# Evaluation of Feature Extraction Techniques for Punjabi Language in ASR system

Jyoti Guglani*
*Department of Electronics & Communication,*
*IMS Engineering College, Ghaziabad(U.P.) 201015*
*jyoti.guglani@imsec.ac.in

***Received:*** *30.04.2019,* ***Accepted:*** *29.05.2019*

**Abstract**

This paper compares the performance of Speech recognition systems for Punjabi language. The Linear Predictive Coefficients, Cepstral Coefficients and Wavelet coefficients have been tested for speaker independent Automatic Speech recognition system for Punjabi language.The database is isolated digits of Punjabi language. The recognition performance of isolated speech recognition system with different feature extraction coefficients has been studied. The recognition performance of Speech recognition system with Perceptual Linear Prediction coefficients is better than Linear Prediction Coefficients and Linear Predictive Cepstral coefficient. The recognition performance of LPCC coefficients is 5.2% better than LPC. The recognition performance of PLP is 11.95% better than LPCC coefficient. MFCC coefficient gives 4.76% high performance than PLP. Wavelet Coefficients gives 16.79% high recognition performance than MFCC. Hidden Markov Model (HMM) classifier is used in all recognition systems and programming is done on MATLAB for training and testing of Punjabi digits.

**Keywords—**Linear Prediction Coefficient,Cepstral Coefficient, Wavelet Coefficient, Hidden Markov Model, Speech Recognition, Punjabi language.

## Introduction

Speech is the one mode of communication. It is a one by which we can sharing our thoughts, emotions and also a way of transferring intelligence from one to another person (Allen, 1994). The Speech recognition is the way by which we can convert a speech signal to a set of text. The recognized text can be the further used in various areas like data entry and document preparation data entry and document preparation. To build a speech recognition system has been a goal of researchers for more than four decades but the desired goal is yet to be achieved (Gong, 1995; Sreenivas and Kirnapure, 1996). A speech recognition system is collection algorithms from different disciplines including Mathematics and Computer Sciences. One of the most difficult aspects of performing research in speech recognition by machine is it's inter disciplinary nature.

A speech recognition system consists of three different blocks; first is database preparation block, than the feature extraction and at last the classification block as shown in Figure 1. The development and evaluation of ASR system is the availability of audio database. The LPC, LPCC, PLP, MFCC coefficients and Wavelet features techniques (Makhoul, 1976 ; Atal, 1974; Hasan *et al.*, 2004; Hermansky, 1990;

Krishan et al., 1994) have been used by various researchers for speech recognition with different languages. In this work Hidden Markov Model (HMM) (Levinson et al., 1983) is used for classification. Satisfactory results are obtained by using all the kinds of feature extraction techniques in combination with this classifier for recognition of speaker independent Isolated Punjabi digits.


**Figure 1:** Block diagram of Speech Recognition System

Database preparation and feature extraction techniques are given in below sections of database preparation, feature extraction followed by Experimental setup and results discussion and the last is conclusions.

**Database Preparation**

Cool software' was used for preparation of this database. A database of twenty four speakers, twelve females and twelve males was made for a total of ten Punjabi digits -*'sifar', 'ikk', 'do', 'tinn', 'char', 'panj', chhe', 'satt', 'atth', 'nau'* (i.e. 'zero', 'one', 'two', 'three', 'four', 'five', 'six', 'seven', 'eight', 'nine'). It was prepared with sampling frequency 16 kHz and 8 bits per sample. Speakers were chosen from different geographical areas of Punjab, different social classes and of different age groups. Each speaker repeat each digit ten times with short pauses between the digits. The ten repetitions of each digit were segmented using wave analzer software. This Database for twelve female and twelve male speakers from different regions of Punjab was prepared by using 'Cool software'. The people from different age group is chosen because of convenience of availability. A reasonable distance was maintained between microphone and the speaker's mouth while recording of audio database recording. The microphone of Sony make is used for recording the database.

**Feature Extraction**

The feature extraction is the method to extract the feature vectors from the speech signal. The recorded speech signal is raw and not suitable to work as input to an speech recognition system; hence the need for feature extraction technique arises. The preprocessing of signal should remove all irrelevant information from signal that includes the various noises and characteristics of the recording devices. The output of feature extraction block acts as the input to the classifier block. The entire scheme for feature extraction using LPC, LPCC, PLP, MFCC and Wavelet Coefficients are shown in Figure 2.

