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

Mari Ostendorf   J. Robin Rohlicek

Boston University   BBN Systems and Technologies Corp.

Objective:

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 will also investigate 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.

Summary of Accomplishments:

- Refined the stochastic segment model to model time correlation using a Markov time structure and using time-dependent parameter reduction.

- Investigated the utility of sentence-level duration phenomena for incorporation in a multi-level algorithm.

- Investigated robust covariance estimates for limited training data.

- Developed a speaker-independent phoneme classification system that achieves 72% accuracy on the TIMIT database.

Plans:

- Implement segment-based and HMM-based phoneme recognition systems to better understand the relative advantages of these modelling techniques.

- Evaluate further refinements to the stochastic segment model such as use of different parameter estimation methods, alternative variable-to-fixed-length transformations, and incorporation of context modelling.

- Implement a multi-level, iterative search algorithm for phomeme recognition.