Robust Speech Recognition Technology
Program Summary[1]

Principal Investigators: Clifford J. Weinstein and Douglas B. Paul
MIT Lincoln Laboratory

The objective of this program is to develop and demonstrate robust, high-performance continuous speech recognizer (CSR) techniques and systems focussed on application in spoken language systems (SLS). The 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 focus of the robust CSR techniques on SLS applications is being facilitated by development and implementation of a well-structured interface between a CSR and a natural language processor (NLP), allowing collaboration with other groups developing NLPs for SLS applications.

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.

The robustness techniques were then adapted to large vocabulary CSR with high performance for both speaker-dependent and speaker-independent tasks on the DARPA Resource Management database.

Recent accomplishments include: (1) development and integration into HMM CSR system of tied-mixture and word-context-dependent phone-modelling techniques; (2) development and demonstration of a voice-controlled flight simulator system — a simple, but complete SLS which integrates the robust CSR into a stressing, real-time task; and (3) development of a proposed specification for a structured interface between CSR and NLP, based on a stack decoder control structure.

Plans for the current program include: (1) continue to improve HMM CSR performance and robustness using tied-mixture techniques and techniques to match the model complexity to amount of training data; (2) develop new acoustic-phonetic modelling and recognition techniques; (3) complete the CSR/NLP interface design, incorporating inputs from other groups, and develop a prototype interface implementation; (4) convert HMM CSR to use a stack decoder control structure, to match the CSR-NLP interface and to allow integration of the Lincoln CSR with an NLP developed at another site.

[1]This work was sponsored by DARPA.