

# SESSION 8B: ROBUST SPEECH PROCESSING

*Jordan R. Cohen, Chair*

Center for Communications Research  
Thanet Road  
Princeton, NJ 08540

## ABSTRACT

Four papers are briefly reviewed.

### 1. The Papers

This session consists of two types of papers. The first two, "Multiple approaches to robust speech recognition" and "Reduced channel dependence for speech recognition" present computational methods for minimizing the acoustic and speaker differences in particular recognizers. The third paper, "Experimental results for baseline speech recognition performance ..." presents preliminary experiments in using an array of microphones for acoustic focusing, while the last, "Phonetic classification on wide-band and telephone quality speech", presents a baseline phonetic recognition result for telephone TIMIT.

In the first paper, the Carnegie Mellon group define several algorithms for jointly compensating for noise and linear filtering in incoming data. Codeword Dependent Cepstral Normalization was found to be advantageous when training with one microphone and testing with another. It was also helpful when used with data from a microphone array. Results were less clear when the algorithm was applied to an auditory front end, but work is continuing.

The SRI paper introduced a long-term filtering algorithm to adjust for acoustic differences between training and test. The best results were found using highpass filtering on channel energies in conjunction with simple noise removal. It was interesting to note that, even after these algorithms, simultaneous recordings through different microphones were quite different.

The Brown paper reports early results on a microphone beam-steering array. They report a series of interesting problems, some solved (microphone mounting), and some not (ceiling reflections). The search for an effective array continues.

Finally, the NYNEX paper reports on comparative phonetic recognition of TIMIT vs NTIMIT. The telephone version of TIMIT appears to induce 1.3 times as many

errors as TIMIT, with a frequency distribution of errors which is expected from the inherent power of the underlying phonemes. This work is offered as a benchmark against which to measure future systems.

### 2. Discussion

Discussion was congenial and to the point. More work in this area will appear in future meetings.