ABSTRACT

Keywords: Automatic speaker recognition (ASR), speaker identification and classification, language identification, data collection and corpora design, Linear prediction cepstral coefficients, Teager energy operator, Teager energy base Mel frequency cepstral coefficients, wavelets and wavelet packets, spline wavelet design, Subband based cepstral coefficients, Wavelet packet cepstral coefficients, polynomial classifier, discriminative training, monolingual, cross lingual and multilingual experiments, spectral resolution in female speech, mimic resistance, effects of coding on ASR performance.

This thesis is concerned with building up speaker recognition systems in Indian languages for tape recorded speech and improving their performance with emphasis on system features. First the baseline ASR system using LP based features (such as LPC and LPCC) and filterbank based features such as MFCC with polynomial classifiers of 2^{nd} or 3^{rd} approximation is built for speaker modeling. The details of experimental setup such as dialectal zones (for Marathi, Hindi, Urdu and Oriya) selected for data collection, corpora design and text material used for recordings in different languages are given. Relative comparison of experiments on speaker identification for monolingual, cross lingual and multilingual modes is made.

The spectral resolution problem associated with female speech is resolved to a large extent employing filterbank based features. The problem of speaker classification and language identification is identified from the standpoint of ASR; and the solution to this problem is accomplished by modifying the structure of polynomial classifier. The work on speaker classification is first supported by spectrogram analysis of voices from rural males followed by experimental results for open set and closed set mode for different Indian languages. For speaker classification, wavelet packet cepstrum and subband cepstrum are employed and the performances have been compared with the performance of MFCC.

Mimic resistance of the system is also evaluated for speaker identification of identical twins and professional mimics. A new performance measure viz. sub-optimal success rate is proposed to take care of misidentification between the twins. A new data fusion technique viz. majority rule is proposed to combine the evidences from different feature sets. Finally, effect of different speech coding standards on the performance of ASR is investigated.