
**Abstract**

Recent work in phonetic speaker recognition has shown that modeling phone sequences using n-grams is a viable and effective approach to speaker recognition, primarily aiming at capturing speaker-dependent pronunciation and also word usage. This paper describes a method involving binary-tree-structured statistical models for extending the phonetic context beyond that of standard n-grams (particularly bigrams) by exploiting statistical dependencies within a longer sequence window without exponentially increasing the model complexity, as is the case with n-grams. Two ways of dealing with data sparsity are also studied; namely, model adaptation and a recursive bottom-up smoothing of symbol distributions. Results obtained under a variety of experimental conditions using the NIST 2001 Speaker Recognition Extended Data Task indicate consistent improvements in equal-error rate performance as compared to standard bigram models. The described approach confirms the relevance of long phonetic context in phonetic speaker recognition and represents an intermediate stage between short phone context and word-level modeling without the need for any lexical knowledge, which suggests its language independence.