Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 2, Issue 22; October-December, 2015, pp. 1896-1899

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

Speech Recognition using Hybrid of GFCC and PLP

Nitasha Gupta[1], A.N Mishra[2] and Usha Sharma[3]

[1]M.tech Student, Department of Electronics and Communication Engineering, Krishna Engineering College, Ghaziabad, U.P
[2]Department of Electronics and Communication Engineering, Krishna Engineering College, Ghaziabad, U.P
[3]Research Scholar, Department of Computer science and Engineering,
Indian School of Mines, Dhanbad, Jharkhand
E-mail: [1]nits12gupta@gmail.com, [2]an_mishra53@rediffmail.com, [3]ushasharma1529@gmail.com

Abstract—*Ever since world began the only way to communicate between humans is language, and the basic medium we use to interact in any language is speech. The speech recognizers make use of a parametric form of a signal to obtain the most important distinguishable features of speech signal for recognition purpose. In this paper we introduce two new techniques which are formed by hybrid feature extraction techniques in Hindi speech recognition through linear discriminant analysis (LDA). Gammatone frequency cepstral coefficients (GFCC), Perceptual linear prediction (PLP) along with their hybrid techniques features for recognition of Hindi isolated has been studied and the corresponding recognition rates are compared. By the combination of these two techniques we have obtained oue two hybrid features named as Bark frequency cepstralcoeficicients (BFCC) and Gammatone perceptual linear prediction (GPLP). The recognition rate obtained by these hybrid feature extraction techniques are better than the conventional feature extraction techniques using LDA.*

Keywords*: BFCC, GFCC, PLP, GPLP, LDA.*

1. INTRODUCTION

The concept of a machine that can recognize the human voice has long an accepted feature in science fiction. Speech recognition is a tool to ease the man-machine interaction. We humans, are able to recognize a speaker's identity when we hear him/her speak, provided that the speaker is known to us that is we have heard enough of his/her speech. While a human system is able to extract the information necessary to identify a speaker under a wide range of conditions, it is significant challenge for a speaker recognition system to extract information from the speech in a meaningful way.

Human perception involves both classification and recognition. The arrival of language is perhaps a good example of the human disposition to classify inherently and recognize patterns. Discovering and recognizing patterns present in the speech is probably the most difficult task in pattern recognition by machine. Speech is the primary communication medium among people. This communication process has a complex structure which consists of not only the transmission of voice but also the gestures, the language, the subject and the capability of the listener. In this respect, the performance of a speech recognizer system heavily depends on how and for which task we designed it. Speech recognition area of science has its roots in the idea of communicating with a machine by voice. Speech can be regarded as an important component to make this communication easier. The ultimate goal of research on automatic speech recognition (ASR) is to build machines that are indistinguishable from humans in the ability to communicate in natural spoken language. In this sense, speech recognition is not a mature science but an emerging one.

The digit recognition task for Hindi language is difficult due to a large number of variability in Hindi dialect. Hindi is a major Indian language belonging to the Indo-European family, which has retro flexion and germination as important feature.

In this research, our work mainly focuses on speech recognition of Hindi digits between 0(shoonya) to 9(nau). The recognition performance of Hindi digits are evaluated and the performances of hybrid feature extraction techniques are compared to conventional feature extraction techniques.

2. FEATURE EXTRACTION

The raw speech signal is complex and may not be suitable for feeding as input to the automatic language identification system; hence the need for a good front-end arises. The task of this front-end is to extract all relevant acoustic information in a compact form compatible with the acoustic models. In other words, the pre-processing should remove all non-relevant information such as background noise and characteristics of the recording device, and encode the remaining (relevant) information in a compact set of features that can be given as input to the classifier. Features can be defined as a minimal unit, which distinguishes maximally close classes. The entire scheme for feature extraction using GFCC, PLP, BFCC and GPLP techniques.

2.1 Gammatone Frequency Cepstral Coefficients(GFCC)

The GFCC is a FFT-based feature extraction technique in speaker identification systems. Fig. 1 depicts the procedure of extracting GFCC feature vectors from speech. Initially, to spectrally flatten the speech signal i.e. to obtain similar amplitude for all frequency components, the speech signal is passed through a pre-emphasis filter, which is a first order FIR digital filter. Speech can be considered to be time invariant over short segments of time. Therefore, speech signal is split into frames of 20ms. Each sample is multiplied by Hamming window, and this windowed signal is passed through gammatone filter bank.

The impulse response of each filter is given by the equation:

$$g_m(t) = t^{n-1}e^{-2\pi b_m t}\cos(2\pi f_{cm}t)$$

where 'n' is the order of the, $'f_{cm}'$ is the center frequency and $'b_m'$ is the attenuation factor of the filter, which is related to the band of the filter, and is decisive factor of impulse response decay rate.

