Speech Recognition in University Classrooms: Liberated Learning Project

Keith Bain
International Project Manager
Saint Mary's University
Halifax, NS B3H 3C3 CANADA
902-496-8741
keith.bain@stmarys.ca

Sara H. Basson, Ph.D
Manager, Business and Market Development,
IBM - T.J. Watson Research Center
Box 218, Yorktown Heights, N.Y. USA

Mike Wald, Ph.D
Director Southern Higher Education Consortium
University of Southampton, New College
The Avenue, Southampton, SO171BJ UK

ABSTRACT
The LIBERATED LEARNING PROJECT (LLP) is an applied research project studying two core questions:

1) Can speech recognition (SR) technology successfully digitize lectures to display spoken words as text in university classrooms?
2) Can speech recognition technology be used successfully as an alternative to traditional classroom notetaking for persons with disabilities?

This paper addresses these intriguing questions and explores the underlying complex relationship between speech recognition technology, university educational environments, and disability issues.

KEYWORDS
Speech Recognition, Accessibility, Higher Education

INTRODUCTION
The Atlantic Center of Research, Access and Support for Students with Disabilities at Saint Mary's University has been responding to the needs of students with disabilities for nearly two decades. A major thrust of the response is understanding the role technology plays in mediating the integration of persons with disabilities into higher education. Since its inception, the Atlantic Centre has advocated and advanced the use of technology to level the playing field for our students. For the past decade the Centre has carefully and critically watched the development of speech recognition technology, believing that one day it may revolutionize the learning experience for students with disabilities. The introduction of true continuous speech recognition products with large, expandable vocabularies engendered a commitment from Saint Mary's to explore the concept further. Thus, a world first initiative- the **LIBERATED LEARNING** Pilot Project (1998) was born.

Assets 2002, July 8-10, 2002 Edinburgh, Scotland

In the fall of 1998, after intensive voice training on computers, three Saint Mary's professors utilized speech recognition software in their courses. Their spoken lectures were digitized and simultaneously translated into text via speech recognition software, then displayed on a large screen at the front of the classroom. Students could not only hear the lecture, but also read the lecture as it was delivered. More importantly, they could also obtain a nearly verbatim transcript in either hard or disk copy for study purposes.

The initial testing of this application for speech recognition was enlightening. This brief exposure to the concept suggested it could indeed provide an alternative to conventional note taking for students with disabilities. Serendipitously, it was noticed that non disabled students used the instantaneous display of the lecture as a reference check for their own notes- the concept gave students access to both auditory and visual learning channels, helping them better integrate the lecture content. They could use the software-generated notes to augment their own notes. Therefore the successful application of speech recognition technology was seen to have valuable implications for every student in the classroom.

Emanating from these humble beginnings and under the leadership of Dr. David Leitch, Saint Mary's forged an international consortium to further refine and research the Liberated Learning concept. Joining Saint Mary's University are IBM Research, Stanford University, California, Ryerson University, Ontario, University of the Sunshine Coast, Australia, Aliant Telecommunications, Alexander Graham Bell Institute at the University College Cape Breton, Nova Scotia, and Durham College, Ontario.

These partners plus associates from the University of Southampton, UK, and the University of Texas at Austin continue to collaboratively develop the technology, understand its impact, and drive future applications.

Liberated Learning Concept

In Liberated Learning courses, specially designed Speech Recognition technology is used to provide greater access to lecture content

- Lecturers engage in a comprehensive training and implementation process to develop a personalized, classroom ready voice profile by "teaching" speech recognition software to understand individual speaking style.
- Lecturer uses a wireless microphone 'connected' to a robust computer system during lectures.
- Specially designed speech recognition software, working in conjunction with IBM's ViaVoice technology, receives digitized transmission of lecturer's speech.
- Using lecturer's voice profile and acoustic information, the software converts spoken lecture into electronic text.
- Text is displayed via projector for class in real time: students can simultaneously see and hear the lecture as it is delivered.
- After the lecture, text is edited for recognition errors and made available as lecture notes for all students through an on-line note system.
- Lecturer's individual voice profile is continuously updated and expanded through intensive system training.

