Robust Continuous Speech Recognition Technology Program Summary*

Clifford J. Weinstein and Douglas B. Paul, Principal Investigators
Lincoln Laboratory, M.I.T.
Lexington, MA 02173-9108

Program Goals

The major objective of this program is to develop and demonstrate robust, high-performance continuous speech recognition (CSR) techniques and systems focused on applications in spoken language systems (SLS). A key supporting objective is to develop techniques for integration of CSR and natural language processing (NLP) systems in SLS applications. The CSR techniques are based on a continuous-observation hidden Markov model (HMM) approach, using tied Gaussian mixtures to model the speech parameters. A stack-decoder control structure is being developed and utilized, both for efficient large-vocabulary recognition, and to facilitate integration of CSR and NLP systems.

Background

The Lincoln program began with a focus on improving speaker stress robustness for the fighter aircraft environment. A robust HMM isolated-word recognition (IWR) system was developed with very high performance under stress conditions. The robust HMM system has since been adapted and extended to large vocabulary CSR. This effort has included the development of a number of new modeling and recognition techniques, and the resulting tied-mixture HMM CSR system achieved state-of-the-art performance for both speaker-dependent (SD) and speaker-independent (SI) recognition on the DARPA Resource Management (RM) database. Current work focuses on extension to the new Wall Street Journal (WSJ) CSR corpus, with vocabularies of 5,000–20,000 words.

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

Recent Accomplishments

Recent accomplishments include: (1) major contributions to the design of the WSJ corpus, and development and implementation of the necessary text preprocessing system to make text and language models available to recording sites and testing sites; (2) development of an efficient stack decoder algorithm for large-vocabulary CSR; (3) development of fast match techniques to expedite large-vocabulary recognition; and (4) application of the tied-mixture, stack-decoder-based HMM CSR, with fast-match to obtain a first set of results on the new WSJ corpus.

Plans

Plans for the current program include: (1) extensive development and test of large-vocabulary CSR techniques of the new CSR corpus; (2) continued development of the tied-mixture HMM CSR system, including adaptive training and recognition techniques, mixture weight smoothing, and improved speaker-independent techniques; (3) further development of the stack-decoder-based HMM, for integration with the CSR/NLP interface system and with NLP systems developed at other sites; and (4) exploration of advanced acoustic modeling techniques for improved recognition.