for all students through making multimedia web resources (e.g. podcasts, vodcasts, etc.) easier to access, search, manage, and exploit for

K. Miesenberger et al. (Eds.): ICCHP 2010, Part II, LNCS 6180, pp. 177–184, 2010.

learners, teachers and other users through supporting the creation of synchronised transcripts, notes, bookmarks, tags, links and images. The paper also argues that the availability of accurate automatic captioning using speech recognition (SR) will encourage the widespread adoption of this technology and so also greatly benefit disabled students. Research has confirmed the importance of captions for searching recordings and reading the transcripts and the value of also being able to personally annotate the recordings (e.g. bookmarks, notes and tags) and search these annotations [1], [2], [3], [4]. Speech recognition has been demonstrated to provide a cost-effective way of automatically creating accessible text captions and transcripts synchronised with audio and video and so allowing audio visual material to be manipulated through searching and browsing the text [5]. Although real-time captioning using phonetic keyboards can provide an accurate live transcription for deaf people, it is often not available because of the cost and shortage of highly skilled and trained stenographers [6]. Real-time SR is required for deaf and hard of hearing students and non native speakers whereas non real-time SR may often suffice for others and will provide more time for processing allowing for more accurate methods of SR to be used. Available SR systems (e.g. Dragon, ViaVoice [7]) can be adjusted to provide the fastest recognition (e.g. real-time) or to provide a slower but more accurate text output. It is possible therefore to provide fast but less accurate real-time recognition in a lecture and then later replace the transcript with one produced using slower but more accurate non real-time recognition. A study of the use of SR captioning in classrooms [8] showed that students with disabilities liked the fact they were not the only people to benefit from the technology as it drew the entire class into a collective learning experience and so making the recording and captioning of lectures standard procedure in universities would be of great benefit.

2 User Needs

User needs analysis has identified many benefits of annotating multimedia recordings in the ways proposed in this paper. It will for example enable learners to: search text transcripts, slides and notes and then replay recordings from that point; read captions to support learning style preference, deafness, or second language; read text descriptions of visual information (in videos and/or slides); use the transcript of lectures to support their understanding of text books and academic articles if they find the more colloquial style of transcribed text easier to follow than an academic written style; insert a bookmark in a recording so as to be able to continue later from where they left off; link to sections of recordings from other resources (e.g. documents, web pages etc.) or share these sections with others; tag and highlight sections of recordings/transcripts they don't understand fully so they can revisit them later for clarification; annotate recordings with notes and URLs of related resources (e.g. documents, websites etc.) at specific places in a recording to clarify issues and support revision; tag recordings using their own terms as a personal index. It will also enable teachers/lecturers to: index their recordings using syllabus topic tags; provide synchronized slides and text captions to accompany podcasts; provide text descriptions

for visually impaired students of visual information; identify topics needing further clarification from the pattern of learners' 'not understood' tags; provide feedback on learner-created recordings of presentations; ask learners to annotate recordings to provide evidence of their group contributions; analyse unstructured tags learners use (folksonomy) to help create structured tags (ontology); tag recordings with URLs of related resources (e.g. documents, websites etc.); link to and use sections of existing multimedia without having to edit the recording. While some proprietary existing systems can synchronise students' notes with teachers' presentations and recordings, they do not provide captioning (e.g. Tegrity [9]). Since no existing technology was found to satisfy all the identified user needs it was decided it was therefore necessary to develop Synote, a new web application that: Works with web multimedia and stores annotations separately in XML format; Synchronises captions, images, tags, links, notes and bookmarks; Enables users to add and search for annotations quickly and easily; Supports private or shared annotations; Is accessible to people with any disability.