The main objective of the Liberated Learning Project is to test applications of speech recognition in actual university classrooms, develop and evaluate a model for using speech recognition in the university environment, and report on the impact of this technological intervention on students with disabilities, faculty, and non disabled students. Furthermore, the project intends to focus global attention on the concept as a method of improving access to learning for people with disabilities. During this three-year applied research project, researchers will thoroughly develop and test multiple applications of speech recognition as a tool to enhance teaching and learning.

DEMOGRAPHIC RATIONAL

To illustrate the potential impact of this teaching and learning tool on students with disabilities, a demographic study of students with a disability, undertaken in Canada by Dr. Leitch in 1998, revealed approximately 7,000 students with a disability were attending the 47 universities surveyed by Canada's McLean's Magazine. In Australia, according to the 1999 statistics produced by the Department of Education, Training and Youth Affairs (DETYA), there were 18,084 students with a disability enrolled at the 39 public universities. Therefore, the immediate implications for speech recognition technology in tertiary education in Australia and elsewhere will be great.

In the US, the sheer number of potential stakeholders exacerbates the need for creative innovations in accessibility. Many ADA analysts believe that this federal law covers more than 50 million people. Various summaries, including those issued by the National Institute on Disability and Rehabilitation Research, indicate that between 15% and 20% of any grouping of randomly selected people can be expected to have those impairments considered as disabilities under federal/state law (source: Louis Harris & Associates, 1994).

Out of the 677,100 higher education students who entered their First Year in 1999/00 in 172 institutions in the UK, 26,720 were known to have a disability.

These demographics do not consider countless individuals who have not self-identified nor have been formally diagnosed with a specific disability. These numbers are likely significant and could be advantaged by multi-modal access to real time information and augmentative notes.

For students with disabilities, it is clear that problems exist with both immediate intake of the lecture material and with notetaking for later study purposes. For example, students who are deaf or hard-of-hearing usually require interpreters or assistive listening devices, and rely upon notetakers. As well, students with certain learning disabilities find it difficult to process information presented orally, and other students are physically unable to take their own notes. International students and English Second Language learners struggle with lecture content delivered in auditory format, typically having greater exposure to English language in print form. Finally, the notetaking skills of non-disabled students are often far from satisfactory.

The Liberated Learning Project is grounded in a paradigm that promotes independence for students with disabilities, unlike conventional approaches to notetaking that have historically sustained a dependence on intermediaries. Furthermore, it is synergistic with universal design principles in that it potentially addresses macro level learning issues for a variety of stakeholders with varying needs.

CHALLENGES

The Liberated Learning Project involves an intricate interaction of technological and human resources. As with any technological application in its infancy, there are obstacles to overcome before the Liberated Learning concept is more readily applicable. Three key challenges captured much of the project's research and development attention:

1) Accuracy of digitized lecture
2) Production of SR generated notes
3) Real time readability of displayed text

Accuracy

SR transcribed word accuracy is arguably the projects' most important critical success factor, whether for display in the classroom, used as lecture notes, or both.

In connection with SR, most references to the measurement of accuracy leave the basis for its determination undefined or stated simply as the percentage of spoken words correctly transcribed into text. This has merit as a general definition for assessing ASR applications such as dictation. However for a number of reasons it is unsatisfactory for assessing accuracy emanating from spontaneous speech, for example, an unread, non-memorized lecture.

It is common to read or to be told by a user of speech recognition, that 98% accuracy is readily achievable. And indeed this is so, under favorable conditions such as dictating or reading selected materials aloud. However, in introducing speech recognition into the classroom, and asking it to recognize a lecturer's spoken lecture, we are asking both the technology and the instructor to undertake a much more challenging application.

For example, most lectures are characterized by extemporaneously generated speech. The dynamism present in this environment inevitably generates false starts, disfluencies, hesitations, ungrammatical constructs, etc. These facets of natural language delivery lead to reduced accuracy. Human factors aside, the interaction of the hardware infrastructure and the inherent design of speech recognition engines limits the effectiveness of introducing high level technical features to the baseline setup. Current speech recognition engines are primarily designed to leverage current commercial grade robustness. They cannot necessarily take advantage of professional grade soundboards, for example. Therefore, efforts to integrate cutting edge associated technologies are somewhat hamstrung by intrinsic speech recognition algorithms.

