Experiments in Landmark-Based Speech Recognition

Mark Hasegawa-Johnson, Sarah Borys, and Ken Chen

Sound to Sense:
Workshop in Honor of Kenneth N. Stevens,
June, 2004.

This paper reviews a few of our experiments that find, aggregate, and apply the acoustic correlates of distinctive features measured relative to landmarks.

The “infogram” is a measurement of the Shannon mutual information between a binary distinctive feature, implemented at a landmark, and the distribution of energies at a remote time-frequency coordinate. Infogram plots show ”hot spots” in time and frequency that pretty reliably match the acoustic correlates predicted by phonetic theory. Although not classifiers themselves, these plots serve as a guide for development of classifier algorithms based on boosting or neural networks.

We are currently testing knowledge-based measurements, mixture Gaussian models, neural networks, support vector machines, and boosting for the purpose of classifying distinctive features at a landmark. Our best place of articulation classifier, to date, is a linear discriminant combination of knowledge-based measurements; stop place accuracy is 76%. Our best manner classifier to date is an RBF support vector machine, observing a window of 11 consecutive short-time spectra, achieving binary manner-class feature accuracies of 80-100%. Experiments are ongoing, and we expect further improvements by the time of the workshop.

As a test of the relative importance of different distinctive feature sets, we used perfect knowledge of various distinctive features to rescore the word lattice output of an HMM. Perfect knowledge of all features (equivalent to perfect knowledge of the phonemes) reduced word error by 11% relative; knowledge of manner features reduced word error by only 2%.