3 Automatic Speech Recognition Captioning

Commercially available SR software (e.g. Dragon, ViaVoice) is unsuitable for transcription of speech in lectures as without the dictation of punctuation it produces a continuous unbroken stream of text that is very difficult to read or comprehend. Liberated Learning (LL) and IBM therefore developed ViaScribe [10] [11] as a SR application that automatically formats synchronized real-time text captions from speech with a visual indication of pauses. Detailed feedback from students, and lecturers [8] showed that this approach enhanced teaching and learning if the text transcription was reasonably accurate (e.g. <15% word error rate). Similar results have been reported by others using the Microsoft SR Engine [12] or Dragon [13]. To improve accuracy and understanding of SR captions an application RealTimeEdit was developed to enable corrections of SR captions to be made in real-time [14]. IBM has also recently developed a speaker independent recognition engine, Attila [15] that can that can be hosted on the web, interfaced with other applications and used to transcribe, edit and display recordings created in a wide range of multimedia file formats. The Attila speech recognition system has only available a US and not a UK English voice and language model and has no simple facility for adding vocabulary or training to individual users' voices and so typically gives word error rates between 15% - 30% for UK speakers using headset microphones. This however compares well with the National Institutes of Standards (NIST) Speech Group reported WER of 28% for individual head mounted microphones in lectures [16]. For speech recognition systems to be able to transcribe lectures more accurately in addition to having 'local' (e.g. UK) language and voice models they need to be designed for education rather than for dictation. For example they need to: Be speaker independent with a speaker dependent training facility; Be customisable for different subject domains; Cope with low quality speech signals and background noise; Recognise or ignore partial words (e.g. hesitations) or

'fillers' (e.g. 'um', 'er'). Commercial rates for manually transcribing and synchronising a lecture recording are typically around £2/minute [17] (rates vary dependent on quality and quantity) and so it would cost about £90 for transcribing and synchronising a 45 minute class. For speech recognition to be used it must therefore cost less than this including licensing, server and maintenance costs. Manual correction of errors will also be required if 100% accuracy is to be achieved. One possible sustainable approach to obtaining accurate transcriptions could be to devise a system where students in the classes themselves corrected errors they found in the transcript, either voluntarily or through being paid or through being given academic credit.

4 Synote Annotation System

A multimedia annotation system called Synote [18] that meets the identified requirements has been developed (supported by the JISC [19])) and trialed (supported by Net4Voice [20]), providing multimedia recordings synchronized with transcripts, slides images and bookmarks (called 'Synmarks' and containing titles, tags, notes, and links) [21]. Synmarks can also be used to provide information about sounds or tone of voice or emotions or synchronised links to videos of sign language translations. Synote has been integrated with Attila to simplify automatic transcription. Figures 1 & 2 show examples of Synote's interface while Figure 3 provides a system overview. When the recording is played the currently spoken words are shown highlighted in the transcript. Selecting a Synmark, transcript word or slide image moves the recording to the corresponding time. Synote also enables searching of the transcripts, Synmarks and PowerPoint slide titles, notes and text content. Users can also synchronise, or transcribe transcripts by hand if they don't have access to suitable SR software that automatically transcribes and synchronises speech and text. Synote is currently being used by teachers and students in universities throughout the world.

