Evaluation and Analysis of Auditory Model Front Ends for Robust Speech Recognition Program Summary*

Richard P. Lippmann, Principal Investigator
Lincoln Laboratory, M.I.T.
Lexington, MA 02173-9108

Program Goals

The purpose of this work is to integrate a number of auditory model front ends into a high-performance HMM recognizer, to test and evaluate these front ends on noisy speech, and to analyze the results in order to develop a more robust front end which may combine features of a number of the current auditory model-based systems.

Background

This project was motivated by the need for improved speech recognition in noise, and by expectation that auditory model front ends could make recognition more robust to noise, microphone variation, and speaking style. The project began in FY91 with the initial goal of implementing, evaluating, and comparing four promising auditory front ends: (1) the mean-rate and synchrony outputs of S. Seneff's auditory model; (2) the ensemble interval histogram (EIH) model developed by O. Ghitza; (3) the IMELDA model due to M. Hunt; and (4) an auditory model developed by J. Cohen. The initial effort has included implementation and integration of the first three models with a Lincoln HMM recognizer, and testing with speech in white noise and in a background of speech babble.

Recent Accomplishments

The Seneff and EIH front ends have been implemented and integrated with an HMM system, and have yielded promising results on the TI-105 stressed speech corpus (a 105-word vocabulary isolated-word corpus which was collected at Texas Instruments), with white noise and speech babble background. As compared with the mel-cepstrum technique, the auditory models have similar performance at high SNR and slightly better performance at low SNR. In order to select the best features from each front end, dimensionality reduction techniques, including Linear Discriminant Analysis (LDA) and Principal Components Analysis (PCA) have been implemented. LDA has been applied to obtain a transformation matrix for IMELDA, with the transformation matrix developed using 100,000 labeled speech frames (1,000 sentences) from the TIMIT corpus. Testing of IMELDA is about to begin on the TI-105 corpus.

Plans

Plans include: (1) experiments on the TI-105 corpus in noise using IMELDA; (2) obtaining Cohen's front end and integrating it with the HMM system; (3) experiments with all front ends on the effect of spectral variability due to talking style in TI-105; (4) exploring the effect of dimensionality reduction in all the front ends; and (5) continuous speech recognition experiments on the Resource Management corpus.

*This work was sponsored by the Defense 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.