# Abstract

Over the last three decades, *mel frequency cepstral coefficients* (MFCC) are used for short-term speech feature extraction in wide range of speech processing applications. Specifically, in *speaker recognition* (SR) technology, MFCC is extensively used due to its considerable performance in both clean and degraded conditions. During these years, significant research has been carried out to find better feature extraction techniques which lead to further improvement in SR performance. In selected cases, the investigated features have performed better than MFCC, but under certain specific conditions. However, the performance obtained using MFCC is almost unparalleled, if different variabilities in training and testing phase are considered. The success of MFCC is due to its inherent characteristics of mimicking different early stages of human auditory perception during its computation process. In this thesis, the focus is on investigating new short-term spectral feature extraction techniques that capture additional useful information while retaining the advantages of MFCC. Here, different techniques are explored to extract relevant information from *mel filter bank log energies* (MFLE) for speaker characterization.

First, motivated by the advantages of multi-band processing, *block transformation* (BT) is investigated for efficient computation of speech features. Two kinds of transformation namely *non-overlapped block transform* (NOBT) and *overlapped block transform* (OBT) are developed where the corresponding features are called as *non-overlapped block transform coefficients* (NOBTC) and *overlapped block transform coefficients* (OBTC). Subsequently, the block sizes are chosen in such a way that the formant frequency zones are processed in an isolated manner, and as a consequence optimum performance is obtained. The proposed features are also uncorrelated as required in diagonal covariance based *Gaussian mixture modeling* (GMM). OBTC with optimal subband partitioning consistently outperforms classical full-band *discrete cosine transform* (DCT) based MFCC in clean and noisy conditions.

Next, two novel feature extraction techniques are studied that represent relative information of subband energies. The relative information are separately extracted using two new linear transformation techniques called as *differential transform* (DT) and *symmetric differential transform* (SDT). The extracted features are referred here as *differential transform coefficients* (DTC) and *symmetric differential transform coefficients* (SDTC). Good SR performance has been achieved with these features. It is found that the relative information based feature conveys complementary characteristics to MFCC or newly investigated OBTC. Thereafter, strength of both the techniques are combined using classifier fusion method to achieve improved recognition performance.

In another work, a novel feature called *temporal spread cepstral coefficient* (TSCC) is investigated from the standard deviation of the subband energies in temporal domain that captures dynamic characteristics of vocal tract. Performance shown by this fea-

ture is comparable to the performance obtained using MFCC. Since this feature conveys complementary temporal information, recognition performance is improved further by combining SR systems of TSCC features and other spectral features. Finally, it is observed that enhanced performance can be achieved with the fusion of proposed features based on block transform (i.e. OBTC), relative information (i.e. DTC or SDTC) and temporal information (i.e. TSCC).

The proposed schemes are evaluated on multiple NIST speaker recognition evaluation (SRE) corpora. To observe the noise-robustness of the proposed features, experiments are also carried out on simulated noisy environment created using NOISEX-92 database. Detail analysis of the noise susceptibility of the proposed schemes have been carried out under different noisy conditions. The newly investigated features based on alternative transformation of DCT are found to be more robust in most cases. In order to get improved SR performance for fused mode in noisy conditions, a new speech quality measure based fusion technique is proposed. This approach gives considerable improvement over existing speech quality measure based fusion technique. Finally, the proposed techniques are validated for state-of-the-art $i$-vector or *total variability* based SR system. The experimental results on large NIST corpora confirm the superiority of the proposed techniques.

**Keywords:** Block Transform, Cepstral Coefficient, Classifier Fusion, Environmental Noise, Feature Extraction, Gaussian Mixture Model (GMM), Mel Frequency Cepstral Coefficients (MFCC), NIST, NOISEX-92, Relative Information, Speaker Recognition, Speech Quality Measure, Temporal Information, Total Variability System.