Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 804 – 811

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Robust Speaker Identification Incorporating High Frequency Features

Latha*

Amrita Vishwa Vidyapeetham, Amrita University, India

Abstract

Speaker identification system identifies the person by his/her speech sample. Speaker Identification (SI) system should posses a robust feature extraction unit and a good classifier. Mel frequency cepstral coefficient (MFCC) is very old feature extraction scheme, which has been regarded as standard set of feature vectors for speaker identification. The mel filter bank used in MFCC method, captures the speaker information more effectively in lower frequencies than higher frequencies. Hence high frequency region characteristics are lost. This problem is solved in the proposed method. The speech signal comprises both voiced and unvoiced segments. The voiced segment includes high energy, low frequency components and unvoiced segment includes low energy, high frequency components. In proposed method, the speech sample is divided into voiced and unvoiced segments. The voiced speech segment is filtered using mel filter bank to generate MFCC from lower frequencies of speech signal and unvoiced speech segment is filtered using inverted mel filter bank to generate IMFCC from higher frequencies of speech signal.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: IMFCC; Inverted Mel Filter Bank; MFCC; Short Time Energy; Voiced and Unvoiced Speech.

1. Introduction

Speaker identification (SI) is the important field of research from past many years. Speaker identification system identifies the person by his/her speech sample. There are two research topics in this field. They are feature extraction and feature matching. A basic speaker identification system is shown in Fig. 1. This system has two different stages. Registration or training is the first stage and the second stage is testing stage. In the training stage, each speaker wish to register has to provide samples of his/her speech to prepare a reference model of all registered speakers. In testing stage, the input speech signal of the speaker claiming the identity, is used to extract the feature and is matched with stored features to get the identification result.

The speaker identification is closed set[1], when it is known that all speakers of interest are included in the speaker model, and is open set when some unknown speaker is not the part of the speaker model. Most of the speaker identification systems are open set. The speaker identification is text dependent[2] if there is constraint on the utterance of the speaker, and it is text independent[3] if there is no restriction on the spoken word.

*Corresponding author. Tel.: +91 -9449107045.
E-mail address: s_latha@blr.amrita.edu

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.064

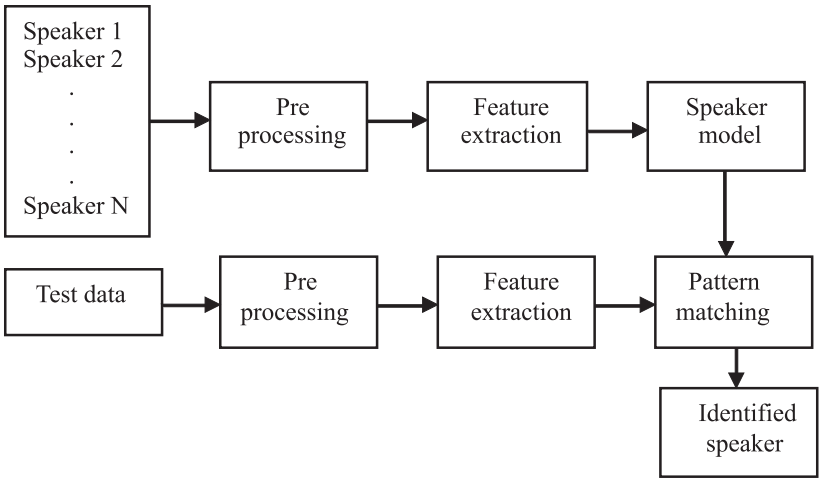

Fig. 1. Speaker Identification System.

1.1 Literature survey

The various feature extraction methods used by many authors include very old basic methods based on spectral averages[4], Pitch, Formats, Linear Predictive Coefficients (LPC)[5], Linear Predictive Cepstral Coefficients (LPCC)[6], Real Cepstral Coefficient (RCC)[6], Mel-Frequency Cepstral Coefficients (MFCC)[6], Perceptual Linear Predictive Coding (PLPC) and Linear Frequency Cepstral Coefficients (LFCC) etc. D.A.Reynolds made a detailed comparison of the feature extraction methods such as PLPC, MFCC, LPCC and LFCC[7] LFCC performance is very poor as it gives equal importance for all frequencies which resulted in increased redundancy. LPCC and PLPC performed better with increased filter orders where as MFCC performance is better with lower order filters also. Based on the literature, it is observed that MFCC outperformed all other methods.

Most frequently used feature matching techniques are Hidden Markov Model (HMM)[2], Gaussian Mixture Model (GMM)[3], Vector Quantization (VQ)[8] and Dynamic Time warping (DTW)[9]. HMM and DTW are the feature matching techniques mostly used for text dependent speaker identification. Different clustering techniques are compared[10] and is observed that the identification accuracy is improved with increased code book size.

1.2 Speaker specific features

In speaker identification the role of parametric representation of the speech signal which is effective in representing speaker specific characteristic is very important in the whole process of identifying the speaker. The identification accuracy is mainly influenced by speaker specific features. If features are more appropriate then the accuracy is high. But selection and extraction of speech features is not an easy process. For an SI system, speech features should occur frequently and naturally in speech signal, should be easy to extract and measure. The robust features are not affected by physical health conditions of the speaker and ambient noise. As we have seen from the literature, frequently used parameters and parametric representations of the speech signal are Pitch, Formats, Linear Predictive Coefficients (LPC), Mel-Frequency Cepstral Coefficients (MFCC), Real Cepstral Coefficient (RCC) and Linear Predictive Cepstral Coefficients (LPCC) etc.

The pitch (reciprocal of pitch period), represents periodicity of relaxation oscillations of the vocal folds of the speaker which intern depends on size and thickness of the vocal folds. Therefore combination of pitch along with other parametric speech features can make a set of robust feature of a speaker.

In Linear predictive coding (LPC) the vocal tract is modeled as an all pole filter (LTI system). The linear prediction method is most accurate method for estimating the parameters which characterize the vocal tract LTI system[11].