

# Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality

Mike Wald

Learning Technologies Group, School of Electronics and Computer Science  
University of Southampton, SO171BJ, United Kingdom M.Wald@soton.ac.uk

*Abstract - Legislation requires that educational materials produced by staff should be accessible to disabled students. Speech materials therefore require captioning and the Liberated Learning Initiative has demonstrated that automatic speech recognition provides the potential to make teaching accessible to all and to assist learners to manage and search online digital multimedia resources. This could improve the quality of education as the automatic provision of accessible synchronised lecture notes enables students to concentrate on learning and enables teachers to monitor and review what they said and reflect on it to improve their teaching. Standard automatic speech recognition software lacks certain features that are required to make this vision a reality. The only automatic speech recognition tool that is being developed to specifically overcome these identified problems would appear to be IBM ViaScribe and investigation of its application in educational environments is occurring through the Liberating Learning Consortium. This paper will describe both achievements and current developments.*

*Index Terms* – accessible multimedia, automatic speech recognition, synchronized speech and text, real time transcription

## INTRODUCTION

Disability legislation reinforces the moral and economic arguments for Universal Usability and Universal Design [1], ensuring websites, software and products are accessible to as many people as possible. This paper explains how the automatic transcription of speech into synchronized text can enhance education for all students and describes achievements and planned developments to turn this vision into reality.

### AWARENESS OF ACCESSIBILITY AND UNIVERSAL DESIGN

In 2002 a letter was sent to President Clinton signed by the presidents of 25 research universities pledging to make universal design and accessibility part of the education provided to computer scientists and engineers at all levels. [2] It is important that all students understand the issue of Universal Design and accessibility as this knowledge could positively influence decisions they make in their future careers. If they will go on to design or create, systems,

programs, web sites, products etc. they will, of course, also require relevant practical knowledge and skills. It is valuable for students not only to be taught about Universal Design and accessibility in theory but also to see it used in practice in the learning and teaching approaches and environment they experience during their studies. Teachers should ideally therefore ‘practice what they preach’ and ‘walk the walk’ as well as ‘talk the talk’ by ensuring their teaching is designed to be as universally accessible as possible. Demonstrating universal design in action can be of great benefit in raising awareness of the issues and making explicit how implementing solutions designed to benefit students with disabilities can benefit all students and how good design can mean that a product will stand more chance of working well for the greatest number of people in the greatest range of situations and with the greatest range of other technologies.

### ACCESS TO SPEECH IN CLASSES

Deaf and hard of hearing students can find it difficult to follow speech through hearing alone or to take notes while they are lip-reading or watching a sign-language interpreter [3]. Although summarised notetaking and sign language interpreting is possible, notetakers can only record a small fraction of what is being said while qualified sign language interpreters with a good understanding of the relevant higher education subject content are in very scarce supply. Although UK Government funding is available to deaf and hard of hearing students in higher education for interpreting or notetaking services, trained stenographers providing a real time verbatim transcript are not normally used because of cost and availability issues.

Students whose first language is not English may also find it difficult to understand the speech of teachers who speak too fast, too indistinctly, have a dialect, accent or do not have English as their first language. Students who find it difficult to write or to spell also can find it difficult to take notes in class. In many situations students with disabilities depend on the goodwill of their fellow students for support (e.g. students who are unable to attend the lecture for mental or physical health reasons).

In many classes all students spend much of their time and mental effort trying to take notes. This is a very difficult skill to master for any student, especially if the material is new and they are unsure of the key points, as it is difficult to

simultaneously listen to what the lecturer is saying, read what is on the screen, think carefully about it and write concise and useful notes.

The automatic provision of a live verbatim displayed transcript of what the teacher is saying, archived as accessible lecture notes would therefore enable staff and students to concentrate on learning and teaching issues (e.g. students could be asked searching questions in the knowledge that they had the time to think). Lecturers would have the flexibility to stray from a pre-prepared 'script', safe in the knowledge that their spontaneous communications will be 'captured' permanently. Lecturers' fears that a transcript of their spontaneous utterances will not look as good as carefully prepared and crafted written notes can be met with the response that students at present can tape a lecture and get it transcribed, and it would be difficult to reasonably refuse this request from a disabled student. An understanding that a verbatim transcript of spontaneous speech is different to carefully crafted written notes is something that would soon be developed and accepted.

