SPEECH RECOGNITION IN LECTURE THEATRES -

LIBERATED LEARNING PROJECT

AN INNOVATION TO IMPROVE ACCESS TO HIGHER EDUCATION

USING SPEECH RECOGNITION TECHNOLOGY

Keith Bain[†], Di Paez[*]
†Liberated Learning Project, St Mary's University, Nova Scotia
Canada
*University of the Sunshine Coast, Queensland, Australia

ABSTRACT - The Liberated Learning Project (LLP) represents a large and complex effort to determine whether automatic speech recognition (ASR) can be successfully used for real-time transcription and display in university lecture theatres. ASR-based notes, which are produced to enhance retention of lecture material by students, will overlap data collected to investigate the use of ASR in the lecture theatre. Stuckless (2000, cited in Konopasky) described the need to collect "data for assessment of accuracy of speech-to-text) for examination by our IBM partner". Leitch (2000, cited in Konopasky) described the need to record outputs like audio-video tapes of lecture and display text, audio and displayed text, audio and software generated notes, unedited, and audio-software generated notes edited.

INTRODUCTION

The **LIBERATED LEARNING PROJECT** (LLP) is a large and complex research project studying two core questions:

1. Can a special application of speech recognition technology successfully digitize lectures to display spoken words as text in university classrooms?

2. Can speech recognition technology can be used successfully as an alternative to traditional classroom notetaking for persons with disabilities?

This paper is intended to inform discussion on the nature and extent of research required to develop and implement the Liberated Learning concept.

CONCEPT AT A GLANCE:

Specially designed Speech Recognition technology used by lecturers in university lectures to improve access for students with disabilities.

- Lecturer develops a personalized voice profile by "teaching" speech recognition software to understand his/her speaking style.

- Lecturer uses a wireless microphone 'connected' to a robust computer system during lectures. An IBM Intellistation computer running 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 notes system.

- Lecturer's individual voice profile is continuously updated and expanded through intensive system training.

Issues relating to the impact of the technology on both students and staff are examined in detail at the pre-technology and post-technology phases of the project, however they are not within the scope of this paper. This paper confines itself to the specific technological issues associated with the development of the software and hardware configurations and its impact on the project, test sites, and users.

PROJECT OBJECTIVES

The main objective of the Liberated Learning Project is to develop and evaluate a model for using speech recognition in the university lecture theatre. Further, 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 project, researchers will thoroughly develop and test multiple applications of speech recognition as a tool to enhance teaching and learning. An effective model of using speech recognition in the university lecture theatre will be developed and refined. The project culminates with an international conference on the importance of speech recognition in the university lecture theatre and will include a comprehensive program for widespread application.

Saint Mary's University, Halifax, Nova Scotia, received major funding in 1999 from the J.W. McConnell Family Foundation in Canada to further this research and refine a unique application of speech recognition technology to assist students with disabilities in the university lecture theatre. Saint Mary's University through Dr. David Leitch, Director of the Atlantic Centre of Research, Access, and Support for Students with Disabilities, is now heading a consortium of Canadian and international university and corporate partners to develop the Liberated Learning concept. These strategic alliances will collaborate on the development and testing of speech recognition technology in the lecture theatre, and study its implications for pedagogy and learning.

Universities involved in developing and testing the concept during the project initially include Saint Mary's University and Ryerson Polytechnic University in Canada, and the University of the Sunshine Coast in Queensland Australia. Additionally, the project has a specific mandate to expand the consortium over its three-year life and multiple universities are poised to join as research associates.

Saint Mary's University began testing the software in lecture theatres during October 2000, and will be followed by the University of the Sunshine Coast and Ryerson Polytechnic Universities in February 2001.

THE NEED FOR THE LIBERATED LEARNING PROJECT

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 in 1995, revealed approximately 7,000 students with a disability were attending one of the 47 universities surveyed by 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.

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. 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 which have historically sustained a dependence on intermediaries. So a major thrust of the research is understanding the role ASR technology can play in mediating the integration of persons with disabilities into higher education.

At present, there is no published, replicable information available which applies directly to the accuracy of automatic speech recognition (ASR) under the control of a lecturer in a lecture theatre situation. This comes as no surprise as the literature contains little information of any kind pertaining to the accuracy of ASR involving a single speaker under conditions of spontaneous speech.

SUPPORT FROM THE CORPORATE SECTOR

To enable the development of ASR for lecture theatre use, the Liberated Learning Project has corporate partners, namely individuals from IBM Research in New York and MTT, (formerly Maritime Telephone and Telegraph) in Canada.

The Project has been assisted by the world's top speech recognition scientists at the IBM - T.J Watson Research Centre in Yorktown Heights, New York. Assistance from the Centre includes:

1. Providing access to the world's top speech recognition scientists;

2. Providing access to IBM's top technical/software experts;

3. Providing access to the software codes for ViaVoice to enable modifications for use in lecture theatres;

4. Copies of the latest IBM speech recognition software;

5. Serving in an advisory role to lecturers as they are trained in the use of ViaVoice technology;

6. Providing updates about advances in speech technology for the project team

7. Working with partners in the Project to design experiments evaluating the efficacy of the technology;

8. Serving as a liaison with IBM Special Needs Division to determine whether the results of the Project can be disseminated more widely in other university settings.