Dr Ross Stuckless, project consultant and professor emeritus at the National Technical Institute for the Deaf in the United States, developed an instrument for a detailed scoring procedure for inter-scorer readability (Word Accuracy sub-test of the Test of Automated Speech Recognition Readability). Dr Stuckless's instrument is designed to test three components of text readability, i.e. word accuracy, sentence markers and speaker changes.

Using this metric, the Liberated Learning project already surpassed its stated benchmark accuracy rate of 90% in a university lecture. However, these accuracy assessments introduced a new complexity in terms of comprehension issues. Certainly, most project researchers agree that certain inherent errors likely impact comprehension more significantly than less significant errors. Consider the following simple example:

Actual words spoken:	I went to the store
Scenario #1, SR transcribed:	I went to the *door*
Scenario #2 SR transcribed:	I went *too* the store

Both scenarios return an accuracy rate of **80%** (4/5 words recognized correctly; word transcription error italicized). However, in the absence of audio cues to aid understanding, for example as experienced by a person with a hearing disability, the difference between the two transcriptions directly affects comprehensibility of the phrase. Quantifying text comprehension will be prominent in subsequent research endeavors.

Producing SR Generated Class Notes

The ease of creating comprehensive class notes from a SR transcription of the spoken lecture is directly proportional to the digitized accuracy. In analysis of many one-hour lectures, speaking rates have varied between one hundred and two hundred words per minute. To illustrate the scope of the editing dilemma, a mean speaking rate of one hundred fifty words per minute translates into 9000 words spoken per lecture. An 80% accurate transcript thus yields 720 recognition errors. If each error takes even only a few seconds to edit, the resulting effort is considerable. Early indications show that the editing process for creating a perfectly accurate, verbatim transcript is roughly a 3:1 ratio of audio data to correction time (1 hour lecture = 3 hours editing). Therefore, achieving high accuracy rates is imperative to ensure timely production of SR class notes.

Editing skills certainly improve with practice and the development of new techniques. Project working groups developed a number of macros applicable in traditional word processors to aid editing tasks. However, even with some efficiency improvement, this process is not sufficient for vast users to adopt the system.

New approaches are being researched.. Faculty are working to identify targeted approaches to editing based on key elements of the lecture. Editors are learning to ignore seemingly insignificant errors, such as homonyms, inadvertently pluralized words, etc. This requires some individual interpretation and discretion about what truly constitutes an important error - one that needs to be corrected to ensure proper comprehensibility.

From a technical perspective, scientists are developing text summarization techniques, which theoretically would reduce the scope of editing requirements. Other approaches include offloading the editing process to students. Allowing students to correct streaming text during the delivery would allow a more perfected transcript to be available immediately at the end of the lecture. Numerous learning, technical, and logistical considerations need to be explored in greater detail in order to implement this solution. However, such creative solutions seem achievable in the very near future.

Readability

Ensuring high accuracy is in and of itself insufficient for ensuring the Liberated Learning concept is an effective learning tool. In the 1998 pilot phase, SR digitized text contained no sentence markers to distinguish independent thoughts. In other words, text flowed together in a continuous stream of words, which quickly enveloped the screen. In the first year of the project, a development team created a new "classroom" speech recognition application that works in conjunction with IBM's ViaVoice technology. Gathering feedback from both faculty and students alike, an iterative software development process has engaged programmers to provide a functional interface. The primary design goal is to create an application capable of delivering readable, accurately displayed text for student use in a lecture dynamic. The first classroom speech recognition application, Lecturer, was successfully tested in 2000.

This ongoing development process continues to evolve as performance data is generated and incorporated in new design schemas. Initially, rapid development using TCL scripting language produces numerous revisions with impressive proof of concept functionality. However, the TCL environment presented a limiting platform for robustness, scalability, and new functionality. As such, a next generation application was needed.

RESEARCH AND DEVELOPMENT

The LLP takes the speech from professors or lecturers in the classroom and transforming that into text, which is displayed on a screen and then stored in an electronic form. How that is done and the technology by which that occurs doesn't really concern professors or students.