### ACCESS TO ONLINE MULTIMEDIA

Legislation requires speech materials to be accessible to disabled learners and tools that synchronise pre-prepared text and corresponding audio files, either for the production of electronic books (e.g. Dolphin [4]) based on the DAISY specifications [5] or for the captioning of multimedia (e.g. MAGpie [6]) are not normally suitable or cost effective for use by teachers for the 'everyday' production of learning materials. This is because they depend on either a teacher reading a prepared script aloud, which can make a presentation less natural sounding, and therefore less effective, or on obtaining a written transcript of the lecture, which is expensive and time consuming to produce. As speech becomes a more common component of online learning materials, the need for synchronised speech and text will therefore increase. Although multimedia materials (e.g. speech, video, PowerPoint files) have become technically easier to create and offer many benefits for learning and teaching, they can be difficult to access, manage, and exploit. If speech could be automatically synchronised with transcribed text files this would provide a cost effective method of assisting learners and teachers to manipulate, index, bookmark, manage and search for online digital multimedia resources that include speech, by means of the synchronised text. Capturing all their presentations in synchronised and transcribed form would also allow teachers to monitor and review what they said and reflect on it to improve their teaching and the quality of their spoken communication.

### ACCESS FOR BLIND AND VISUALLY IMPAIRED OR DYSLEXIC STUDENTS

Although speech synthesis can provide access to some text based materials for blind, visually impaired or dyslexic students, it can be difficult and unpleasant to listen to for long periods and cannot match synchronised real recorded speech in conveying 'pedagogical presence', attitudes, interest,

emotion and tone. It is also difficult to automatically synthesise spoken descriptions of pictures, mathematical equations, tables, diagrams etc. in a way that can be easily understood by blind students and so recorded speech, synchronised with a text description can be of benefit for these situations.

### ACCESS TO PREFERRED MODALITY OF COMMUNICATION

Teachers may have preferred teaching styles involving the spoken or written word that may differ from learners' preferred learning styles (e.g. teacher prefers spoken communication, student prefers reading). Speech, text, and images have communication qualities and strengths that may be appropriate for different content, tasks, learning styles and preferences. Speech can express feelings that are difficult to convey through text (e.g. presence, attitudes, interest, emotion and tone) and that cannot be reproduced through speech synthesis. Images can communicate information permanently and holistically and simplify complex information and portray moods and relationships. Students can usually read much faster than a teacher speaks and so find it possible to switch between listening and reading. If you are distracted or lose focus it is easy to miss or forget what has been said whereas text reduces the memory demands of spoken language by providing a lasting written record that can be reread. Synchronised multimedia enables all the communication qualities and strengths of speech, text, and images to be available as appropriate for different content, tasks, learning styles and preferences. Some students, for example, may find the more colloquial style of verbatim transcribed text from spontaneous speech easier to follow than an academic written style.

### DEVELOPMENT OF A TOOL TO AUTOMATICALLY TRANSCRIBE AND SYNCHRONISE SPEECH

#### *Feasibility Trials*

Feasibility trials of using automatic speech recognition (ASR) software to provide a real time verbatim displayed transcript in lectures for deaf students in 1998 by Dr Wald in the UK [7] and St Mary's University, Nova Scotia in Canada identified that existing commercially available software was unsuitable for transcribing the speech of the lecturer and displaying it in the class for the deaf students. The main problem was that, as the dictation of punctuation does not occur in normal spontaneous speech in lectures, even if the system recognised the speech accurately, the transcribed text appeared as a continuous stream of words and was very difficult to read and understand. The trials however showed that reasonable accuracy could be achieved by interested and committed lecturers who spoke very clearly and carefully after extensively training the system to their voice by reading the training scripts and teaching the system any new vocabulary that was not already in the dictionary.

### *Liberated Learning Project*

Based on these feasibility trials the international collaborating Liberated Learning Project was established by SMU in 1999 funded by a Canadian charity and since then Dr Wald has been working with IBM and Liberated Learning (coordinated by Saint Mary's University, Nova Scotia, Canada) to demonstrate that ASR can make speech accessible to all. [8] Lecturers wear wireless microphones enabling them to move around as they speak and the transcribed text is edited for errors and available for students on the Internet.

