# Accessible Lectures: Moving Towards Automatic Speech Recognition Models Based on Human Methods

Miltiades Papadopoulos
Accessibility Research Centre

University of Teesside
United Kingdom

M.Papadopoulos@tees.ac.uk

Dr. Elaine Pearson
Accessibility Research Centre

University of Teesside
United Kingdom

E.Pearson@tees.ac.uk

## ABSTRACT

The traditional lecture remains the most common method of teaching and while it is the most convenient from a delivery point of view, it is the least flexible and accessible. This paper responds to the challenge of meeting the needs and access requirements of students with disabilities by urging further adaptations in the learning environment. The aim of this work is to explore the way speech recognition technology can be employed in the University classroom to make lectures more flexible and accessible. The concluding section explores the concept of an ASR model, based on principles derived from studies of human methods of recognition, in order to increase their performance and efficiency.

## Categories and Subject Descriptors

K.3.1: [**Computer and Education**]: Computer Uses in Education – *Computer-assisted instruction (CIA)*.

I.2.7: [**Artificial Intelligence**]: Natural Language Processing – *Language models*.

## General Terms

Human Factors, Performance.

## Keywords

Accessibility, Automatic Speech Recognition (ASR), Human Speech Perception (HSP).

## 1. INTRODUCTION

The evolving landscape of learning technology has placed an emphasis on new educational approaches and pedagogies, with a focus on the adoption of flexible methods of teaching, nonetheless the traditional lecture remains the most common method of delivery [2]. Institutions consider face-to-face delivery as the best way of meeting the expectations of their students and, in addition, most school leaver entrants to university are perceived to have an expectation that teaching would be carried out using a combination of lectures and tutorials [8]. On the other hand, the Special Educational Needs and Disability Act [7]

requires all services to be accessible to students and implies that all staff, academic and support, have a responsibility for providing a learning environment in which disabled students are not disadvantaged. The aim of this study is to explore the way speech recognition technology can be employed in the University classroom to make lectures more flexible and accessible. This work aims to propose an ASR model based on principles derived from studies of human methods of recognition for improved performance.

## 2. MACHINE RECOGNITION IN THE UNIVERSITY CLASSROOM

A number of innovative approaches have been adopted to supplement lectures through real time captioning and audio recordings to make them more flexible. The Liberated Learning Project, have utilised Automatic Speech Recognition technology, in order to improve the flexibility of the traditional lecture and meet the needs of people with disabilities [1]. Additionally, the Villanova University Speech Transcriber (VUST) system was designed to improve the accessibility of computer science lectures using real-time speech recognition software. The study was conducted at the Applied Computing Technology lab at Villanova University and evaluated the impact of the VUST system on the effectiveness of a portable, centralised, laptop-based ASR system designed to augment note-taking, by deaf and hard of hearing students, in the college classroom [4].

The results of these studies suggest that reasonable accuracy rates could only be achieved by committed lecturers after extensive training [4; 5; 9]. ASR systems involve extensive training, need to get used to each speaker's voice and must learn a new vocabulary. In a lecture situation this would be a combination of social and subject specific language. Most of the research in the area of ASR and the way it can be embedded within teaching methodologies focuses on the way it can be deployed in the university environment. Although, in many cases, the results have been unconvincing, research fails to explore new approaches towards more efficient systems.

## 3. TOWARDS HUMAN METHODS

Despite the impressive technological advances and the substantial progress that has been made in the area of automatic speech recognition, the performance of ASR systems is still below the levels required for accurate transcriptions of lectures. Current systems are largely based on statistical approaches,

mainly Hidden Markov Models, and although they have reached a level of maturity, their performance is still much worse than that of humans. This has generated an interest in the ASR community to think about developing innovative techniques based on principles derived from studies of human methods of recognition [6].

In order to gain understanding of the main human perception mechanisms and examine whether human methods can be adapted by automatic speech recognition models, a review of the processes of human speech perception is required. Unfortunately, a complete understanding of all the processes related to speech perception is not viable. In addition, not all human processes can be adopted by computer systems. Even the simplest mechanisms in the brain cannot be easily monitored and applied to speech recognition technology. Therefore, an overview of the most applicable processes of human and machine recognition was produced in an attempt to compare their most relevant steps. Identifying the main differences and similarities of human and automatic processes is a step towards new ASR. The key processes of human speech perception can be divided into the following aspects: Signal Analysis, Speech Units, Speech Segmentation, Speech Variability, and Linguistic Scoring. A preliminary model is proposed according to the findings to date (Figure 1). The model is based on the current statistical ASR models, enriched with relevant human methods.
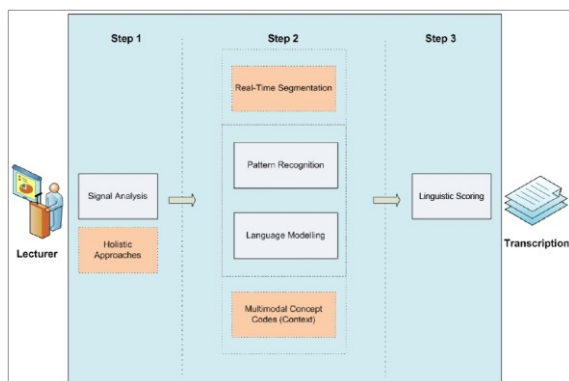


