

From Sound to Sense: 50+ Years of Discoveries in Speech Communication
June 11 - June 13, 2004 at MIT

[F13] **Feature-based pronunciation modeling for automatic speech recognition.** Karen Livescu & James Glass, Computer Sci. and Artificial Intelligence Lab, MIT, Cambridge, MA, USA.

The accurate modeling of pronunciation variation has been cited as a serious obstacle for the automatic recognition of spontaneous speech. Typical approaches to modeling this variation are based on either hand-crafted or data-derived rules for phonetic substitutions, insertions, and deletions. These methods can predict many of the common types of variation, but have had only modest success in improving recognition performance. In addition, phone-based models typically do not account for the full extent of variation seen in spontaneous speech, both because of the paucity of training data for automatically-derived rules and because of the risk of increased inter-word confusability. One aspect of phone-based models that may help to explain these drawbacks is that they do not account for one of the underlying causes of variation, namely the semi-independent evolution of sub-phonetic features.

We investigate an approach to pronunciation modeling based on explicitly representing the evolution of multiple linguistic feature streams. In this approach, pronunciation variation is viewed as the result of asynchrony between feature streams and changes in feature values, rather than the substitution, insertion, and deletion of entire phones. We have implemented a flexible feature-based pronunciation model using dynamic Bayesian networks. As a proof of concept, we have performed word recognition experiments using phonetic transcriptions of utterances from the Switchboard corpus as input to the model. The experimental results, as well as the model's qualitative behavior, suggest that this is a promising way of accounting for the types of pronunciation variation often seen in spontaneous speech.