*A. Linear Prediction Coefficient*

The Linear prediction technique (Makhoul, 1976) is used to extrat the coefficients by minimizing the mean square error between the input sample and the estimated sample. The auto-correlation method or

 

the covariance method are used for extraction of these features. The transfer function *H(z)* of system is given by the equation:

$$H^{*}z)=\frac{G}{A(z)}=\frac{G}{\left(1+a_1z^{-1}+a_2z^{-2}+,,,+a_pz^{/n}\right)} \quad \ldots(1)$$

The values $a_i$ are the prediction coefficients and *G* are the gain of the vocal tract excitation system.

The Pre-processing of acoustic signal incudes filtering, normalization and mean sustraction. The digitized speech signal is pre-emphasized with transfer function represents by, $H(z)=1-0.98z^{-1}$. By the mismatch between the training and testing conditions of ASR system, it is recommented to reduce the variations in the data that does not carry relevant information about signal. For that purpose the normalization techniques were applied. During the process, every speech sample of the speech signal is divided by the maximum amplitude sample value. To remove the DC offset substact the mean from speech signal.

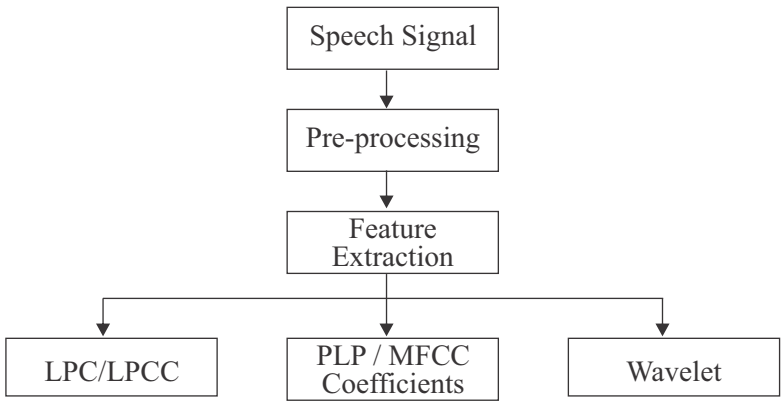


**Figure 2:** Block diagram representation of Feature Extraction techniques

After front end processing of the speech signal, Punjabi digit samples were divided into time frames of 25ms. The second frame starts after time period of 15ms of first frame and overlaps the first frame by time duration of 10ms. This process continue until all the frames of speech samples were addressed. Each speech frame was multiplied by a hamming window and 13$^{th}$ order LPC analysis was performed.

*B. Linear Prediction Cepstral Coefficient*

A very important LPC parameter set, directly derived from LPC coefficients are the cepstral coefficients (Makhoul, 1976; Atal, 1974). The cepstral coefficients are the FT presentation of the log magnitude spectrum. A special feature of cepstral coefficient is orthogonality (Atal, 1974).
The recursion used for LPCC features from LPC is as follow
s:

$$a_0=\ln u^2 \quad \ldots(2)$$

$u^2$ is the gain term in the LPC model.

$$c_m = a_m + \sum_{k=1}^{m-1} (k/m) c_k a_{m-k}, \quad / £ m Å p \qquad ...(3)$$

$$c_m = \sum_{k=1}^{m-1} (k/m) c_k a_{m-k} * \quad m @ p \qquad ...(4)$$

*C. Perceptual Linear Prediction:*

Perceptual Linear Prediction coefficients having the characteristics from LPC as well as MFCC. The PLP features (Hermansky, 1990) can be extracted by the steps shown in Figure 1. The speech signal is pre-emphasized and normalized similarly as in extraction of LPC. The frame duration and overlapping are also taken same as in LPCC feature extraction. The power spectral estimated for the windowed speech signal. The power spectrum is integrated within critical band filter responses which are overlapped. The higher frequencies of band-pass filter is applied in such a way that pass-band in the domain of critical band modulations is established to a range of frequencies that appears to be required for speech signal estimation. To reduce the effects of amplitude variations cube root of spectral amplitudes are extracted. The Inverse Discrete Fourier Transform is applied afterwards. The Durbin's algorithm is used for obtaining 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.

