Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 2, Number 12; July-September, 2015 pp. 50-53

http://www.krishisanskriti.org/acsit.html

Suitable Feature Extraction Technique for Hindi Digit Speech Recognition

Megha Yadav[1], Usha Sharma[2] and A.N Mishra[3]

[1]M.tech Student, Department of Electronics and Communication Engineering, Krishna Engineering College, Ghaziabad, U.P
[2]Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand,
[3]Department of Electronics and Communication Engineering, Krishna Engineering College, Ghaziabad, U.P
E-mail: [1]meghayadav038@gmail.com, [2]ushasharma1529@gmail.com, [3]an_mishra53@rediffmail.com

Abstract—*Automatic speech recognition (ASR) has made great strides with the development of digital signal processing hardware and software. But despite of all these advances, machines cannot match the performance of their counterparts in terms of accuracy and speed, specially in case of speaker independent speech recognition. So today significant portion of speech recognition research is focussed on speaker independent speech recognition problem. The reasons are its wide applications which are user-friendly for the convenience for the general public. Although many interactive software applications are available, the user of these applications are limited due to language barriers. Hence development of speech recognition system in local language will help anyone to make use of this technological advancement. In this paper, we mainly focused on Gammatone frequency Cepstral coefficient that are derived by applying a cepstral analysis on a Gammatone filterbank responses. This paper evaluates comparative recognition performance of Linear Prediction Coefficients (LPC), Predictive Linear Precoding (PLP) and Mel Frequency Cepstral Coefficient (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC). These features have been tested for speaker independent isolated Hindi digits recognition. Our evaluations show that the GFCC feature performs considerably better than conventional acoustic features.*

Keywords*: ASR, GFCC, LPC, MFCC, PLP, Hindi digit recognition.*

1. INTRODUCTION

In everyday listening conditions, the acoustic input reaching our ears is often a mixture of multiple concurrent sound sources. While human listeners are able to segregate and recognize a target signal under such conditions, robust automatic speech recognition remains a challenging problem. A major challenge for automatic speech recognition (ASR) relates to significant performance reduction in noisy environments. Plenty of research has been focused on robust speech recognition to mitigate this problem.

To tackle this robustness problem, speech enhance-ement methods, such as spectral subtraction, have been utilized for robust speech recognition. These methods tend to perform well when noise is stationary. Hence, they have limited ability to handle novel interferences.

On the contrary, human listeners are capable of recognizing speech when input signals are corrupted by noise. The human ability in these complex acoustic environments is accounted for by a perceptual process called auditory scene analysis (ASA).

In this paper, we evaluate the various feature extraction based speech recognition methods for Hindi digits. Our work mainly focuses on speech recognition of Hindi digits between 0(shoonya) to 9(nau). Along this we have also studied that between feature extraction techniques such as linear predictive Coding (LPC), Predictive Linear Pre-coding (PLP), Mel Frequency Cepstral Coefficient(MFCC) and Gamm-atone Frequency Cpestral Coefficient(GFCC), GFCC performs better in noisy environment.

2. FEATURE EXTRACTION

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (much data, but not much information) then using certain techniques the input set of features is reduced. Here LPC, PLP, MFCC and GFCC based feature extraction techniques are used for feature extraction.

3. SPEECH FEATURE EXTRACTION TECHNIQUES

3.1 Linear Predictive Coding (LPC)

The basic idea behind the LPC analysis is that a speech sample can be approximated as a linear combination of past speech samples. To compute features, the first step of all the three Feature Extraction technique is pre-processing which was applied on all the speech samples. The major steps in pre-processing are pre-emphasis filtering and normalization. The digitized speech signal is passed through first-order FIR filter to spectrally flatten the signal and to make it less susceptible to finite precision effects. To compensate for the variation in the amplitudes of different speech samples of same digit, all

the speech samples are divided by the sample with highest amplitude. Each sample is multiplied by an N-sample Hamming window, and this windowed frame is passed to perform short term auto correlation.

