PERFORMANCE ANALYSIS OF MFCC AND LPC TECHNIQUES IN KANNADA PHONEME RECOGNITION

[1]Kavya.B.M, [2]Sadashiva.V.Chakrasali
Department of E&C, M.S.Ramaiah institute of technology , Bangalore, India
Email:[1]kavyabm91@gmail.com ,[2]Sadashiva.c@msrit.edu

***Abstract*- Speech is one of the oldest and the most natural used means of information exchange among the human beings. Accurate Phoneme recognition forms the backbone of most successful speech recognition systems. A collection of techniques exists to extract the relevant features from the steady state regions of phonemes both in time and frequency domains. Here we build an automatic phoneme recognition system based on Hidden Markov Model (HMM) which is dynamic modeling scheme. The Mel-Frequency Cepstrum Coefficients (MFCC) and Linear Predictive Coding (LPC) techniques are used for feature extraction. The performance comparison of these two techniques with HMM classification is done to achieve better performance with high recognition rate and low computational complexity. MFCC features with HMM gives the high recognition rate while LPC with HMM is computationally less complex. The major advantage of comparing these two techniques is that they improve reliability of the system. This has been carried out on five Kannada phonemes.**
Keywords: Automatic Speech Recognition, HMM, MFCC, LPC, Kannada.

1. INTRODUCTION

Speech and hearing, is the man's most used means of communication. For centuries people have tried to develop machines that can understand and produce speech as humans do so naturally. Speech recognition can be defined as the process of converting an acoustic signal captured by microphone to set of words. Automatic Speech Recognition (ASR) is one of the fastest developing fields in speech science and engineering. The Automatic Speech Recognition System of any language must be able to recognize spoken sentences, words, syllables, and phonemes of that language. Speech technology is the technology of today and tomorrow with the developing number of methods and tools for better implementation. Speech recognition has a number of practical implementations for both serious and fun or entertainment works. ASR has an interesting and useful implementation in expert systems, a technology whereby computer can act as a substitute for human expert. In country like India where there are so many dialects variation this technology helps in reducing human staff trained in different languages.

To demonstrate these concepts, we have built a database of 5 Kannada phonemes. Each phoneme is recorded 120 times out of which 90 is used for training while 30 for testing with a sampling rate of 8 kHz. Hence total we have 450 phonemes for training and 150 phonemes for testing.

2. METHODOLOGY

I. *Construction of database*

A speaker dependent system is built. All the samples were recorded from native Kannada speakers both for training and testing. Audacity software is used to record phonemes. And they are stored in .WAV format. Details of the database are shown in Table 1.

Sl. No.	Phonemes used in speech recognition	No. of samples taken
1	Short Vowel /a/	120
2	Short Vowel /i/	120
3	Short Vowel /ou/	120
4	Diphthong /ai/	120
5	Diphthong /au/	120

Table 1: Phonemes used in speech recognition

II. *Pre-processing*

Since the recordings were taken under normal conditions with background noise, it is important to remove these noises. And then normalization is done.

III. *Feature extraction*

The goal of feature extraction is to represent any speech signal by a finite number of measures of the signal. Here Mel frequency cepstrum coefficients (MFCC) and linear predictive coding (LPC) coefficients are used for feature extraction.

Mel frequency cepstrum coefficients (MFCC): The Mel-Frequency Cepstrum (MFC) is a representation of short-term power spectrum of sound. The MFCCs are coefficients that collectively makeup MFC. The difference between the cepstrum and the Mel frequency cepstrum is that in MFC, the frequency bands are equally spaced on the mel scale. The Mel frequency scale is linear frequency spacing below 1000Hz and logarithmic spacing above 1000Hz. In other words frequency filters are spaced linearly at low frequencies and logarithmically at high frequencies which is used to capture the phonetically important characteristics of speech. This is the important property of human ear. MFCC mimics the human auditory system. Fig. 1 shows the flowchart of MFCC method feature extraction.

Step 1: After normalizing the signal, the signal is converted into frames.

