A Tutorial on
Automatic Speech Recognition for Wireless Mobile Devices

R. C. Rose and S. Parthasarathy
AT&T Labs-Research, Florham Park, NJ 07932

SUMMARY

This tutorial will cover topics in automatic speech recognition (ASR) that are specific to the implementation of ASR systems on wireless mobile devices. Emphasis will be placed on algorithms and architectures that relate to ASR in the context of the mobile devices themselves, the environments they are used in, the networks they are connected to, and the target applications that they support. The tutorial will have two major parts. The first part will consist of a survey of existing mobile ASR applications, architecures, and supporting technology. The second part will present algorithms that have been developed to make ASR systems more robust, more efficient, and more flexible for mobile applications.

The first part of the tutorial is meant to provide a broad background and to motivate the techniques presented in part two. No background in signal processing, statististics, or speech recognition technology is assumed for this portion of the tutorial. The survey will begin with a discussion of ASR based applications using mobile devices along with a discussion of the issues affecting ASR performance that are specific to mobile applications. Examples of issues that are specific to mobile ASR include rapidly varying acoustic environments, bursty wireless channels with a variety of coding and compression algorithms, and interaction of voice input with other interface modalities. A range of system architectures will be described along with examples of how different levels of ASR functionality can be implemented in different parts of the network. A survey of supporting technology including specialized transducers, specialized devices,and language resources for mobile applications will also be presented.

The second part of the tutorial covers topics from the speech recognition literature that relate to performance and implementation of ASR for mobile device applications. Some background in signal processing, statistical modeling, and ASR is assumed for this part of the tutorial. There are three major technology areas that will be covered. First, robust ASR techniques are described. This will include a brief background of robust acoustic modeling techniques, a description of particular sources of distortion that are characteristic of mobile acoustic environments and channels, and discussion of robust algorithms that have been proposed in the literature for mobile applications. Second, techniques for efficient ASR implementation will be described. These include techniques for reducing size and increasing speed for acoustic models and recognition networks. Finally, the last technology area is related to the use of word lattices generated by the speech recognizer for applying additional knowledge sources available from the task domain. These techniques include concatenating and re-scoring lattices obtained from multiple utterances in system initiated interactions. They also include techniques for decomposing word lattices so that portions of a continuous utterance can be isolated, weighted, and associated with individual fields in an interactive scenario.

SYLLABUS

I. Survey of existing applications, architectures, and supporting technology

1. **ASR applications on handheld, mobile devices**

Portable Devices	processing power and user interfaces
Network Connectivity	protocols, bandwidth, and voice/data channels
Potential Applications	data entry, information retrieval, command-and-control, and dictation

2. **Architectures for Implementing ASR Services on Mobile Devices**

Distributed Functionality: ASR engine, feature analysis, dialog, and applications databases

System Requirements: distributed vs. embedded implementations, computational complexity, latency, security and consistency of application databases

3. **ASR issues unique to mobile applications**

   Device Related Issues: microphone characteristics and interaction of voice with other input modalities

   Channel Issues: channel fading and error characteristics, speech coding and compression algorithms, network delays

   Environment Issues: background noise and effect on speaking style

4. **Supporting Technology: Transducers and Specialized Devices**

II. Algorithms for Robust, efficient, flexible ASR Applications

1. **Fundamentals of ASR and Acoustic Modeling**

   ASR Algorithms: HMM-based speech recognition

   Robust Acoustic Modeling: discriminative feature space transformations and feature space normalization; model adaptation; combination of speech and background models; recovery of missing features

2. **Robust Algorithms Applied to Mobile Applications**

   Adaptation: exploiting side information such as speaker identity and transducer/device type

   Background Modeling: nonstationary models, estimation algorithms, speech detection

   Multi-Modal Input: joint decoding of pen and speech input

3. **Efficient ASR Implementations**

   Computation: fixed-point arithmetic, VQ, tied-distributions, Gaussian selection

   Memory: small memory footprint, optimized decoder networks

4. **Lattice Based Constraint Modeling**

   Language Constraints: combination of multiple word lattices; dynamic language model generation

   Salient Information Extraction: confidence-based lattice segmentation; extraction of salient fragments of the utterance

5. **Survey of current projects involving ASR on mobile devices**

   Research projects: information access using speech and pen on a PDA (MiPad, MATCH); interaction with a *forms* interface on a PDA using speech

   Commercial embedded ASR engines

   Commercial special purpose devices with multimodal interfaces

Biographical Sketches

Richard Rose received B.S. and M.S. degrees from the Electrical and Computer Engineering Department at the University of Illinois. He obtained a Ph.D. E.E. degree from what is now known as the The Center for Signal and Image Processing (CSIP) at the Georgia Institute of Technology in 1988 with a thesis in speech coding and speech analysis. From 1980 to 1984, he was with Bell Laboratories, now a division of Lucent Technologies working on signal processing and digital switching systems. From 1988 to 1992, he was a member of the Speech Systems and Technology group, now called the Information Systems Technology Group, at MIT Lincoln Laboratory working on speech recognition and speaker recognition. He has been with ATT since 1993, and is currently a principal member of technical staff in the Speech and Image Processing Services Laboratory at ATT Labs - Research in Florham Park, NJ.
Mr. 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 has served as an adjunct faculty member of the Georgia Institute of Technology. He was elected as an at large member of the Board of Governers for the Signal Processing Society during the period from 1995 to 1997, and served as membership coordinator during that time. He was an associate editor for the IEEE Transactions on Speech and Audio Processing from 1997 to 1998, and is currently a member of the IEEE SPS Speech Technical Committee. He spent the spring of 1996 at Furui Lab at NTT in Tokyo. He is the author of over 60 refereed papers and is an inventor on 10 patents. He is a senior member of the IEEE and is member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.

S. Parthasarathy received M.S. and Ph.D. degrees from the Electrical Engineering Department at the University of Rhode Island. His masters research was in the area of robotic control and his doctoral work focussed on robust parameter estimation. He joined AT&T Bell Laboratories, Murray Hill, NJ, in 1987 and worked on articulatory speech synthesis and the problem of automatically estimating articulatory parameters from the speech signal. From 1990 to 1993, he was with Bell-Northern Research in Montreal, working on large vocabulary speech recognition for telephony applications. He then returned to AT&T Bell Laboratories (now AT&T Labs-Research) and has been involved in research on speech and speaker recognition. He is a member of the IEEE and serves on the Speech Technical Committee of the IEEE Signal Processing Society.