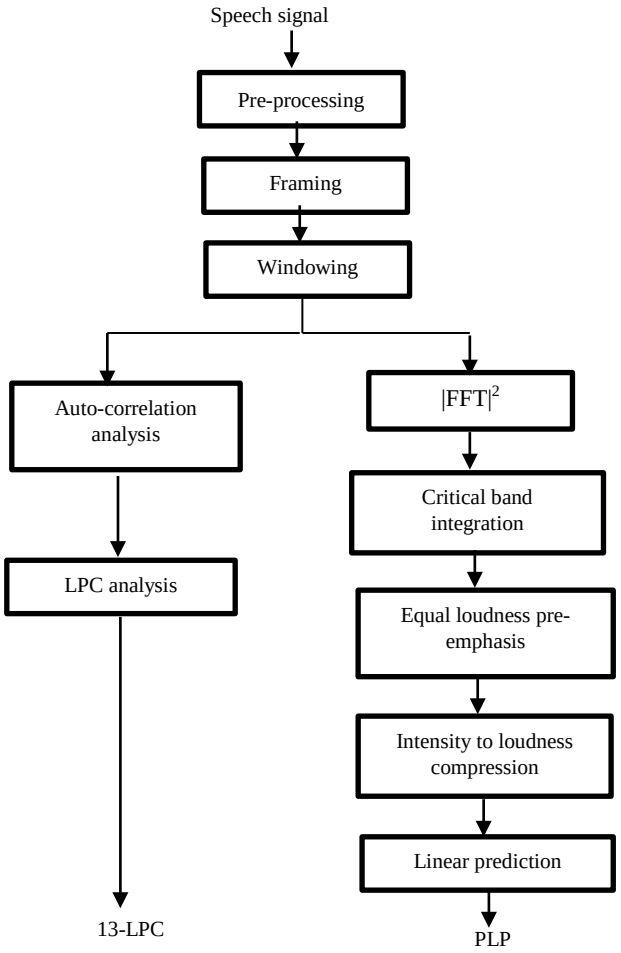

Fig. 1: Feature extraction using LPC and PLP

The highest autocorrelation value is the order of LPC analysis, i.e., p which has been taken 13. First thirteen LPC coefficients were chosen for each frame. Finally thirteen LPC coefficients were selected for each sample of Hindi words by applying vector quantization on features of all frames of each sample of each Hindi digit

3.2 Predictive Linear Precoding

PLP inherits its characteristics from LPC as well as MFCC. PLP features can be obtained by following the steps as shown in Figure1. The speech signal is pre-emphasized and normalized similarly as in LPC. The frame duration and overlapping are also taken same as in LPC. The power spectral estimate for the windowed speech signal is computed. The power spectrum is integrated within overlapping critical band filter responses. For obtaining PLP, trapezoidal shaped filters are applied at 1-bark intervals. Restrictions are imposed on higher frequencies of band-pass filter in such a way that pass-band in the domain of critical band modulations is established to a range that appears to be required for speech intelligibility. The elements of critical band spectrum are explicitly weighted. To reduce the effects of amplitude variations for spectral resonances cube root of spectral amplitudes are taken after that IDFT is performed. Durbin's algorithm is used for computing the PLP coefficients. Again 13 features are taken for each speech sample by applying vector quantization on the features of all frames of a sample.

3.3 Mel Frequency Cepstral Coefficient

The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. The Mel frequency scale has linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Fig. 2 depicts the procedure of extracting MFCC feature vectors from speech. The speech is pre-emphasis and normalized by using same procedure used in LPC and PLP. After that, the Fast Fourier transform (FFT) is used to convert each frame of N samples from time domain into frequency domain. Then, Compensation for non- linear perception of frequency is implemented by the bank of traingular band filters with the linear distribution of frequencies along the so called Mel-frequency range. Linear deployment of filters to Mel-frequency axis results in a non-linear distribution for the standard frequency axis in hertz. The Mel spectrum coefficients and their logarithm are real numbers. Hence they can be converted to the time domain using the discrete cosine transform (DCT). The result is the Mel Frequency Cepstral Coefficients.

