

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

- Demonstrated improved phoneme classification performance (23% reduction of error) with left-context-dependent, gender-dependent segment models.
- Developed a new segment model for recognition based on a dynamical system representation of time correlation. Since classical approaches to system identification (parameter estimation) are too costly, an alternative algorithm was developed based on the Estimate-Maximize algorithm. Phoneme classification results show a significant improvement over previous approaches to time correlation modeling.
- Jointly with BBN, developed a methodology for integrating recognition systems based on rescoreing the N-best sentence hypotheses and combining scores to rerank the sentences. Demonstrated improved word recognition performance by combining the BBN Byblos system results with the BU segment model rescoreing. Implemented a segment model rescoreing system using context-dependent models which gives encouraging initial results.
- Applied Bayesian techniques for speaker-adaptation to the stochastic segment model. Adapting distributions means alone, with 30 sentences, yields a 16% re-

duction in phoneme recognition error relative to speaker-independent performance.

- Investigated iterative search algorithms for word recognition.

PLANS FOR THE COMING YEAR

- Extend dynamical system model results by considering different correlation structures and confirming that classification performance extends to recognition applications.
- Investigate the use of context-dependent models in segment-based word recognition in the N-best rescoreing formalism.
- Investigate parallel algorithms for segment recognition implemented on the Connection Machine.
- Extend algorithms for speaker adaptation which include variance adaptation.
- Continued investigation of iterative and multi-level algorithms including: incorporation of global constraints and features, iterative search for word recognition, and progressive application of phoneme context, word and grammar constraints.