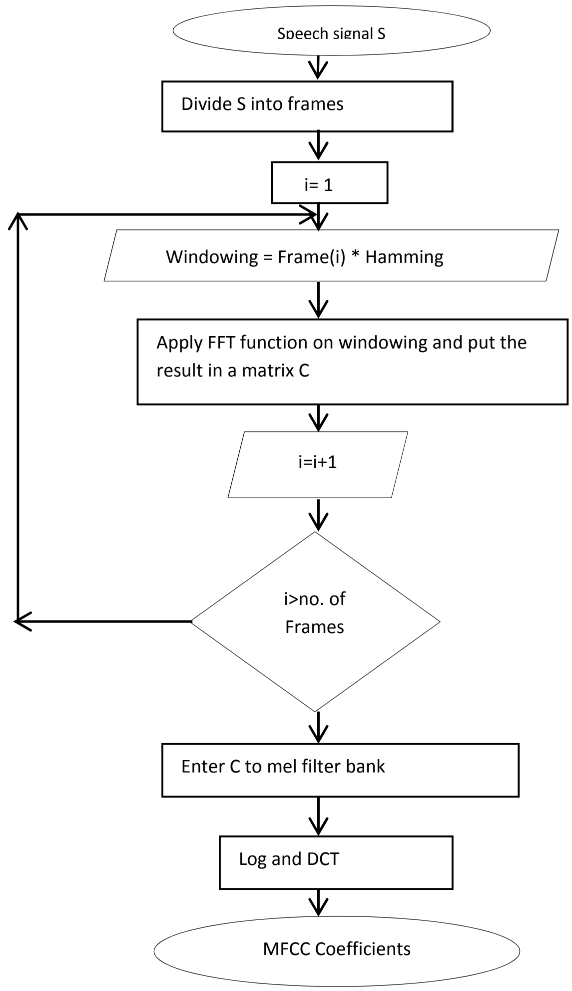

Fig. 1: MFCC method feature extraction.

Step 2: For every frame windowing is done. This minimizes the signal discontinuities at the beginning and end of the frame. Minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. Hamming window is used whose form is

$$w(n) = 0.53836 - 0.46164\ \cos\left(\frac{2\pi n}{N-1}\right) \text{------> 2.1}$$

Where N- No. of samples in a frame

n- Total no. of frames

Step 3: Apply Fast Fourier transform (FFT) to each frame. This step converts samples from time domain to frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier transform (DFT). It is defined over set of N samples$\{x_n\}$,

$X_n = \sum_{k=0}^{N-1} x_k\, e^{\frac{-2\pi kn}{N}}$, N-1 ------------------------> 2.2

Step 4: Once for all frames the FFT is calculated, the values are entered to Mel filter bank. The approximate formula for calculating mel frequency for given f frequency in Hz is

$mel(f) = 2595 * log_{10}\left(\frac{1+f}{700}\right)$-----------------------> 2.3

Step 5: Final step where we convert the log Mel spectrum back to time. The result is called Mel Frequency Cepstrum Coefficients (MFCC).we convert them back to time domain using discrete cosine transform (DCT). Therefore MFCC's is calculated as

$\tilde{c}_n = \sum_{k=1}^{K} \log \tilde{S}_k \cos\left[\frac{n(k-0.5)\pi}{K}\right]$------------------> 2.4

By applying the above procedure for each frame, a set of MFCCs is calculated. Fig. 2 shows the MFCC of phoneme /a/.

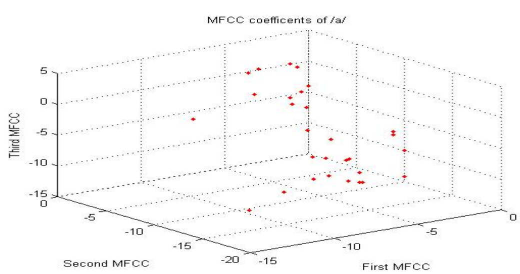

Fig. 2: MFCC of /a/

<u>Linear predictive coding (LPC):</u> Linear predictive coding is a technique used mostly in speech processing to estimate basic speech parameters like pitch, formants and spectral envelope of the speech signal, in compressed form, using the information of linear predictive model. LPC is one of the most useful methods for encoding good quality speech at low bit rates. The coefficients of current sample are generated by the linear combination of the past samples using autocorrelation or auto covariance method.