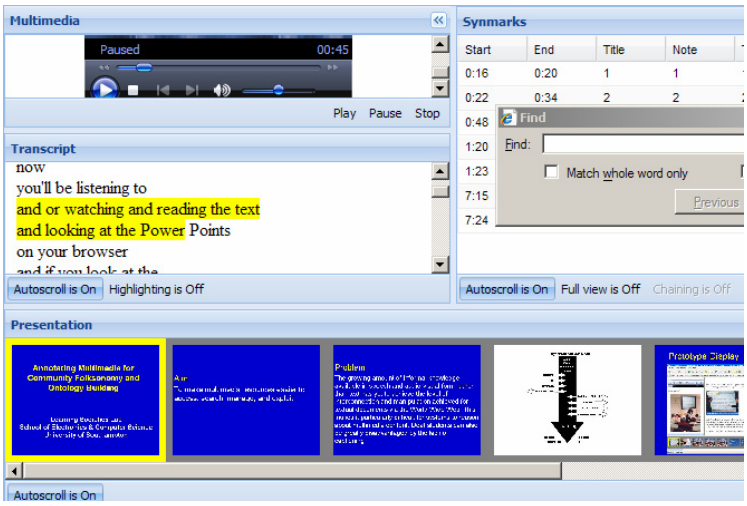

Fig. 1. Synote Interface

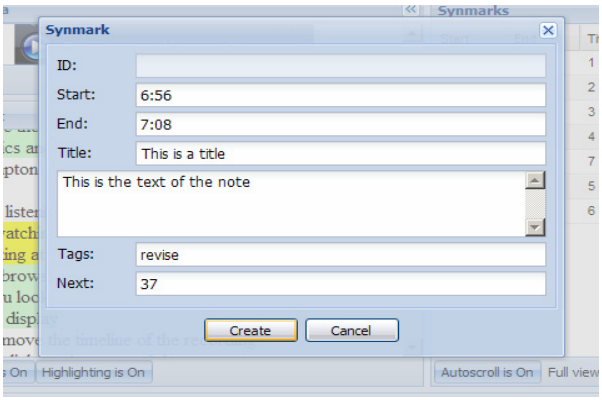

Fig. 2. Creation of Synmarks

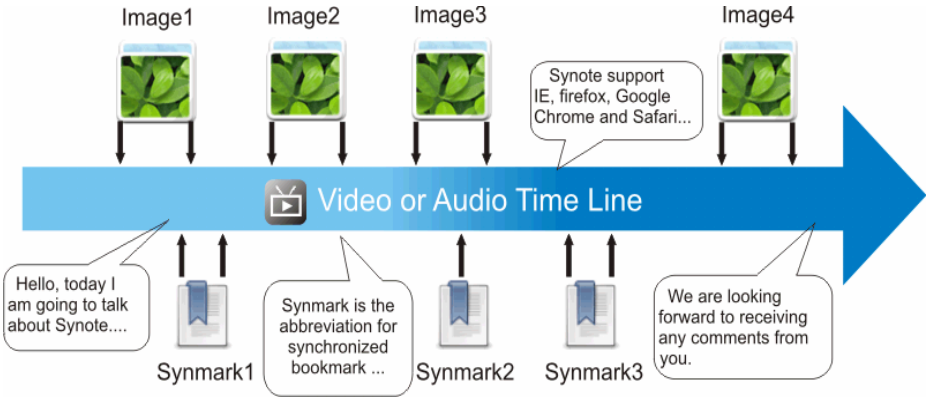

Fig. 3. System Overview

Since 2007 Dr Wald has used Synote with over 40 recordings of his lectures with synchronised transcripts and slides for his teaching of approximately 200 students on 5 undergraduate and postgraduate modules in The University of Southampton's School of Electronics and Computer Science (ECS). At the time of writing there are over 500 recordings available on Synote (most with synchronised transcripts) for teachers and students to use for teaching and learning. These include guest lectures presented by leaders in their field. Questionnaires and interviews with students and teachers who have used Synote have shown that both students and teachers like the system and have found it enhances learning and teaching and would like more recordings of classes to be made available in this way. In the University of Southampton School of Electronics and Computer Science (ECS) 5 classes with approximately 200 students (two 1st year BSc classes, one 3rd year BSc class and two Masters Degree classes) were surveyed, with 101 students filling in questionnaires about their experience with Synote. The questionnaire results showed that Synote's design to support a wide range of browsers was justified and appreciated with 54% using Internet