The first approach the LLP adopted was to use a high end workstation (IBM Intellistation) and that worked well but it was not very portable, and so the next approach that was adopted was to use a laptop computer.

An approach that the project is now looking at is to use a network system and process the speech remotely from the classroom. The speech from the lecturer will go over the network and be processed somewhere else than the classroom and the text is then returned back into the classroom and displayed on the screen.

So why take this approach what are the advantages?

Firstly, there is no need for every single professor in every single classroom to actually own the latest computer system. They also do not have to carry it in with them as the up-to-date high performance processing system with the latest recognition engines can be stored somewhere else and accessed over the network. Another benefit is that there is no need for every professor to be a technical wizard as the technical wizards can be somewhere else on the network sorting out any problems.

The next benefit is that there is no need for professors to worry about whether their text and speech data has been saved or have students worry that all this valuable information has disappeared because again, someone somewhere else on the network can worry about that.

With a network system it is possible for any student to have his or her own display. It can be wireless and customized and personalized to how they want text to appear including the font size, the colour & how it scrolls.

It is also possible to have real time editing and correction, which means that students can walk away with the correct version of the text and do not have to wait until some time later to get access to this.

Since the system is working over a network it is possible to have more than one speech recognition engine running at the same time. Firstly this might allow the use of both speaker dependent recognition, where the speaker has enrolled and trained the system how to recognize their own voice, and speaker independent recognition, where the system will be able to recognize anyone in the room. This would mean, that in interactive group sessions, contributions, questions and comments from the room would be transcribed directly into text and would not have to be repeated by the professor.

Speech-recognition systems work by calculating how confident they are that a word that has been spoken, has been recognized in their dictionary. It is possible with more than one speech recognition engine running on the network to compare the scores to find the best recognition.

If a dictionary based system decides that it is unlikely that the word that has been spoken is in its dictionary it is still forced to throw up a word even though the system knows that probably isn't the correct word. If it were possible to present a Phonetic display of the spoken word it would give the person a clue as to what the word might have been.

It would also be possible to share the language models, the vocabulary, and the content between professors, so if more than one professor was teaching a course or a subject area they could use another professor's language models and training of the system.

IBM NETSCRIBE

In the spring of 2001, Saint Mary's University and IBM Research signed a Joint Study agreement to pursue these opportunities. Combining Liberated Learning discoveries with IBM's work from a similar project in France called Lipcom, programmers at IBM developed a prototype called **Netscribe.** Netscribe opened the door to a number of exciting opportunities for new classroom applications. Over the summer, The Liberated Learning project team hosted a number of IBM's top speech recognition scientists and demonstrated network-based operation, a client/server environment where text could be displayed on multiple, individually customizable outputs, speaker independent software, and investigated streaming speech recognition text to the Internet. The project research and development teams are currently investigating testing and conversion strategies for these and other exciting applications. Actual classroom tests of network based speech recognition will occur in the 2002 academic semester at numerous Liberated Learning campuses.

CONCLUSION

The Liberated Learning concept may potentially revolutionize educational access for persons with disabilities. The Liberated Learning Project has already resulted in dramatic increases in the knowledge and experience base with respect to potential educational applications for speech recognition. The success of the efforts of the Liberated Learning Project team will encourage the continued support from the corporate sector as well as help in expanding the consortium of universities engaged in the Project. Members of the team are confident that the Liberated Learning Project will receive widespread acceptance as a model for universities to better accommodate students in the classroom.

Thus, we believe the project's mission of enabling universal equal access to information will be realized through the use of a new technology, through the ongoing support of our many partners, and through pioneering research and development.

REFERENCES

1. Bain, K. Paez, D. Speech Recognition in Lecture Theatres. *Proceedings of the Eighth Australian International Conference on Speech Science and Technology.* Canberra, Australia (2000)
2. Leitch, D. Canadian Universities: The Status of Persons with Disabilities. Saint Mary's University, Nova Scotia (1998)
3. Leitch, D. MacMillan, T. Liberated Learning Project: Improving Access for Persons with Disabilities in Higher Education Using Speech Recognition Technology; *Year II Report.* Saint Mary's University, Nova Scotia (2001)
4. Stuckless, R. Assessing the word accuracy of text produced form an instructor's use of ASR in the college classroom (2000)