Robust Speech Recognition Technology
Program Summary
Principal Investigators: Clifford J. Weinstein and Douglas B. Paul
MIT Lincoln Laboratory

The major objective of this program is to develop and demonstrate robust, high-performance continuous speech recognizer (CSR) techniques and systems focused on application 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, which has previously demonstrated high performance for normal speech and robustness for stressed speech. The motivation is that current state-of-the-art CSR systems must be improved in performance and robustness for advanced SLS environments, with variabilities including those due to spontaneous speech, noise, and task-induced stress. The effort in CSR/NLP integration is focused on development of a structured CSR/NLP interface, which will allow effective collaboration with and between other groups developing NLP and/or CSR systems.

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 99% accuracy under stress conditions, representing more than an order-of-magnitude reduction in error rate relative to a baseline HMM system. A robust CSR system was then developed and integrated into a voice-controlled flight simulator — a simple, but a complete SLS involving a stressing, real-time task. The robust HMM recognition system was then adapted and extended to large vocabulary CSR. This effort has included development of a number of new modelling and recognition techniques which have resulted in state-of-the-art performance for both speaker-dependent (SD) and speaker-independent (SI) recognition on the DARPA Resource Management (RM) database.

Recent accomplishments include: (1) development of tied mixture techniques using observation pruning, which when incorporated with other improvements into the HMM recognizer, have yielded performance for both SD and SI training which is equivalent to the best reported on the October 1989 RM test set; (2) development and implementation, in initial prototype form, of a structured CSR/NLP interface including the required protocols, the stack decoder control structure, and initial CSR and NLP simulators to allow testing; (3) development of effective and efficient stack decoder search strategies for continuous speech recognition; and (4) analysis of the interaction between true source, training, and testing language models, and the effect of this interaction on performance testing of CSR systems; and (5) development of a new phonetic context model, the semiphone, which produces results similar to traditional triphone systems, but which allows significant reduction in the number of models which must be trained.

Plans for the current program include: (1) continue to improve HMM CSR performance using tied-mixture techniques, new semiphone techniques to reduce the number of states which must be trained, and new acoustic-phonetic modelling and recognition techniques; (2) complete the development of the new stack decoder control structure, and convert the HMM CSR to this control structure; (3) complete the prototype implementation of the CSR/NLP interface, and collaborate with other groups in application of this interface to integration of CSR and NLP systems developed at different sites.