

ROBUST CONTINUOUS SPEECH RECOGNITION

PIs: John Makhoul and Richard Schwartz
makhoul@bbn.com, schwartz@bbn.com

BBN Systems and Technologies
70 Fawcett Street
Cambridge, MA 02138

PROJECT GOALS

The primary objective of this basic research program is to develop robust methods and models for speaker-independent acoustic recognition of spontaneously-produced, continuous speech. The work has focussed on developing accurate and detailed models of phonemes and their coarticulation for the purpose of large-vocabulary continuous speech recognition. Important goals of this work are to achieve the highest possible word recognition accuracy in continuous speech and to develop methods for the rapid adaptation of phonetic models to the voice of a new speaker.

RECENT RESULTS

- Ported the BYBLOS system to the Wall Street Journal (WSJ) corpus. We found that the techniques that we had developed for recognition of the ATIS corpus worked quite well without modification on the WSJ corpus.
- Performed several key experiments on the WSJ corpus. We verified our conjecture that a speaker-independent system trained on a small number of speakers has about the same word error rate as a system trained on a large number of speakers, assuming the same total amount of training speech. This is the first time that this result has been performed in a well-controlled way for large vocabulary speech recognition. We also verified that training the system separately on each of the speakers and averaging the resulting models results in essentially the same performance as training on all of the data at once. These results have wide ranging implications for data collection and system design.
- We have shown that, for large vocabulary recognition, a speaker-independent system will have about the same error rate as a speaker-dependent system when the speaker-independent system is trained on about 15 times as much speech as the corresponding speaker-dependent system.
- We showed that a simple blind deconvolution method for microphone independence, in which the mean cepstrum is subtracted from each cepstrum vector, is somewhat better than the RASTA method.
- Developed a new algorithm for microphone independence which uses a codebook transformation, based on selection among several known microphones. The algorithm reduced the word error rate for unknown microphones by 20% over using blind deconvolution alone.
- In the Nov. 1992 speech recognition test on the ATIS domain, our BYBLOS system continued to give the best results of all sites tested, with a 30% reduction in word error over last year. In our first test on the WSJ corpus, our system had the second lowest error rates.
- Chaired the CSR Corpus Coordinating Committee.

PLANS FOR THE COMING YEAR

For the coming year, we plan to continue our work on improving speech recognition performance both on the Wall Street Journal corpus and on the spontaneous ATIS speech corpus. We plan to explore different parameterizations of the speech signal and new models for microphone and speaker adaptation.