Abstract

This paper describes a technique for unsupervised audio segmentation. Main objective of the work presented in this paper is to study the performance of audio segmentation system using metric-based method. The system first classifies the audio signal into speech and nonspeech signal using variance of zero crossing rate. The feature Line spectral pair is used for automatically detecting the speaker change point. Hotelling T2 distance metric is used in the
first stage for coarse speaker change detection. The Bayesian information criterion (BIC) is used in the second stage to validate the potential speaker change point detected by the coarse segmentation procedure to reduce the false alarm rate. Database of four files containing the speech recorded from different combinations of male and female speakers mixed with nonspeech signal such as music and environmental sound are used for segmentation. The database-file with one male and one female gives the best performance with F1 measure of 0.9474.

References


**Index Terms**

Computer Science

Engineering and Technology

**Keywords**

Speaker segmentation  LSP  audio segmentation  VZCR