*D. Wavelet Based Feature Extraction*

The preprocessing of speech signal is done intially. In Wavelet based feature extraction technique (Krishnan, 1994) the pre-emphasized signal is decomposed into twenty four bands by applying different wavelets. The composed energy of all wavelet coefficients in each sub-band is obtained. Logarithm of the energy of all the 24 frequency bands is calculated. After that Discrete Cosine Transformation is done as a result the coefficients are arranged in the ascending order of the weightage of the information composed in the coefficients. First 13 coefficients are selected out of 24 DCT coefficients.

**Experimental Setup and Results**

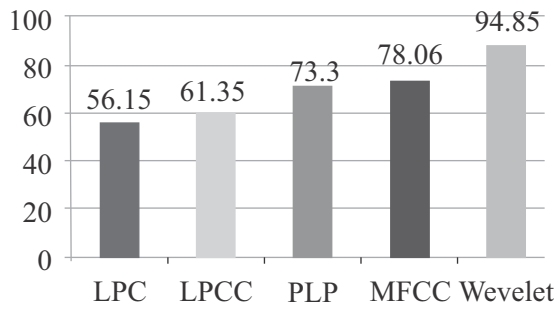


**Figure 3**: Comparative study of Linear prediction based features

By using LPC, LPCC, PLP, MFCC and Wavelet coefficients, 13 features are extracted for each sample of

 

all 24 Punjabi digits. The features of all the twenty four speakers were stored for future processing after extraction. The feature vectors of first twenty speakers were used for the purpose of training and the feature vectors of last four speakers of each digit were used for testing the classifier. Comparative speaker independent ASR system for Punjabi language shows results for the LPC, LPCC, PLP, MFCC and Wavelet Coefficient feature extraction techniques with HMM are shown in Figure 3. The recognition performance of LPCC Coefficients is 5.2% better than LPC while the recognition performance of PLP is further improved by 11.95% than that of LPCC Coefficients. One of the reasons for better performance of LPCC features is that these are cepstral features. MFCC coefficient gives 4.76% better than PLP coefficient. Wavelet coefficient gives 16.79% better than MFCC. The Cepstral features are useful because they operate in a domain in which the excitation function and the vocal tract filter function are separable. Wavelet coefficient gives better efficiency than other techniques.

**Conclusion**

The size of twenty four speakers was selected based on convenience. The system provides satisfactory results for speaker independent cases. The recognition performance with PLP Coefficients features was better than that with LPC and LPCC features. The recognition performance with MFCC features was further improved than that with PLP Coefficients features. Wavelet Coefficient gives best recognition performance among all feature extraction techniques.

**References**

Allen, J. B. 1994. How do humans process and recognize speech. IEEE Transactions on Speech and Audio Process, 2(4), 567–576.

Gong, Y. 1995. Speech recognition in noisy environments: A survey. Speech Communication, 16(3), 261–291.

Sreenivas, T., Kirnapure, P. 1996. Codebook constrained Wiener filtering for speech enhancement. IEEE Transactions on Speech & Audio Process, 4, 383–389.

Makhoul, J. 1976. Linear prediction: A tutorial review, Proc. of IEEE, **63(**4), 561-580.

Atal, B. S. 1974. Effectiveness of Linear Prediction characteristic of the speech wave for automatic speaker identification and verification. Journal of the Acoustical Society of America, 55(6), 1304–1312.

Hasan, M.R., Jamil, M., Saifur Rahman, M.G. 2004. Speaker identification using Mel frequency Cepstral Coefficients. 3rd International conference on Electrical and Computer Engineering, Dhaka, Bangladesh, pp. 565-568.

Hermansky, H. 1990. Perceptual linear prediction (PLP) analysis of speech. Journal of Acoustic Society America 87, 1738-1752.

Krishnan, M., Neophytou, C. P., Prescott, G. 1994. Wavelet Transform Speech Recognition Using Vector Quantization, Dynamic Time Warping and Artificial Neural Networks. International Conference on

Spoken Language Processing, Yokohama, Japan, pp. 1-3.

Levinson, S. E., Rabiner L. R., Sondhi, M. M. 1983. Speaker independent isolated digits recognition using Hidden Markov Model. IEEE International Conference on Acoustics, Speech and Signal Processing- Proceedings (ICASSP), Boston, pp. 1049-1052.

 