Explorer, 31% using FireFox, 12% using Google Chrome and 3% using Safari. The results also confirmed that Synote was easy to use as over 80% of the respondents didn't need to read or listen to the Synote guides and rated Synote as 4 or 5 on a 5 point scale of ease of use with the remaining percentage rating it as 3. The design of Synote to provide synchronised slides, video, audio and transcripts was also appreciated with over 80% respondents rating Synote as 4 or 5 for being useful overall as well as its slides, video and audio and transcript also each being useful with averages ratings of between 4.1 and 4.5. A slightly smaller percentage rated the Synmark option and the print out option as being useful with an average of 3.7 and 3.3 respectively on the 5 point scale used. One student commented that since there was no experience in using these facilities they were not being used but if Synote was used regularly then it would become second nature to use all these facilities. Ninety seven percent of the students wanted all their lectures to be presented on Synote. Forty four students in two of these five classes identified in more detail how using Synote affected their learning by indicating whether Synote improved or reduced or provided no change in the aspects of their learning shown in Table 1.

Table 1. Affect on Learning

	improve	no change	reduce
learning	95%	5%	0%
attention	61%	34%	5%
motivation	50%	50%	0%
efficiency	77%	23%	0%
enjoyment	66%	32%	3%
results	69%	31%	0%
notetaking	53%	30%	18%
attendance	13%	45%	42%

A few students commented that some students' notetaking skills might not develop if all classes were recorded and presented on Synote. Some students commented that if they were ill or had serious problems they would in the past normally still have tried to come in to classes because they were worried they would get behind in their work and would be unable to catch up, whereas with Synote they could stay at home and get well secure in the knowledge that they would not miss anything as they could learn using the Synote recording. This is clearly an important issue during flu epidemics. Students stated that it was important that that ALL lectures were recorded so they didn't find that the one lecture they missed hadn't been recorded. Of these 44 students 37% identified themselves as native speakers, 26% as fluent, 28% as having good language knowledge and 9% as having little language knowledge Non native speakers in particularly commented how valuable Synote was for them as it was sometimes difficult to understand lecturers' speech and note-taking in a foreign language was very difficult for them. One commented that they could get words not understood in the transcript translated by Google. Of these 44 students 7% identified themselves as having hearing disabilities, 2% visual disabilities, 11% learning disabilities and 7% other disabilities. Two overseas students wrote *"Synote gives a very clear understanding of module …It was very useful for me especially as I am a non English*

native speaker" and *"I think Synote is a very good way to listen to lectures. If for example we miss the lecture we can actually listen to it in our own time or if we didn't really understand the lecture we can go back to it and listen to it carefully. I also like the highlighted part whenever the lecturer speaks on the text so we can't actually get lost within long texts."* Other students wrote *"Synote is very useful for students in general, I think at present all they need is to get used to something like Synote. It will then become second nature"* and *"Synote is the best system I have ever seen for assistive technology it is very useful for me to understand what the lecturer taught after class I hope all school majors could integrate this system thanks".* Synote is used and valued by all students so non native speakers and disabled students feel more included and don't need to walk to the front of the class to ask the teacher if they can record the lecture on their personal digital recorder. Also the quality of recording from a teacher's wireless headworn microphone is significantly better than from small personal digital recorders placed at the front of the class. Students did not like retyping handwritten notes they had taken in class into Synote after the recording had been uploaded and so Synote has recently been enhanced so that synchronised notes taken live in class on mobiles or laptops using Twitter [22] [23] can be automatically uploaded into Synote. Google announced in November 2009 that they plan to use speech recognition to caption YouTube videos; a feature that has been available for nearly two years for students using Synote.

5 Conclusion

This paper has put forward the following premises and arguments: Real-time captioning can support deaf students and non native speakers; Transcripts obtained from captioning, can support disabled students who find difficulty taking notes (e.g. dyslexic, motor impaired, visually impaired, hearing impaired etc.); Manual real-time captioning using stenographers is expensive and there is a shortage of people trained to undertake this task; If captioning was seen to provide a great benefit to ALL students then the cost of captioning per student benefiting would decrease; If automatic captioning using SR could help provide captions of the required accuracy then this would help overcome the shortage of stenographers; An application such as Synote that encouraged universities to caption all audio and video recordings in order to enhance the learning and teaching benefits of multimedia recordings provides great benefit to deaf and disabled students and non native speakers. Currently available commercial SR systems (e.g. Microsoft, Nuance) do not however make a synchronised transcript output available to the user as the text and speech are only temporarily available in a proprietary synchronized format for correction purposes.