3.4 Gammatone frequency cepstral coefficient

The most significant behind using this technique is the introduction of asymmetric filters to replace the triangular filters of the Mel filter-bank in previous section. It was studied that these filters better app-roximate the filtering done in the basilar membrane. During implementation of this technique, initial steps such as pre-processing, framing and windowing are similar to previous section. Resultant signal then further passed through gammatone filter bank. Then the Equal loudness is applied to each filtered output. Further, a logarithmic and Discrete Cosine transform is applied to get GFCC features of input signal. GFCC has a fine resolution at low frequency as compared to MFCC. MFCC works with a log while GFCC works with cube root. Cube roots provide more robustness to GFCC as compared to that of logs in MFCC. Hence, GFCC is more robust in noisy environment than MFCC. Similar to previous techniques first 13 coefficient were choosen from each frame.

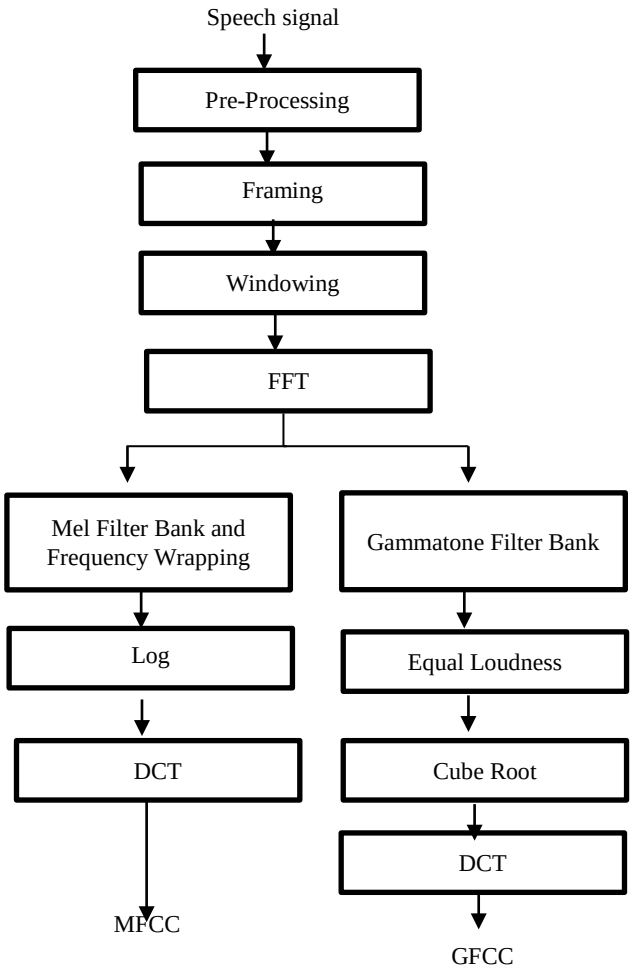

Fig. 2: Block diagram of MFCC and GFCC

4. CLASSIFICATION

After extracting the features and removing irrelevant information, there comes classification or modelling or pattern matching. In our study we have used Linear Discriminant Analysis (LDA), it is a well- known technique in statistical pattern classification for improving discrimination and compressing the information contents (with respect to classification) of a feature vector by a linear transformation.

5. DATABASE

It is a clean isolated Hindi digits database of twenty four speakers. A database of twenty-four speakers, eighteen females and six males for a total of ten Hindi digits ("Shunya", "Ek", "Do", "Teen", "Chaar", "Paanch", "Che", "Saat", "Aath" and "Nau") was prepared with sampling frequency 16 kHz and 16 bits per sample. Speakers were chosen from different geographical areas of India, different social classes and of different age groups (18-27 years). Every speaker was asked to repeat each digit ten times with short inter-digit pauses. Further, all ten repetitions of each digit were segmented manually. The age group of 18-27 years was chosen as students of different dialects in this age group were easily available. A distance of 2-6 inch was maintained between microphone and the speaker at the time of database recording. Two different microphones (Sony make) were used for recording the database.

