Abstract

A language identification technique using multiple single-language phoneme recognizers followed by n-gram language models yielded top performance at the March 1994 NIST language identification evaluation. Since the NIST evaluation, work has been aimed at further improving performance by using the acoustic likelihoods emitted from gender-dependent phoneme recognizers to weight the phonotactic likelihoods output from gender-dependent language models. We have investigated the effect of restricting processing to the most highly discriminating n-grams, and we have also added explicit duration modeling at the phonotactic level. On the OGI Multi-language Telephone Speech Corpus, accuracy on an 11-language, closed-set, language identification task has risen to 89% on 45-s utterances and 79% on 10-s utterances. Two-language classification accuracy is 98% and 95% for the 45-s and 10-s utterances, respectively. Finally, we have started to apply these same techniques to the problem of dialect identification.