Abstract

The objective is to model the dominating speaker-specific source in the time-domain at different levels, namely, Subsegmental, segmental and supra-segmental. The speaker-specific source information contained in the LP residual. Hence, LP residual contains different speaker-specific information at different levels. At each level features are extracted using proposed method called Hidden Markov models (HMM) and it is compared with existing Gaussian Mixture model (GMM). The experimental results demonstrates that the performance of Subsegmental level is more than the other two levels. However, the evidences from all the three levels of processing seem to be different and combine well to provide improved performance than the state-of –art speaker recognition system and demonstrating different speaker information captured at each level of processing. Finally, the combined evidence from all the three levels of processing together with vocal tract information further improves the speaker recognition performance. Experiments were conducted on TIMIT database using Gaussian Mixture Models (GMM's) and Hidden Markov models (HMM's). Comparing both results the proposed model HMM is better than the existing model GMM.
- Huang, W., Chao, J., & Zhang, Y. (2008). Combination of pitch and MFCC, GMM super vectors for speaker verification. In IEEE int. conf. on audio, language and image process (ICALIP) (pp. 1335–1339).
- Nist speaker recognition evaluation plan (2003). In: *Proc. NIST speaker recognition*
workshop, College Park, MD.


Index Terms

Computer Science
Pattern Recognition

Keywords

Subsegmental segmental suprasegmental LP residual Hidden Markov models
(HMM)