MTT, Canada are also partners in the Project as they see unlimited opportunities in the advancement of this technology for the purpose of enhancing the global learning environment. MTT provide the sophisticated telecommunications support structure necessary for the collaboration of individuals and institutions in Canada, the United States, Britain and Australia. The role of MTT is to provide an efficient, reliable and effective system for archiving, posting and accessing software generated lecture notes.

IMMEDIATE 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 universally applicable.

Accuracy:

Word accuracy is a major component in the readability of the text of a lecture, and in turn its value to the student, whether it is displayed in the classroom to be read in real time, used as lecture notes, or both.

In connection with ASR, 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. In fact, many scientists suggest that the greatest accuracy improvements in this phase of the project may come about because of speaker dependent adjustments, such as elocution lessons for faculty, as opposed to strictly technical advances.

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 has developed an instrument for a detailed scoring procedure for inter-scorer readability (Word Accuracy sub-test of the National Technical Institute for the Deaf 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. Dr Stuckless's experience in scoring using these previously documented procedures will undertake a random sample of 2,000 consecutive spoken words from one hour lectures videotape recorded by each of the participating professors to derive a word accuracy score.

To enable accurate scoring, a lecture is videotaped is then dubbed to an audiocassette. A copy of the original, unedited text generated automated speech recognition in tandem with the lecturer's voice is then analyzed and compared to derive the accuracy score.

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, digitized lectures displayed as 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 this project phase, a team of developers have 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.

Secondly, the application must enhance faculty performance and ease of use. 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 functionality. Long term strategies include migration to a richer

programming language, for example, Java, as software development tools improve and core engine access through ViaVoice API (application program interface) is extended.

Hardware Development:

Determining the optimal mix of associative technologies: sound card, operating system, microphone technology, memory, storage, etc remains a particularly challenging aspect. Moreover, improving the technical model's cost efficiencies - an important task when considering more universal application of the Liberated Learning concept, must also be considered. Finally, the system must be easy to use. The current baseline system is likely overly complex from an end user perspective.

Ensuring as many elements of the research, as is possible, are standardised to ensure universal comparison may prove difficult. Different production schemes exist for seemingly identical technical components between countries and even between regions. To illustrate this point, three recent workstations with identical specifications were purchased for testing. During final configuration, researchers noticed a discrepancy in BIOS settings among the "identical" systems, which negatively impacted on available system resources for speech processing. These extremely subtle hardware discrepancies can only be discovered through intricate, multivariance testing.

Data Sharing:

A primary project goal was to provide a robust system for archiving, editing, posting, and accessing speech recognition software generated notes. The Liberated Learning On-line Notes System is a java-based web application designed to facilitate easy and efficient uploading and downloading of raw speech data and class notes generated by Liberated Learning concept. At the heart of the system is a simple and easy to learn HTML interface that provides the full functionality needed by users - students, lecturers, and researchers.

To facilitate this interchange, the project developed a robust SQL Database environment, incorporating Java Servlets and Java Applets system architecture to upload speech recognition data and allow students to download edited notes in various formats. Essentially, the Java technology communicates with the server housed database and presents HTML pages to authorized users. The Java servlets performs user authentication to ensure that the user is only given access to files where they have been specifically granted permission. Collection of data pertaining to system usage and other measures also enable researchers to analyze the use and success of the system. This shared on-line notes system must continue to evolve as the project expands.

One pressing challenge is how researchers separated by huge geographical boundaries battle existing bandwidth infrastructure. Team members must be able to exchange and analyze speech data files exceeding 100Mb, sometimes in real time. Plans to incorporate video analysis further exacerbate this communication dilemma. Thus, challenges associated with scalability and enhanced functionality relative to the extant research methodology must be considered.

CONCLUSION

The Liberated Learning Project will result 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 lecture theatres.

BIBLIOGRAPHY

Coco, D.S. Bagnall, J. (2000) Improving Access for Persons with Disabilities in Higher Education using Automated Speech Recognition Technology, (PEPNet 2000 Conference, Denver, USA.)

Konopasky, R.J. (2000) Liberated Learning Project: ASR-Based Class Notes, An Introduction and Framework for Discussion, (unpublished)

Leitch, D. (2000) Liberated Learning Project, Research Issues, An Introduction and Framework for Discussion (unpublished)

Leitch, D. MacMillan, P. (2000) Proposed Research Guidelines for the Liberated Learning Project, Part One: Conceptualization (unpublished)

Spicer, G. (2000) Functional Description for the Initial Phase of the Liberated Learning Project (unpublished)

Stuckless, R. (2000) Assessing the word accuracy of text produced from an instructor's use of ASR in the college classroom (unpublished)

Stuckless, R. (2000) Benchmarking the accuracy of ASR in the college classroom: A proposal (unpublished)