To make the Liberated Learning vision a reality, a prototype application, Lecturer, was developed in 2000 in collaboration with IBM to provide a readable display by automatically breaking up the continuous transcribed stream of text based on the silence of pauses in the speech. Text formatting was adjustably triggered by the pause length with short and long pause timing and markers corresponding, for example, to the written phrase and sentence markers 'comma' and 'period' or the sentence and paragraph markers 'period' and 'newline'. Spontaneous speech does not however have the same structure as carefully constructed written text (e.g. people don't speak in complete sentences) and so the transcribed text does not lend itself easily to being punctuated in the same way as normal written text. A more readable approach was achieved by providing a visual indication of pauses showing how the speaker grouped words together (e.g. one new line for a short pause and two for a long pause: it is however possible to select any symbols as pause markers)

Successful trials with Lecturer demonstrated the feasibility of the Liberated Learning concept and Lecturer was replaced in 2001 by IBM ViaScribe [9]. Both applications used the Via Voice 'engine' and its corresponding training of voice and language models and both automatically provided text displayed in a window and stored, synchronised with the speech, for later reference.

### *File Formats*

As with commercially available ASR software, Lecturer used its own proprietary format for synchronising the speech with the text to allow the person editing the text to replay the speech to make any corrections. When the text was edited, speech and synchronisation could be lost (as also occurs with commercially available ASR software). ViaScribe, however was designed to use a standard file format enabling synchronised audio and the corresponding text transcript and slides to be viewed on an Internet browser or through media players that support the SMIL 2.0 standard for accessible multimedia. Editing could now be undertaken without loss of speech or synchronisation.

### *Accuracy Rates*

Analysis of accuracy scores using the NTID Test of Accuracy and Readability [10] gave a mean accuracy of 77% with a standard deviation of 9.58% for 17 lecturers at 8 institutions [11]. These lecturers varied in their lecturing experience, abilities, familiarity with the lecture material and the amount of time they spent on improving the voice and language

models and so it would appear to be reasonable with experience and training to aim for accuracies of 85% and above.

Re-voiced ASR is sometimes used for live television subtitling in the UK [12] and in classrooms in the US [13] but the Liberated Learning project aimed to research how students with disabilities could remain independent of the support of an intermediary and so did not use somebody to 're-voice' the speech of lecturers who had not undertaken the ASR training process, or to punctuate the text by speaking the punctuation.

It was observed that lecturers' accuracy rates were lower for spontaneous speech than scripted speech and informal investigation suggested that this might be because:

- vocabulary may be introduced that is not in the dictionary;
- the speaker is not as fluent and may hesitate or stumble over words;
- the rate of delivery varied more in a spontaneous than scripted situation;
- ASR is not optimised for recognition of a specific speaker's spontaneous speech as it is based on generic language models created from written documents.

It was also observed that lecturers' ASR accuracy rates were lower in classes compared to those achieved in the office environment. This has also been noted elsewhere [14]. Informal investigations have suggested this might be because the rate of delivery varied more in a live classroom situation than in the office resulting in indistinct word boundaries (i.e. the end of words being run into the start of subsequent words.)

Teaching subjects that use very specialist vocabulary and structures (e.g. computer programming or mathematics) initially gave lower accuracy rates, although these did improve with extensive training of the voice and language models. Mathematical expressions did not automatically appear as equations.

### *Student and Teacher Feedback*

Detailed feedback from 44 students with a wide range of disabilities [11] and interviews with lecturers showed that both students and teachers generally liked the Liberated Learning concept and felt it improved teaching and learning as long as the text was reasonably accurate (i.e. >85%). The majority of students used the text as an additional resource to verify and clarify what they heard and many students developed strategies to cope with errors in the text.