**Figure 1: Initial Work on the Proposed Model**

During the Signal Analysis step, continuous systems analyse acoustic signals as discrete segments and label them according to their acoustic properties. Holistic processes, employing words or syllables as seen on Human Speech Perception, could be employed as an alternative. Subsequently, acoustic models are utilised to calculate transition probabilities between the acoustic segments. A pronunciation lexicon assists the process. During this stage real-time segmentation processes could be included. HSP is also affected by many other multimodal concept codes, which are actively involved in the context of a sentence [3]. Similar processes could also be adopted by ASR models.

## 4. SUMMARY & FUTURE WORK

There is a need to enhance the learning experience for disabled students and those studying in a foreign language. Current systems cannot yet transcribe efficiently in such challenging environments. Additional factors, such as the acoustic quality in the room, the recording quality and the background noise may have a significant effect on the transcriptions' accuracy rates.

Acoustic and phonetic knowledge could be incorporated into the current statistical models, in an attempt to deal with these issues, minimise performance degradation and therefore, improve the accuracy rate of transcription.

Various areas need further investigation; a complete review of current models and their processes needs to be produced. Therefore, a set of experiments has been planned. Part of the experiments has already been conducted and was intended as a preliminary study of the state of ASR systems, in order to establish baseline figures for trained and untrained systems, native and non-native speakers, and laboratory and classroom situations. The rest of the experiments will collect additional data in genuine lecture situations, in an attempt to obtain more concrete results and statistically valid figures. In addition, a detailed analysis of human processes and their mechanisms will be conducted, in order to identify processes that could be used in speech technology.

## 5. REFERENCES

[1] Bain, K., Basson, S. and Wald, M. 2002. Speech Recognition in University Classrooms: Liberated Learning Project. In Proceedings of the 5th International ACM Conference on Assistive Technologies (Edinburgh, Scotland, July 8-10, 2002). Assets '02. ACM, New York, NY, 192-196. DOI=http://doi.acm.org/10.1145/638249.638284

[2] Demetriadis, S. and Pombortsis, A. 2007. e-Lectures for Flexible Learning: A Study on Their Learning Efficiency. Journal of Educational Technology & Society. 10, 2, 147-157.

[3] Dusan, S. and Rabiner, L. 2005. Can Automatic Speech Recognition Learn More from Human Speech Perception. In Proceedings of the 3rd Romanian Academy Conference on Speech Technology and Human-Computer Dialogue (Cluj-Napoca, Romania, 2005). Romanian Academy.

[4] Kheir, R. and Way, T. 2007. Inclusion of Deaf Students in Computer Science Classes Using Real-Time Speech Transcription. In Proceedings of the 12th Annual SIGCSE Conference on Innovation & Technology in Computer Science Education (Dundee, Scotland, June 25-27, 2007). ITiCSE '07. ACM, New York, 192-196. DOI=http://doi.acm.org/10.1145/1269900.1268860

[5] Leitch, D. and MacMillan, T. 2003. Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research. Year III Report. Saint Mary's University, Nova Scotia.

[6] Lickley, R., McKelvie, D. and Bard, E. 1999. Comparing Human & Automatic Speech Recognition Using Word Gating. In Proceedings of the ICPhS Satellite Meeting on Disfluency in Spontaneous Speech (UC Berkeley, California, July 30, 1999). 23-26.

[7] SENDA. 2001. Special Educational Needs And Disability Act. http://www.opsi.gov.uk/acts/acts2001/20010010.htm

[8] Smith, A., Ling, P. and Hill, D. 2006. The Adoption of Multiple Modes of Delivery in Australian Universities.

Journal of University Teaching and Learning Practice. 3, 2, 67-81.

[9] Wald, M. 2004. Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality. In Proceedings of the 34th ASEE/IEEE Frontiers in Education Conference (Savannah, GA, October 20-23, 2004).