The logarithm is applied to each of the filter output to stimulate the human perceived loudness given certain signal intensity and to separate the excitation (source) produced by the vocal cords and the filter that represents the vocal tract. Since the log-power spectrum is real, Discrete cosine transform (DCT) is applied to the filter outputs which produces highly uncorrelated features and results coefficients being more concentrated at lower indices. The result is gammatone frequency cepstral coefficients. First 13 features are taken for each speech sample by applying vector quantization on the features of all frames of a sample.

2.2 Perceptual Linear Prediction (PLP)

PLP models the human speech based on the concept of psycophysics of hearing. PLP discards irrelevant information of the speech and thus improves speech recognition rate. Fig. 2 depicts the procedure of extracting PLP feature vectors from speech.

Initially, to convert the speech signal from time domain to frequency domain FFT is being used. A frequency wrapping into the Bark scale is applied. There is a conversion from frequency to bark, which is a better representation of the human hearing resolution in frequency. The sampled speech signal is pre-emphasized by the simulated equal-loudness curve which gives non equal sensitivity of human hearing at different frequencies and simulates the sensitivity of hearing at about 40-dB level. Intensity-loudness power law is an approximation to the power law of hearing and simulates the non linear relation between the intensity of sound and its perceived loudness. After that inverse FFT of the auditory results in the autocorrelation coefficients of the speech. The PLP coefficients are then obtained using the Levinson-Durbin recursion algorithm. Again 13 features are taken for each speech sample by applying vector quantization on the features of all frames of a sample.

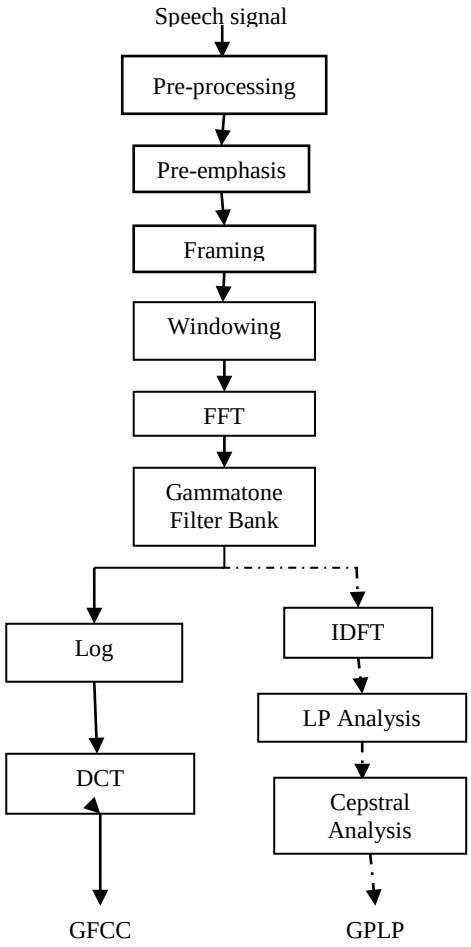

Fig. 1: Feature extraction using GFCC and GPLP

3. HYBRID FEATURES

In this experiment two main blocks as shown in Fig. 1 and 2 were interchanged to develop two hybrid feature extraction techniques. The interest is to see the influence of the spectral processing on the different cepstral transformation. The Fig. 1 and 2 shows the steps of parameterization for the basic method and besides PLP and GFCC the way of computing the hybrid techniques has been shown by dashed arrow in figures.

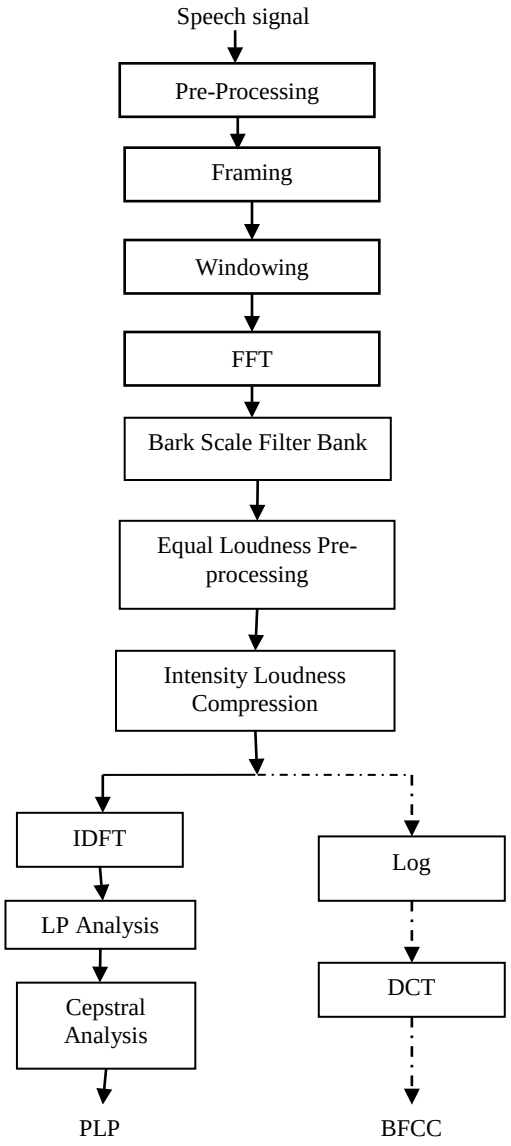

Fig. 2: Block diagram of PLP and BFCC

3.1 Gammatone Perceptual Linear Prediction(GPLP)

In this approach instead of using Bark filter bank, Gammatone filter bank has been applied to compute GPLP. The signal is pre-emphasized before the segmentation and FFT spectrum is processed by Gammatone scale filter bank. The resulting spectrum is converted to the cepstral coefficients using LP analysis with prediction order of 13 followed by cepstral analysis.