References

1. Whittaker, S., Hyland, P., Wiley, M.: Filochat handwritten notes provide access to recorded conversations. In: Proceedings of CHI 1994, pp. 271–277 (1994)
2. Wilcox, L., Schilit, B., Sawhney, N.: Dynomite: A Dynamically Organized Ink and Audio Notebook. In: Proc. of CHI 1997, pp. 186–193 (1997)

3. Chiu, P., Kapuskar, A., Reitmeief, S., Wilcox, L.: NoteLook: taking notes in meetings with digital video and ink. In: Proceedings of the seventh ACM international conference on Multimedia (Part 1), pp. 149–158 (1999)
4. Brotherton, J.A., Abowd, G.D.: Lessons Learned From eClass: Assessing Automated Capture and Access in the Classroom. ACM Transactions on Computer-Human Interaction 11(2) (2004)
5. Bain, K., Hines, J., Lingras, P.: Using Speech Recognition and Intelligent Search Tools to Enhance Information Accessibility. In: Proceedings of HCI International 2007. LNCS, Springer, Heidelberg (2007)
6. Wald, M.: An exploration of the potential of Automatic Speech Recognition to assist and enable receptive communication in higher education. ALT-J, Research in Learning Technology 14(1), 9–20 (2006)
7. Nuance, http://www.nuance.co.uk/
8. 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 (Retrieved February 7, 2007)
9. Tegrity, http://www.tegrity.com/
10. ViaScribe, http://www.liberatedlearning.com/technology/index.shtml
11. Bain, K., Basson, S., Wald, M.: Speech recognition in university classrooms. In: Proceedings of the Fifth International ACM SIGCAPH Conference on Assistive Technologies, pp. 192–196. ACM Press, New York (2002)
12. Kheir, R., Way, T.: Inclusion of deaf students in computer science classes using real-time speech transcription. In: Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education (ITiCSE 2007). ACM SIGCSE Bulletin, vol. 39(3), pp. 261–265 (2007)
13. Bennett, S., Hewitt, J., Mellor, B., Lyon, C.: Critical Success Factors for Automatic Speech Recognition in the Classroom. In: Proceedings of HCI International (2007)
14. Wald, M.: Creating Accessible Educational Multimedia through Editing Automatic Speech Recognition Captioning in Real Time. International Journal of Interactive Technology and Smart Education: Smarter Use of Technology in Education 3(2), 131–142 (2006)
15. Attila, http://www.liberatedlearning.com/news/AGMSymposium2009.html
16. Fiscus, J., Radde, N., Garofolo, J., Le, A., Ajot, J., Laprun, C.: The Rich Transcription 2005 Spring Meeting Recognition Evaluation. National Institute Of Standards and Technology (2005)
17. Automatic Sync Technologies, http://www.automaticsync.com/caption/
18. Synote, http://www.synote.org
19. JISC, http://www.jisc.ac.uk/fundingopportunities/funding_calls/2007/07/circular0207.aspx
20. Net4Voice, http://www.net4voice.eu
21. Wald, M.: A Research Agenda for Transforming Pedagogy and Enhancing Inclusive Learning through Synchronised Multimedia Captioned Using Speech Recognition. In: Proceedings of ED-MEDIA 2007: World Conference on Educational Multimedia, Hypermedia & Telecommunications (2007)
22. Twitter Synote, http://twitter.com/synote
23. ECS news, http://www.ecs.soton.ac.uk/about/news/2812