Abstract

This thesis documents our investigations on feature analysis and design of compensators for speaker recognition under stressed conditions. Any condition that causes a speaker to vary his or her speech production from normal or neutral condition is called stressed speech condition. Stressed speech is induced by emotion, high workload, sleep deprivation, frustration and environmental noise. In stressed condition, the characteristics of speech signal are different from that of normal or neutral condition. Due to changes in speech signal characteristics, the performance of speaker recognition system may degrade under stressed speech conditions. In this work, the problem of speaker recognition under stressed condition has been dealt with three broad approaches.

First, six speech features (mel-frequency cepstral coefficients (MFCC), linear prediction (LP) coefficients, linear prediction cepstral coefficients (LPCC), reflection coefficients (RC), arc-sin reflection coefficients (ARC) and log-area ratios (LAR)), which are widely used for speaker recognition, are analyzed for evaluation of their characteristics under stressed condition. Statistical techniques such as Probability density function (pdf) characteristics, F-ratio test and Kolmogorov-Smirnov (K-S) test are used for evaluation of speaker discrimination capability of the features under stressed condition. Two classifiers, Vector Quantization (VQ) classifier and Gaussian Mixture model (GMM), are used to evaluate the speaker recognition results with different speech features. Relative ranking of the features based on their performance in speaker recognition under stressed condition is evaluated. This analysis help select the best feature set for speaker recognition under stressed condition.

Secondly, sinusoidal model is used for analysis of speaker recognition system under stressed condition. Sinusoidal model has three different parameters or features. These are amplitude, frequency and phase features. With the assumption that these three features may carry different information of a speaker, the performance of the system can be improved by suitably combining the three features at the classifier stage. In this work, we propose evaluation of confusion probability from confusion matrix for analysis of independent information with sinusoidal amplitude feature (SAF) and sinusoidal frequency feature (SFF). Results show that there may be considerable amount of independent information between the sinusoidal amplitude feature (SAF) and sinusoidal frequency feature (SFF). Improvement in speaker verification is observed by combining the amplitude and the frequency features.

Finally, four novel compensation techniques are proposed and evaluated for improvement of speaker recognition under stressed condition. The compensation techniques are speaker and stressed informa-

tion based compensation (SSIC), compensation by removal of stressed vectors (CRSV), cepstral mean normalization (CMN) and combination of MFCC and sinusoidal amplitude (CMSA) features. Compensation techniques improve speaker recognition results for speech signals under stressed conditions. Speech data from SUSAS database corresponding to four different emotions or stressed conditions, Angry, Lombard, Question and Neutral, are used for analysis of speaker recognition under stressed condition.