WHERE ARE THE SPEECH PRODUCTION MODELS IN OUR SPEECH PROCESSING SYSTEMS?

Richard Rose

Department of Electrical and Computer Engineering
McGill University
Montreal, Quebec

RESUMEN

The conjecture that automatic speech recognition and speech synthesis systems should profit from speech production-oriented representations of speech has motivated research in a number of areas over several decades. The focus of this research has ranged from attempts to construct complete multidimensional articulatory models to the development of systems that detect and integrate acoustic-phonetic features in more traditional ASR systems. It is believed by many researchers that this class of approaches may provide the only solution to some of the fundamental problems in existing state of the art systems. These problems include the limitations imposed by the use of canonical dictionaries in ASR and by the inability to maintain robust performance as portions of the speech spectrum are masked by noise. The impact of these limitations has become ever more apparent as academic and industrial laboratories focus on task domains that evoke increasingly more spontaneous speech in increasingly difficult acoustic environments. It is generally agreed, however, that the major advances in ASR have come from the development of more powerful statistical modeling formalisms which incorporate only minimal speech production or perceptual knowledge. This talk will attempt to explain why speech production-oriented approaches are important and at the same time explain why the development of efficient formalisms that directly incorporate knowledge of actual articulatory dynamics for ASR is extremely difficult. In light of these difficulties, it will address the progress and potential success of efforts to incorporate speech production knowledge into existing statistical modeling formalisms.

Biografía: Richard Rose

Richard Rose received B.S. and M.S. degrees in Electrical Engineering from the University of Illinois, and Ph.D. E.E. degree from the Georgia Institute of Technology. From 1980 to 1984, he was with Bell Laboratories working on signal processing and digital switching systems. From 1988 to 1992, he was with MIT Lincoln Laboratory working on speech recognition and speaker recognition. He was with AT&T from 1992 to 2003, specifically in the Speech and Image Processing Services Laboratory at AT&T Labs - Research in Florham Park, NJ after 1996. Currently, he is an associate professor of Electrical and Computer Engineering at McGill University in Montreal, Quebec. Professor Rose served as a member of the IEEE Signal Processing Society Technical Committee on Digital Signal Processing from 1990 to 1995, and was on the organizing committee of the 1990 and 1992 DSP workshops. He was elected as an at large member of the Board of Governors for the Signal Processing Society during the period from 1995 to 1997. He served as an associate editor for the IEEE Transactions on Speech and Audio Processing from 1997 to 1999 and is currently a member of the editorial boards for the Speech Communication Journal and for the IEEE SPS Newsletter. He was a member of the IEEE SPS Speech Technical Committee (STC) from 2003 to 2005 and was founding editor of the STC Newsletter. Most recently, he was the Co-chair of the IEEE 2005 Workshop on Automatic Speech Recognition and Understanding.