CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION TO SPEAKER VERIFICATION

Research on speaker recognition, including identification and verification, has been an active area for several decades. Speaker recognition refers to recognizing persons from their voice. The goal is to have a machine that can automatically identify a particular person or verify a person’s claimed identity from his/her voice. Each person has his/her own characteristics and manner of speaking. Also, physically, vocal tract shapes and larynx sizes are different.

Speaker recognition includes speaker identification and speaker verification. Speaker identification search for a person in the database which requires one to many comparisons. Speaker verification checks whether the claimed identity is true or not. This in turn requires one to one comparison. As one of the techniques in biometrics, speaker recognition can be used in many access control applications such as network security, phone transactions, room access, etc. Speaker Verification using the unique characteristics of the speaker provides good security against unauthorised access to important information. Examples include telephone-based applications like voice dialing; mobile banking and teleshopping where users access a service remotely from any location; and on-site applications like access control of a system and access control of a secured area or any material where users are physically present at a point of entry.

On-site applications of speaker verification system may offer much freedom in the choice of data collection equipment, the design of user
interfaces, and means of controlling user behavior and the background environment. This facilitates the collection of high-quality speech samples under controlled conditions to allow for accurate speaker verification. But with physically present users, there are also other biometric technologies to choose from, such as fingerprint sensors, face detectors and iris scanners (Jain et al 2006).

With telephone-based applications, on the other hand, speaker verification has the advantage over other biometric technologies. In that a microphone and a speech transmission channel are already available in every telephone, while using other biometric traits than voice would require attaching additional sensor devices to telephone instruments. But designers of speaker verification systems have little possibility to influence what telephone instruments are used, resulting in varying input signal characteristics between users and possibly even between calls by a single user. They also have little possibility to control users or their acoustic environment, and user interfaces are often limited to speech, audio and button type interfaces. Therefore, compared to on-site applications, telephone-based applications of speaker verification generally need to deal with a wider range of inter-session variability not directly related to speakers' voice characteristics. Both identification and verification can be classified into text-independent and text-dependent applications based on whether or not the person is required to speak pre-determined words or sentences.

1.2 OBJECTIVE OF THESIS

The present thesis deals with several different topics in text-independent automatic speaker verification system. The objective is to develop a SVS system that achieves high verification accuracy with low storage requirements. This proposed system combines the advantages of different modelling methods to improve the performance of text-independent
speaker verification systems. It requires two groups of speakers, the enrolled target speakers and the non target speakers or background speakers.

The quest for the better speech parameterization leads to various features which are extracted from speech. However, only some of the features are important. An ideal feature should (Rose & Reynolds 1990, Wolf 1972).

- Have large inter speaker variability and small intra speaker variability.
- Be easily extractable from the speech signal
- Have invariable speaker characteristics over long time
- Not be susceptible to mimicry by imposters
- Be robust to noise and distortion

Mainly, we have concentrated in two processes involved in the speaker verification system. The first one is feature extraction and the selection of best feature or best combination of various features which provides good performance. The second one is modelling the speaker using different soft computing techniques and statistical methods.

Front-end or feature extractor is the first component in an automatic speaker verification system. Feature extraction transforms the raw speech signal into a compact but effective representation that is more stable and discriminative than the original signal. Since the front-end is the first component in the chain, the quality of the later components such as speaker modelling and pattern matching is largely determined by the quality of the front-end. In other words, classification can be at most as accurate as the features. The dimension of the feature vector and the number of features determines the computation speed of the system. Several feature extraction
methods have been proposed, and successfully exploited in the speaker verification task. Based on physical interpretation, features are classified into (i) voice source features (ii) short-term spectral features (iii) spectro temporal features (iv) prosodic features and (v) high level features. The selection of feature depends on the application, best representation of speech or speaker characteristics and complexity in extracting it from speech. Among these features, short term spectral features are easy to extract and provides good performance in the speaker verification system. Hence, in this thesis, our attention is on the short term spectral features like LPCC (Linear Predictive Cepstral Coefficients), MFCC (Mel Frequency Cepstral Coefficients), PLP (Perceptual Linear Predictive Coefficients), RASTA-PLP (RelAtive SpecTrAl Perceptual Linear Predictive Coefficients), Wavelet based residual coefficients and scalogram based features. Initially, the common and famous LPCC and MFCC are used for the analysis of speaker verification system. Then the modification is done in the extraction of RASTA-PLP in the mel-scale frequency using melfilters. Similarly, the wavelet based residual coefficients also provides greater performance than the MFCC based system.

1.3 THESIS CONTRIBUTIONS

The thesis investigates the advantages and drawbacks of the existing methodologies of text-independent speaker verification, and proposes methods that could lead to an improved performance.

The major contributions of the thesis are:

- A new method of deriving the Gaussian mixture model (GMM) parameters using Learning Vector Quantization (LVQ) algorithm is proposed. The LVQ based GMM training showed a significant improvement of the equal error rate and
higher convergence rate when compared to the classical EM based GMM training.

- The improvement in the learning process of LVQ based GMM is achieved by introducing Fuzzy and Wavelet membership functions in the training algorithm. The proposed method of training the parameters of GMM using Fuzzy LVQ (FLVQ) and Wavelet Fuzzy LVQ (WLVQ) for speaker verification is investigated.

- At the classification level, the use of fuzzy membership function in Support vector machine (SVM) using GMM supervectors is proposed and is found to perform comparably to GMM-SVM. The performance of the system is analyzed with various kernel functions in SVM and various fuzzy membership functions in Fuzzy SVM and compared with GMM-UBM (GMM-Universal Background Model) based system.

- Wavelet Neural Network (WNN) and Adaptive Fuzzy Wavelet Network (AFWN) based speaker verification have also been implemented and their performance has been compared with conventional Artificial Neural Networks such as BPN, SOFM(Self Organizing Feature Map) and RBFN(Radial Basis Functional Network)
1.4 OUTLINE OF THE THESIS

Chapter 1 introduces speaker verification system and the contributions of the thesis.

Chapter 2 includes an overview of speaker verification system with its applications and explains the state of the art technique used in speaker verification system.

Chapter 3 describes a text-independent speaker verification system using GMM-UBM. In this work, the best features out of individual and combination of various features are selected using Genetic algorithm.

Chapter 4 describes Adaptive Wavelet Neuro Fuzzy Inference System (AWNFIS) based speaker verification system. It includes analysis and the improvements in the performance of speaker verification system when wavelet membership function is used in ANFIS systems.

Chapter 5 presents the proposed method of modeling the speaker using AFWN and WNN. In this work, the performance of these systems have compared with the Neural network trained with DWT feature and DWT-MFCC.

Chapter 6 deals with various Vector Quantization (VQ) algorithms used in the training process of GMM. It explains the method of reducing the computational complexity of EM algorithm for GMM and improving the performance of speaker verification system. Also, the comparison between various algorithms is discussed in this chapter.

Chapter 7 deals with Fuzzy LVQ and Wavelet LVQ algorithms used for training the GMM. It improves the performance of speaker verification system because of the use of membership values of wavelet basis functions.
Also, the comparison between proposed method and GMM-UBM with EM algorithms is discussed in this chapter.

Chapter 8 presents the current state of the art technique, SVM- with super vectors of GMM in modeling the speaker. Our contribution in this work is usage of fuzzy SVM for modeling and analysis of system performance with various membership functions and also wavelet function as the membership function in the fuzzy SVM.

Chapter 9 presents in detail the results and analysis of speaker verification system using two speech corpora. Work on experiment design related to those two speaker verification corpora is described. Conclusion of the thesis with a summary of the contributions of this research is also included.