Segment-Based Acoustic Models for Continuous Speech Recognition

Mari Ostendorf *J. Robin Rohlicek*

Boston University
Boston, MA 02215

BBN Inc.
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 modeling is focussed on stochastic, segment-based models that capture the time correlation of a sequence of observations (feature vectors) that correspond to a phoneme, hierarchical stochastic models that capture higher level intra-utterance correlation, and multi-pass search algorithms for implementing these more complex models. This research has been jointly sponsored by DARPA and NSF under NSF grant IRI-8902124 and by DARPA and ONR under ONR grant N00014-92-J-1778.

RECENT RESULTS

- Implemented different auditory-based signal processing algorithms and evaluated their use in recognition on the TIMIT corpus, finding no performance gains relative to cepstral parameters probably due to the non-Gaussian nature of auditory features.
- Improved the score combination technique for N-Best rescoring, through normalizing scores by sentence length to obtain more robust weights that alleviate problems associated with test set mismatch.
- Further investigated agglomerative and divisive clustering methods for estimating robust context-dependent models, and introduced a new clustering criterion based on a likelihood ratio test; obtained a slight improvement in performance with an associated reduction in storage costs of a factor of two.
- Extended the classification and segmentation scoring formalism to handle context-dependent models without requiring the assumption of independence of features between phone segments (using maximum entropy methods); evaluated different segmentation scores with results suggesting more work is needed in this area.
- Evaluated a new distribution mapping, which led to an 8% reduction in error on the development test set but no improvement on other test sets.
- Investigated the use of different phone sets and probabilistic multiple-pronunciation networks; no improvements were obtained on the RM corpus, though there may be gains in another domain.
- Extended the two level segment/microsegment formalism to application in word recognition using context-dependent models; evaluated the trade-offs associated with modeling trajectories vs. (non-tied) microsegment mixtures, finding that mixtures are more useful for context-independent modeling but representation of a trajectory is more useful for context-dependent modeling.
- Investigated the use of tied mixtures at the frame level (as opposed to the microsegment level), evaluating different covariance assumptions and training conditions; developed new, faster mixture training algorithms; and achieved a 20% reduction in word error over our previous best results on the Resource Management task. Current SSM performance rates are 3.6% word error on the Oct89 test set and 7.3% word error on the Sep92 test set.

PLANS FOR THE COMING YEAR

- Continue work in the classification and segmentation scoring paradigm; demonstrate improvements associated with novel models and/or features.
- Port the BU recognition system to the Wall Street Journal (WSJ) task, 5000 word vocabulary.
- Develop a stochastic formalism for modeling intra-utterance dependencies assuming a hierarchical structure.
- Investigate unsupervised adaptation in the WSJ task domain.
- Investigate multi-pass search algorithms that use a lattice rather than N-Best representation of recognition hypotheses.