Abstract

This paper describes challenging task of closed set text-independent speaker identification in emotional and whispered speech environments. In the first phase of the work, speaker identification system is developed using neutral speech and tested using speech samples comprising of six basic emotions of anger, happiness, sadness, disgust, neutral and fear. The performance is analyzed using Mel frequency cepstral coefficients (MFCC), Line spectral frequencies (LSF), and temporal energy of subband cepstral coefficients (TESBCC) feature sets. The second phase of work involves the process of speaker identification system in whispered speech environment. The performance of the speaker identification system degrades drastically for whisper speech utterances. A new feature called temporal Teager energy based subband cepstral coefficients (TTESBCC) is proposed. The comparison of the performance of MFCC, TESBCC, weighted instantaneous frequency (WIF) and TTESBCC feature sets is done for this process. A novel classifiers fusion technique is developed and its performance is compared with that of the individual classifiers. Two databases with speech utterances of thirty
nine speakers recorded in the six basic emotions and speech utterances of twenty five speakers in whispered speech are used for experimentation. The speech utterances for database were recorded in Indian language –Marathi. It is observed fusion of classifiers considerably enhances the speaker identification accuracy in both emotional and whispered speech environments.

References


Index Terms

Computer Science | Signal Processing
Keywords

Speaker identification, whispered speech, temporal Teager energy based subband cepstral coefficients, emotional environment, classifier fusion.