Table 1: Hindi Digits, English Digits and their Pronounciation

Hindi Digits	Hindi Pronounciation	English Digits	English Pronounciation
०	Shoonya	0	Zero
१	Ek	1	One
२	Do	2	Two
३	Teen	3	Three
४	Chaar	4	Four
५	Paanch	5	Five
६	Che	6	Six
७	Saath	7	Seven
८	Aath	8	Eight
९	Nau	9	Nine

6. EXPERIMENTAL RESULTS

For this work, the analyses are done with speaker independent isolated speech recognition for Hindi digit under Matlab environment. The same speech samples were used for different experiments with different feature extraction and pattern matching algorithms. In this experiment, the dataset is divided into training and testing data where 60% data is given for training and the remaining 40% data are given for testing. The same speaker's data are used for the experiments. The performance of speech recognition systems are mainly specified in terms of accuracy of matching. The robust recognition test involves comparative study with four features: LPC, MFCC, PLP, GFCC. For each feature extraction technique we have obtained different accuracy efficiency as described in chart below for LPC(51.85%), PLP (67.23%), MFCC(64.89%) and GFCC(69.51%).

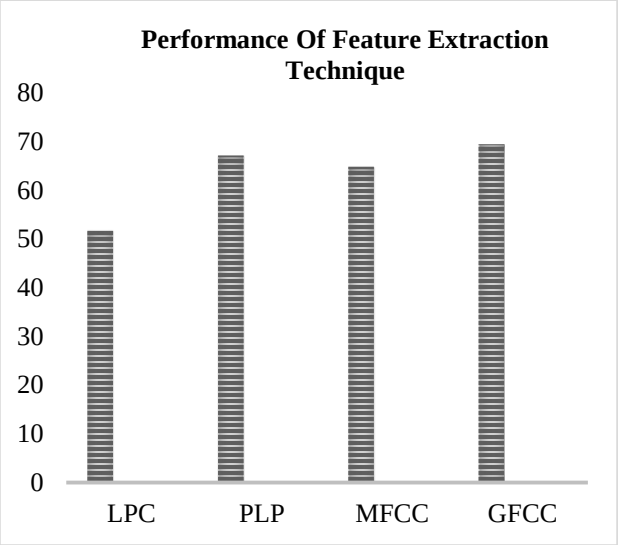

Fig. 3: Performance chart of feature extraction technique

7. CONCLUSION AND FUTURE WORK

The main objective of this research work is to provide a detailed comparative analysis and implementation of the most popular speech feature extraction techniques for speaker independent Hindi digit speech recognition system. The most popular speech feature extraction and pattern matching techniques were implemented and analysed. Totally, four feature extraction algorithms namely MFCC, LPC, PLP and GFCC are implemented and its performances were deeply observed. The potential pattern matching algorithm (LDA) is used for speech recognition. By investigating these feature vectors along with the recognition techniques it was found that the GFCC features gave better results and outperformed the other algorithms for all the speech recognition techniques. Highest word recognition accuracy is achieved with GFCC techniques for both training and testing data. Based on the satisfactory results and metrics of this technique the feature combination method will be proposed in future.

8. ACKNOWLEDGMENT

I wish to express my sincere gratitude to Prof. A. N Mishra for his constant guidance throughout the course of the work and many useful discussions which enabled me to know the subtleties of the subject in proper way.

REFERENCES

[1] Bhupinder Singh, Rupinder Kaur, Nidhi Devgun, Ramandeep Kau, "The process of feature extraction in automatic speech recognition system or computer machine interaction with humans", *IEEE International Journal Of Advance Research In Computer Science And Software Engineering*, Vol. 2, Feburary 2012.

[2] R. Schluter, I. Bezrukov, H. Wagner, H. Ney, "Gam-matone features and feature combination for large vocabulary speech recognition", *IEEE International Conference On Acoustic Speech And Signal Processing*, Vol. 4, April 15-20 2007, pp. 649-652.

[3] Yang Shao, Zhaozhang Jin, DeLiang Wang, " An auditory-based features for robust speech recognition", *IEEE International Conference On Acoustic Speech And Signal Processing*, April 19-24 2009, pp. 4625-4628.

[4] Vimala. C, Radha. V, " Suitable feature extraction and speech recognition technique for isolated Tamil spoken words" *IEEE International Journal Of Computer Science And Information Technologies*, Vol. 5 (1), 2014, pp. 378-383.