

ROBUST CONTINUOUS SPEECH RECOGNITION

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

BBN Systems and Technologies, 10 Moulton St., Cambridge, MA 02138

OBJECTIVES

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 our BYBLOS speech recognition software to run on Silicon Graphics Inc. workstations, in addition to Sun workstations. In the process, we consolidated our programs and modularized them to make them more easily portable to other sites and applications.
- Developed methods for modeling spontaneous speech effects in the Airline Travel Information System (ATIS) domain.
- Developed a novel method for silence modeling in spontaneous speech, especially to deal with the problem of missing silences in the transcriptions. The method works iteratively by hypothesizing silences everywhere in the grammar, performing recognition, correcting the transcriptions based on the recognition, and then retraining.
- Developed a trigram language model for the ATIS domain. This was possible because of the availability of sufficient

training data. Because of the potential large amount of computation associated with trigram grammars, we achieved large savings in computation by simply rescoring an N-best list that was produced using a bigram grammar.

- In the most recent speech recognition test on the ATIS domain, using data collected from five different sites, our BYBLOS system achieve a 9.6% average word error rate over all utterances. This performance was the best among all sites tested.
- The N-best paradigm allows the simple and modular integration of many knowledge sources. Recently, we have combined our BYBLOS HMM system to another system at Boston University based on Stochastic Segment Models. The hybrid system improved speech recognition performance over our state-of-the-art HMM system.
- A hybrid system consisting of our BYBLOS system and another phonetic modeling technique using neural networks, called Segmental Neural Networks, has also succeeded in improving performance over our HMM system by reducing the error rate by 25%.

PLANS FOR THE COMING YEAR

For the coming year, we plan to continue our work on improving speech recognition performance on spontaneous speech in the ATIS domain. In addition, we plan to start work on the new Wall Street Journal continuous speech recognition corpus. Our research into improved modeling will include the exploration of new acoustic features and methods for rapid speaker adaptation.