
Abstract

This paper describes an acoustic class dependent technique for text-independent speaker identification on very short utterances. The technique is based on maximum likelihood estimation of Gaussian mixture model representation of speaker identity. Gaussian mixtures are noted for their robustness as a parametric model and their ability to form smooth estimates of rather arbitrary underlying densities. Speaker model parameters are estimated using a special case of the Expectation-Maximization (EM) algorithm and a number of techniques are investigated for improving model robustness. The system was evaluated using a 12 reference speaker population from a conversational speech database and achieved 80% average text-independent speaker identification performance for a 1 second test utterance length.