LPC coefficients are obtained by applying some procedures on the speech sample. First the autocorrelation is applied on the windowed frames. Every frame which is windowed is auto correlated by the p^{th} order. Once the autocorrelation coefficients are calculated, the Levinson – Durbin algorithm is used to find the LPC coefficients.

At the beginning, first coefficient of the first column (of LPC coefficients) is calculated as follows

$A_{(i,i)} = \left[\sum_{k=0}^{i} B_{k+2} \quad A_{(i-k,i)}\right]/E_i$ --------> 2.5

E(1) = B(1)

Where A = matrix of LPC coefficients

B = Vector of the autocorrelation coefficients

E = Vector of the energy of the prediction error

*E(i+1) = (1-A(i,i)^2) * E(i)* -----------> 2.6

The above equation is used to calculate the E(i) values.

In the second stage the second coefficient of the second column is calculated using equation 2.5 and the rest coefficients of the second column is calculated using the below equation.

$A_{(i,j)} = A_{(i,j-1)} + A_{(j,j)} * A_{(j-i,j-1)}$ -------------> 2.7

So on the procedure is calculated to find all the coefficients. The last column of the matrix A gives the coefficients. Fig. 2 shows LPC coefficients of phoneme /a/.

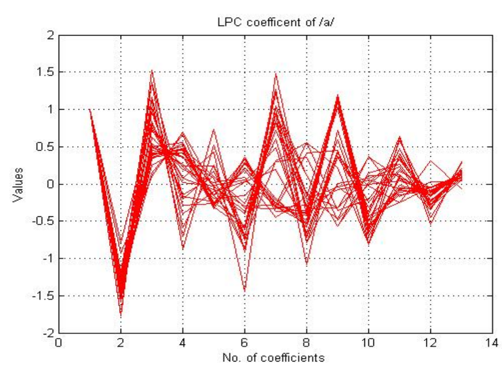

Fig. 3: LPC coefficients of /a/

IV. *Recognition using HMM*

Hidden markov model is the dynamic modeling scheme which is used to recognize the phonemes. A finite state machine with probabilistic state transitions is a markov model. It follows markov property and Markov property states that probability distribution of future states depends only upon the present state. HMM is

similar to markov model except that the states are hidden. Viterbi algorithm is used to find the most likely sequence of hidden states.
Features extracted are trained with HMM and finally recognition of the phonemes is done.

3. RESULTS AND DISCUSSION

In case 1 where MFCC is used for feature extraction and HMM is used for classification, the accuracy in each class is shown in Table 2.

Class	Match	Accuracy (%)
/a/	18/30	60
/i/	29/30	96.67
/ou/	29/30	96.67
/ai/	30/30	100
/au/	30/30	100

Table 2: Accuracy using MFCC and HMM

Total system efficiency using MFCC and HMM = 90.67%

In case 2 where LPC is used for feature extraction and HMM is used for classification, the accuracy in each class is shown in Table 3.

Class	Match	Accuracy (%)
/a/	19/30	63.33
/i/	30/30	100
/ou/	26/30	86.67
/ai/	28/30	93.33
/au/	18/30	60

Table 3: Accuracy using LPC and HMM

Total system efficiency using LPC and HMM = 80.67%

Fig. 4 and 5 shows all the MFCC and LPC coefficients of all 5 phonemes.

We can see that MFCC with HMM works better compared to LPC with HMM in terms of overall system efficiency. MFCC has many steps to compute coefficients (Windowing, FFT, DCT) while LPC (Autocorrelation) has fewer steps. This makes MFCC computationally complex compared to LPC.

