# Session 12: Speech Recognition II

**Jordan Cohen, Chair**
IDA Center for Communication Research
Thanet Road
Princeton, NJ 08540

This session was the final formal meeting, and contained the "*et cetera*" papers. These all discussed speech recognition algorithms, but these differed from the standard HMM's reported during most of the meeting. Each of these papers comes to grips with an aspect of training data, uncertainty in the recognition, and number of parameters to be set.

In the first paper, Jim Glass sketched some of the avenues pursued in the MIT SUMMIT work. They have been working on input normalization, different segmental representations, and issues in experimental phonetics. The auditory model requires careful normalization (a condition also explored by IBM in several ICASSP papers), and a non-linear adaptive technique was described without results. Boundary classification was discussed, and a modified metric for computing edges was shown to give improved phonetic recognition. In addition, an MLP classifier was developed and tested using an augmented boundary representation. Every time additional information was included in the MLP classifier performance improved, but overall the best classification was not as good as that reported by Phillips in JASA.

Ron Cole then presented his recent work on spoken alphabet recognition using MLP techniques and a variety of acoustical and segmental features. He reported reasonable error rates, and success in performing name retrieval from a database using spelled input. Secondary networks were used to resolve acoustic ambiguity, and were not universally successful in this task. Unfortunately Ron did not present performance figures on Brown's substantial E-set corpus, so it is difficult to calibrate his work with previous results.

Matt Lennig presented a review of the INRS work with large vocabulary recognition for isolated English using HMM techniques. He outlined the development of trigram models, microsegmental models, and duration models in the traditional framework. The most important recent development was switching from a VQ front end to a mixture of Gaussians. Performance was in the range of 7% word error with the best system using an 86,000 word vocabulary.

The final paper was a presentation by Martin Russell reviewing the RSRE results on the Airborne Reconnaissance Task. The speaker dependent task uses HMM techniques with highly stylized text. The most interesting observations from this work were observed variations in performance as a function of the number of free parameters in the model – an obvious peak in performance can be seen at 40-80 thousand parameters. Martin observes that with fewer parameters the system cannot capture the speech variability, while with more parameters the training data in his corpus is inadequate to set the model.

These papers taken together were a refreshing look at non-ARPA funded speech recognition projects, together with a few novel ideas from MIT. The audience was quiet and undemanding.

Charles Wayne ended the session on a positive note, remarking that there continued to be progress in speech recognition, and that the natural language effort had passed an historic milestone in its quantitative tests of the OAG data. He wished everyone more success in the future.