## FUTURE DEVELOPMENTS

Liberated Learning research and development has also included improving the recognition, training users, simplifying the interface, and improving the display readability. Some of these developments have been trialled in the laboratory and some in the classroom. Current research and development activities include:

- new trials in UK, China, Japan in addition to continuing trials in US, Canada and Australia

- a new integrated speech recognition engine (Lecturer and Viascribe required the ViaVoice ASR engine);
- removing the need for training by using language models based on individual's recorded and transcribed spontaneous speech rather than on generic written documents (This should also help ensure better accuracy for a speaker's specialist subject vocabularies and structures);
- personalised individual wireless displays to enable students to read the transcribed text in the format they prefer and to highlight, annotate and save sections when appropriate.
- 'real time editing' enabling one or more 'editors' to correct recognition errors as they occur.

Although it can be expected that developments in ASR will continue to improve accuracy rates, the use of a human intermediary to improve accuracy through re-voicing and/or correcting mistakes in real time as they are made by the ASR software could sometimes help compensate for some of ASR's current limitations.

### CONCLUSION

Synchronised ASR multimedia enables academic staff to take a proactive rather than a reactive approach to teaching students with disabilities by providing a cost effective way of making their teaching accessible. This can improve the quality of education for all students by assisting learners to manage and search online digital multimedia resources and enable students to concentrate on learning rather than notetaking and enable teachers to improve their teaching through reflection. It can also be valuable in demonstrating practically to all students the principles and benefits of accessibility and Universal Design. The potential of using ASR to provide automatic transcription of speech in higher education classrooms has been demonstrated by the Liberated Learning Initiative and the only ASR tool that can provide a real-time transcription display, synchronisation and editing would appear to be IBM ViaScribe.

### REFERENCES

- [1] Shneiderman, B. "Universal Usability" *Communications Of The ACM* May 2000, Vol.43, No.5 <http://www.cs.umd.edu/~ben/p84-shneiderman-May2000CACMf.pdf> last accessed 2005-03-04
- [2] Olsen, F. "25 Universities Pledge to Increase Research in Computing for the Disabled" *The Chronicle of Higher Education* September 22, 2000 <http://chronicle.com/free/2000/09/2000092202t.htm> last accessed 2005-03-04
- [3] Wald, M. "Hearing disability and technology", *Access All Areas: disability, technology and learning*, JISC TechDis and ALT, 2002, pp. 19-23.
- [4] <http://www.dolphinlink.co.uk/audio/products/EasePublisher/index.htm> last accessed 2005-03-04
- [5] <http://www.daisy.org> last accessed 2005-03-04
- [6] <http://ncam.wgbh.org/webaccess/magpie/> last accessed 2005-03-04
- [7] Wald, M. "Developments in technology to increase access to education for deaf and hard of hearing students". In: *Proceedings of CSUN Conference Technology and Persons with Disabilities*. California State University Northridge, 1999.
- [8] Bain, K. Basson, S. Wald, M. "Speech recognition in university classrooms". In: *Proceedings of the Fifth International ACM SIGCAPH Conference on Assistive Technologies*. ACM Press, 2002, pp. 192-196.
- [9] [http://domino.research.ibm.com/comm/wwwr\\_seminar.nsf/pages/sem\\_abstract\\_292.html](http://domino.research.ibm.com/comm/wwwr_seminar.nsf/pages/sem_abstract_292.html) Last Accessed 2005-01-31
- [10] Leitch, D. MacMillan, T. "Improving Access for Persons with Disabilities in Higher Education Using Speech Recognition Technology". *Liberated Learning Project Year II Progress Report*. 2001 Saint Mary's University, Nova Scotia.
- [11] Leitch, D. MacMillan, T. "Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research". *Year III Report*. 2003 Saint Mary's University, Nova Scotia.
- [12] Lambourne, A. Hewitt, J. Lyon, C. Warren, S. "Speech-Based Real-Time Subtitling Services". *International Journal of Speech Technology*, 7, 2004. pp.269-279, Kluwer Academic Publishers
- [13] Francis, P.M. Stinson, M. "The C-Print Speech-to-Text System for Communication Access and Learning". In: *Proceedings of CSUN Conference Technology and Persons with Disabilities*. California State University Northridge, 2003.
- [14] Bennett, S. Hewitt, J. Kraithman, D. Britton, C. "Making Chalk and Talk Accessible" *ACM SIGCAPH Computers and the Physically Handicapped*, Issue 73-74, 2002, pp.119-125.