SESSION 3: CONTINUOUS SPEECH RECOGNITION*

Douglas B. Paul, Chair

MIT Lincoln Laboratory
Lexington, MA, 02173

The papers in this session focus on techniques for and applications of large-vocabulary continuous speech recognition. The technique oriented papers discuss techniques for channel compensation, fast search, acoustic modeling, and adaptive language modeling. The applications oriented papers discuss methods for using recognizers for language identification, speaker identification, speaker-sex identification, and keyword spotting.

In "Efficient Cepstral Normalization for Robust Speech Recognition," Liu et al. discuss several preprocessors for channel (including microphone) compensation. Several of these techniques cover only channel equalization and several also account for additive noise. The authors obtained the their best unknown-microphone performance using a technique that accounts for both the equalization and the additive noise.

In "Comparative Experiments on Large Vocabulary Speech Recognition," Schwartz et al. describe several aspects of the BBN recognition system. They briefly describe their use of forward-backward N-best search. They also found a number of small modeling improvements which add up to a significant total improvement in performance. Finally, they describe their results on channel compensation—which are not completely in agreement with the results of the previous paper.

"An Overview of the SPHINX-II Speech Recognition System" by Huang et al. describes the CMU SPHINX-II recognition system. It describes their feature set, their use of tied-mixture (semicontinuous) pdfs, their statewise-clustered phone models (senones) and their search strategy. It also describes a technique for combination of the acoustic and language model probabilities which does not assume statistical independence between the two information sources.

Murveit et al. describe the search strategy used in the SRI recognizer in "Progressive-Search Algorithms for Large Vocabulary Speech Recognition." This progressive search strategy performs the search several times, initially using inexpensive coarse models and then progressively more detailed and expensive models on each iteration. Information from each iteration is used to produce a smaller word network to constrain the search space of the next iteration.

In "Search Algorithms for Software-Only Real-Time Recognition with Very Large Vocabularies," Nguyen et al. describe the techniques used at BBN to achieve real-time recognition of a 20K word task. The techniques center on using a very fast approximate forward search. Information saved from this forward search is then used to constrain a backwards A* search. This backwards search is inherently fast and can provide an N-best sentence list for more detailed reevaluation.

Gauvain and Lamel, in "Identification on Non-Linguistic Speech Features," apply a phonetic recognizer to several other purposes. By using multiple phone sets running independently in parallel, they use the output likelihoods to identify speaker sex, speaker identity, and the language. In each case the phone sets are matched to the aspect to be identified.

"On the Use of Tied-Mixture Distributions" by Kimball and Ostendorf discusses the use of tied Gaussian-mixture pdfs, which have been shown to yield good recognition performance in standard HMM recognizers at a number of sites. They discuss the application of tied mixtures to their stochastic segment recognition models and show improved performance over a non-mixture based system.

In "Adaptive Language Modeling Using the Maximum Entropy Principle," Lau et al. describe a new method for recognition-time adaptation of the of the language model based upon the recent past. The technique uses "trigger" words that signal an increased probability for other words in the near future. They report a greater reduction in perplexity than that obtained by the use of a "caching" adaptive language model.

In "Improved Keyword-Spotting Using SRI's DECIPHER (TM) Large-Vocabulary Speech-Recognition Sys-

*This work was sponsored by the Advanced Research Projects Agency. The views expressed are those of the author and do not reflect the official policy or position of the U.S. Government.

tem," Weintraub describes use of a large-vocabulary recognizer to a keyword-spotting task. He shows significantly improved performance over the traditional technique of searching for only the keywords against a background of unknown words.

Peskin et al., in "Topic and Speaker Identification via Large Vocabulary Continuous Speech Recognition," describe the use of the Dragon large-vocabulary recognizer to perform both topic and speaker identification. The technique described here uses a topic and speaker-independent recognizer to produce a word sequence. This word sequence can then be economically rescored using topic-dependent language models for topic identification or speaker-dependent acoustic models for speaker identification. The authors report good performance on both tasks.