ROBUST CONTINUOUS SPEECH RECOGNITION

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

BBN Systems and Technologies
70 Fawcett St.
Cambridge, MA 02138

1. 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.

2. RECENT RESULTS

During the last year, we have:

- Developed a new 5-pass decoding algorithm that allows us to incorporate trigram language models and cross-word coarticulation models directly within the N-best search. The new decoder is considerably faster than the previous one and results in slightly higher accuracy.

- Participated in the December 1993 ARPA evaluations. On the baseline hub test, we achieved a 14.3% word error rate. Our result for the primary test in which we expanded the vocabulary and grammar was 12.3%, which was substantially better than any result produced by an ARPA site, and second only to one other result.

- In a spoke test for outlier speakers, our overall results show that the baseline performance for speakers with foreign accents is 4 times worse than that for native speakers. By using speaker adaptation, the error rate was reduced by more than a factor of 2.

- In a spoke test for known alternate microphones, our recognition performance with the boom microphone in the cross-channel condition did not degrade much relative to the control condition.

- In the spoke for spontaneous dictation, we increased the vocabulary from 20K to 40K words, and also added about 1000 words that occurred in the spontaneous training data but not in the original vocabulary. This reduced the word error from 26% to 20%.

- Considered several powerful models to use in search algorithms, including segmental neural networks (under a separate effort), a 13-state phoneme model, and a stochastic segment model (in collaboration with Boston University). The combination of all of the models produced the lowest error rate.

- Began exploring a new method for system adaptation to speakers, called auto-adaptation. This method will improve performance by making appropriate use of the information that a whole utterance is spoken by the same speaker in a single environment.

- Performed experiments to better understand issues relating to microphone independence. We developed a technique in which the training is performed with a single high quality microphone, and the test utterance with the unknown microphone is transformed to resemble the training microphone as much as possible. We found that our algorithm was able to classify the microphone into the correct microphone class about 98% of the time, and the resulting normalization reduced the word error rate by 33%.

- Chaired the CCCC (CSR Corpus Coordinating Committee), and participated in other committees. The CCCC was responsible for developing the "hub and spokes" paradigm for the evaluation of CSR systems.

3. PLANS FOR THE COMING YEAR

We will continue our work on improving speech recognition performance both on the Wall Street Journal corpus and on the spontaneous ATIS speech corpus. Work will focus on improved phonetic models, adaptation methods, and robustness against different acoustic channels and new vocabularies and grammars.

445