

Gupta, V., Kenny, P., Ouellet, P., Boulianne, G. and P. Dumouchel. "Multiple Feature Combination to Improve Speaker Diarization of Telephone Conversations" In *Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU2007)*. Kyoto, Japan, December 9-13, 2007.

Abstract

We report results on speaker diarization of telephone conversations. This speaker diarization process is similar to the multistage segmentation and clustering system used in broadcast news. It consists of an initial acoustic change point detection algorithm, iterative Viterbi re-segmentation, gender labeling, agglomerative clustering using a Bayesian information criterion (BIC), followed by agglomerative clustering using state-of-the-art speaker identification methods (SID) and Viterbi resegmentation using Gaussian mixture models (GMMs). We repeat these multistage segmentation and clustering steps twice: once with MFCCs as feature parameters for the GMMs used in gender labeling, SID and Viterbi re-segmentation steps, and another time with Gaussianized MFCCs as feature parameters for the GMMs used in these three steps. The resulting clusters from the parallel runs are combined in a novel way that leads to a significant reduction in the diarization error rate (DER). On a development set containing 30 telephone conversations, this combination step reduced the DER by 20%. On another test set containing 30 telephone conversations, this step reduced the DER by 13%. The best error rate we have achieved is 6.7% on the development set, and 9.0% on the test set.

Keywords : speaker diarization, speaker segmentation and clustering, BIC clustering, SID clustering.