3.2 Bark Frequency Cepstral Coefficient (BFCC)

BFCC is the process where we combine PLP processing of the spectra and cosine transform to get the cepstral coefficients. Instead of using Gammatone filter bank, Bark filter bank has been applied and equal loudness pre-emphasis with intensity to loudness power law has been applied to the GFCC like features. Only first 13 cepstral features of each windowed frame of speech utterances were taken.

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 (withrespect 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 groupwere 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 RESULT

This section presents the experimental evaluation of GFCC, PLP, GPLP, BFCC features for speaker independent speech recognition. In this experiment the data is divided into training

and testing data where 60% data is given for training and 40% data is given for testing. The same speaker's data are used for experiment. The performance of this speech recognition systems are mainly specified in terms of accuracy of matching. Using these feature extraction techniques we conducted speaker independent speech recognition on testing set with the result shown in the chart. We can see in clean testing condition GFCC features (65.51%) generated comparable result to PLP features (67.23%) and hybrid techniques as GPLP features (69.42%) generate comparable result to BFCC features (61.107%).

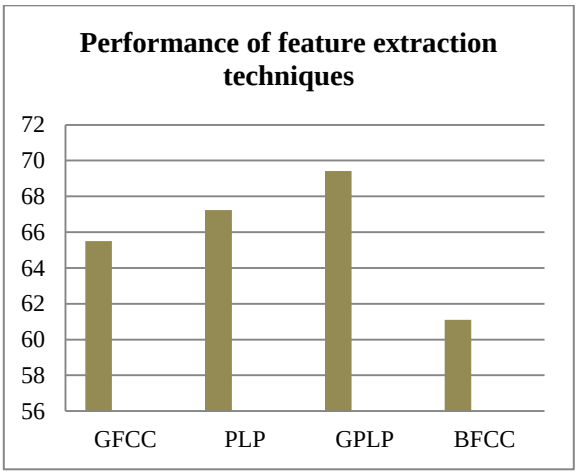

Fig. 3: Performance of GFCC, PLP, GPLP, BFCC

7. CONCLUSION

In this work four different feature extraction techniques are used for speech recognition. So, the language identification system provides satisfactory results by these four feature extraction techniques, GFCC, PLP, BFCC, GPLP with LDA classifier. BFCC has shown less identification performance as compared to GFCC and also shown the least performance among all the techniques. GPLP shown much better performance as compared to the PLP because it is more invariant to fixed spectral distortion and channel noise.

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 subtitles of the subject in proper way.

REFERENCES

[1] P.K.Sahu, Anirban Bhowmick, Mahesh Chandra,"Hindi vowel classification using GFCC and formant analysis in sensor mismatch condition",*WSEAS Transactions on system*, vol 13, 2014, pp. 130-143.

[2] Sharmila, Dr.Achyuta N. Mishra, Dr.NeetaAwasthy, " Hybrid Features for Speaker Independent Hindi Speech Recognition",*International Journal of Scientific & Engineering Research*, Volume 4, Issue 12, December-2013.

[3] R. Schluter, I. Bezrukov, H. Wagner, H. Ney, "Gamma tone features and feature combination for large vocabulary speech recognition", *ICASSP* 2007.

[4] A.N Mishra, Mahesh Chandra, Astikbiswas, S.N Sharan "Hindi phoneme-viseme recognition from continuous speech", *International journal of signal and imaging system engineering*, vol, 6no. 3,2013.

[5] H. Hermansky, "Perceptual linear prediction (PLP) analysis of speech", *Journal of Acoustic Society America*, vol. 87, pp. 1738-1752, 1990.

[6] H. Hermansky and N. Morgan, "RASTA processing of speech", *IEEE Transaction on Speech and Audio Processing*, vol. 2, no. 4, pp. 578-589, 1994.

[7] Vinay Kumar, "Statistical approach towards Hindi language words", HAL *inria-00114544*, version 1, 2006.

[8] K. Samudravijaya, "Hindi speech recognition", *J. Acoustic Society of India*, vol. 29, issue1, pp. 385- 393, 2000.

[9] R. P. Lippman, "Speech recognition by machines and humans", *Speech Communication*, vol. 22, pp. 1–15, 1997.

[10] Y. Gong, "Speech recognition in noisy environments: A survey", *Speech Communication*, vol. 16, no.3, pp. 261–291, 1995.