Segment-Based Acoustic Models with Multi-level Search Algorithms for Continuous Speech Recognition

Mari Ostendorf *J. Robin Rohlicek*

Boston University BBN Inc.
Boston, MA 02215 Cambridge, MA 02138

PROJECT GOALS

The goal of this project is to develop improved acoustic models for speaker-independent recognition of continuous speech, together with efficient search algorithms appropriate for use with these models. The current work on acoustic modelling is focussed on stochastic, segment-based models that capture the time correlation of a sequence of observations (feature vectors) that correspond to a phoneme. Since the use of segment models is computationally complex, we are investigating multi-level, iterative algorithms to achieve a more efficient search. Furthermore, these algorithms will provide a formalism for incorporating higher-order information. This research is jointly sponsored by DARPA and NSF.

RECENT RESULTS

- Developed methods for robust context modeling for the stochastic segment model (SSM) using tied covariance distributions, and investigated different regions of tying using clustering techniques. On the RM Oct 89 test set, improvements reduced the error rate of the SSM by a factor of two (9.1% to 4.8% word error), and the current BBN-HMM/BU-SSM combined system achieves 3.3% word error.
- Determined that linear models have predictive power similar to non-linear models of cepstra within segments, and explored different models of the statistical dependence of cepstral coefficients in the context of a dynamical system (DS) model.
- Evaluated the dynamical system model in phoneme recognition (as opposed to classification in previous work) using the split-and-merge search algorithm. The DS model outperforms the independent-frame model on the TIMIT corpus.
- Reformulated the recognition problem as a classification and segmentation scoring problem, which allows more general types of classifiers and nontraditional feature analysis. Demonstrated that for equivalent feature sets and context-independent models, the two methods give similar results.
- Investigated duration models conditioned on speaking rate and pre-pausal location, and improved performance by increasing the weight of duration by including the duration probabilities separately in the N-best score combination.
- Analyzed the behavior of recognition error over the weight space for HMM and SSM scores in the N-best rescoring paradigm. Addressed the problem of local optima with a grid-based search, determined that the relative weights for the HMM and SSM scores are similar, and discovered a significant mismatch problem between training and test data.
- Extended Bayesian techniques for speaker adaptation and evaluated these in the RM word recognition task, achieving 16% reduction in error using 3 minutes of speech with simple mean adaptation techniques. Covariance adaptation techniques seem to require more speakers for training the priors.
- Developed a multi-level stochastic model of speech that can take advantage of multi-rate signal analysis; evaluating the model for the two-level case with cepstral features shows improved performance over a single-level model.

PLANS FOR THE COMING YEAR

The plans for the coming year reflect the fact that this grant ends in summer 1992.

- Continue work in the classification and segmentation scoring paradigm: demonstrate improvements associated with novel models and/or features, and extend the probabilistic framework to allow context-dependent models.
- Extend context modeling through further exploration of clustering and to recently developed DS or multi-level variations.
- Implement different auditory-based signal processing algorithms, and evaluate their usefulness for recognition through a series of experiments on the TIMIT corpus.