

ACOUSTICS2008/1832
**Broader range of training voices improves performance of HMM
model of phonemic identification**

J Parchment

University of Arizona, Dept. of Linguistics, Douglass Building, Room 200E, Tucson, AZ 85721, USA

A model proposed by Lin (2005) learns phonetic categories from waveform input. Recorded speech from a set of male talkers is divided into training and test sets. The training set is separated into phonemes, subjected to cepstral analysis, and used as the input to a Hidden Markov Model, which clusters the phonemes into phonemic categories. After this unsupervised learning process, the model is then able to accurately identify speech segments in the test set, showing that relevant acoustic information is captured by the model.

The current study explores the outcome when a model of this type is trained on a range of talkers differing in sex and vocal tract configuration. Preliminary results suggest that this approach can improve performance when testing is generalized to a wider range of new talkers. However, too wide a range of training voices reduces accurate categorization, while too narrow a range reduces generalizability. Continuing efforts seek to quantify the optimum range of training voices and to identify the variables that can predict the degree of improvement in performance on test voices. This work has implications for automatic speech recognition models as well as for issues of speaker normalization.