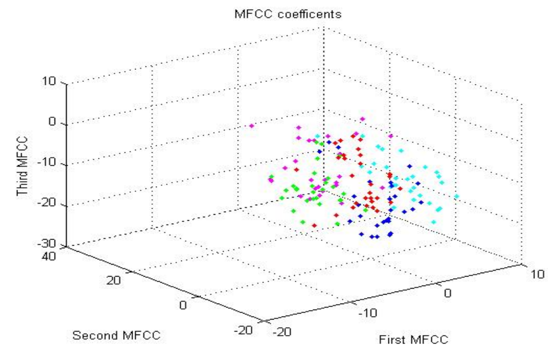

Fig. 4: All the MFCC coefficients in 3D plot

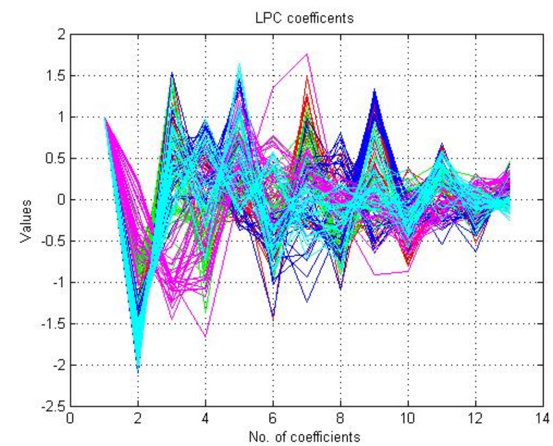

Fig. 5: All the LPC coefficients

4. CONCLUSION

In this work, a speaker dependent Hidden Markov Modeling for Kannada phoneme recognition was done. Feature was extracted in two different methods. That is Mel Frequency Cepstrum Coefficients and Linear Predictive Coding Coefficients. Mel frequency cepstrum coefficients with Hidden Markov Model gave 10% better efficiency compared to linear predictive coefficients with Hidden Markov Model. But simulation results show that computations in case of Mel frequency cepstrum coefficients is more compared to linear predictive coefficients computations. MFCC is computationally complex compared to LPC. So based on the application the feature extraction technique can be chosen with some relaxation of accuracy and efficiency. In this work only five Kannada phonemes were modeled. This should be further expanded to all. So that we can have a complete Kannada phoneme recognition system which further helps in building word recognition system.

5. REFERENCES

[1] Prashanth Kannadaguli, Ananthakrishna Thalengala, "Phoneme Modeling for Speech Recognition in Kannada using Hidden Markov Model", IEEE international conference on Signal processing, Informatics, Communication and engineering(SPICES), Feb, 2015

[2] Thomas F. Quatieri, "Discrete Time Speech Signal Processing Principles and practice", Pearson Education, 2011

[3] L.R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signals", Pearson Education, 2005

[4] Peri BhaskaraRao, "Salient phonetic features of Indian languages for Speech Technology", Sādhanā, Vol. 36, Part 5, Oct 2011, pp. 587–599.

[5] Hemakumar G., Punitha P., "Speech Recognition Technology: A Survey on Indian Languages", International Journal of Information Science and Intelligent System, Vol. 2, No. 4, 2013

[6] Yusnita M. A., "Phoneme-based or Isolated-word modeling Speech Recognition System?", IEEE 7th International Colloquium on Signal Processing and its Applications, 2011

[7] Hemakumar G., "Acoustic Phonetic Characteristics of Kannada Language", International Journal of Computer Science Issues, Vol. 8, Issue 6, No. 2, Nov 2011

[8] Youngjik Lee, Kyu-Woong Hwang, "Selecting Good Speech Features for Recognition", ETRI Journal, Vol. 18, No. 1, April 1996

[9] N K Narayanan, T M Thasleema, V Kabeer, "Malayalam Vowel Recognition Based On Linear Predictive Coding Parameters and k-NN Algorithm", International Conference on Computational Intelligence and Multimedia Applications, 2007

[10] Prashanth Kannadaguli, Vidya Bhat, "A Comparison of Bayesian Multivariate Modeling and Hidden Markov Modeling (HMM) based approaches for Automatic Phoneme Recognition in Kannada", Recent and Emerging trends in Computer and Computational Sciences (RETCOMP), 2015