Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

LPC and MFCC Analysis of Assamese Vowel Phonemes

Bhargab Medhi*, Prof P. H. Talukdar
Department of Instrumentation &
Gauhati University, Assam, India

Abstract— A speech signal contains many levels of information. Speech conveys the information about the language being spoken, the emotion, gender, and the identity of the speaker. Features parameters extracted from speech are very useful for speaker recognition as well as speech recognition. In this paper, the features LPC and MFCC are computed of Assamese vowel phonemes which will be helpful to develop Assamese Automatic speaker recognition (ASR) system. We create a small database for eight Assamese vowel phonemes, each phoneme is repeated 10 times, spoken in isolation by 10 speakers of equal number of male and female. Thus our database consists of 800 phonemes.

Keywords— Phoneme, Frame, LPC, LPCC, MFCC

I. INTRODUCTION

The Assamese is a New Indo-Aryan language spoken in the state of Assam. The Assamese people pronounce it as 'Ôxômiya'(IPA:[ɔxɔmija]). There are thirty two essential phonemes in Assamese language out of which eight are vowel phonemes and twenty four consonant phonemes. Vowels are classified as front, mid, or back, corresponding to the position of the tongue hump[1]. The Assamese vowel scripts with corresponding IPA symbols are presented in TABLE1:

Table 1 Assamese vowel phoneme

Phonemes IPA	Assamese Scripts (Graphemes)	Example	Meaning
/i/	ই, ঈ	কিতাপ, হাতী	Book, Elephant
/ɛ/,/e/	এ	এখন	One
/a/	আ	আম	Mango
/ɔ/,/ɒ/	অ', অ	ল'ৰা, বহল	Boy,Width
/o/	ও	ওখ	Tall
/u/	উ , ঊ	উৰণ	Fly

LPC is a way of encoding the information in a speech signal into a smaller space for transmission over a restricted channel. LPC encodes a signal by finding a set of weights on earlier values that can guess the next signal value. Linear predictive coding (LPC) is a tool used in audio signal processing and also for speech processing which represents the spectral envelope of a digital speech signal in compressed form with the information of a linear predictive mode The result of LPC analysis is a set of co-efficient a[1….k] and an error signal e(n), the error signal will be as small as possible and represents the difference between the predicted signal and the original[3,4].

MFCC is also a feature widely used in automatic speech and speaker recognition introduced by Davis and Mermelstein in the year 1098's.The main point to understand about speech is that the sounds generated by a human are filtered by the shape of the vocal tract including tongue, teeth etc. The shape determines what sound comes out. If we can determine the shape accurately , this should give us an accurate representation of the phoneme being produced. The shape of the vocal tract manifests itself in the envelop of the short time power spectrum, and the job of MFCCs is to accurately represent this envelop[9].

II. LPC

LPC (Linear predictive coding) is the most useful method for encoding good quality speech at a low bit rate[13,14]. It gives extremely accurate estimates of speech parameters. Since a very small error can distort the whole spectrum, LPC has to be tolerant of transmission errors[9]. The figure1 shows the block diagram of LPCC:

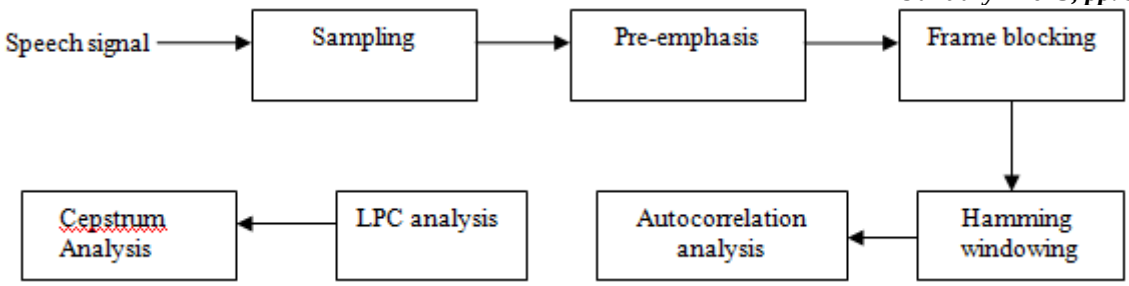

Fig1: The block diagram of LPCC

A. Speech sampling

Higher sampling frequency or more sampling precision gives higher recognition accuracy. It is not necessary to increase the sampling rate beyond 11,025 Hz or the sampling precision higher than 8 bits.

B. Pre-emphasis

The speech waveform which is digitized has a high dynamic range and it suffers from additive noise. So pre-emphasis is used to spectrally flatten the signal so as to make it less susceptible to finite precision effects in the processing of speech[9,10]. The most widely used pre-emphasis is the fixed first-order system. The calculation of pre-emphasis is shown as follows.

$$H(z) = 1 - az^{-1} \qquad 0.9 \le a \le 1.0$$

The most common value for a is 0.95 (Deller et at; 1993). A Pre-Emphasis can be expressed as

$$\hat{s}(n) = s(n) - 0.95s(n-1)$$

C. Frame blocking

According to Rabiner (1993), the speech signal is said to be stationary when it is examined over a short period of time. In order to analyze the speech signal, it has to be blocked into frames of N samples, with adjacent frames being separated by M samples. If M ≤ N, then LPC spectral estimates from frame to frame will be quite smooth. On the other hand if M > N there will be no overlap between adjacent frames[5,6,7].

D. Windowing

Each frame is windowed in order to minimize the signal discontinuities or the signal is lessened to zero at the starting and ending of each frame. If window is defined as w(n), then the windowed signal is

$$\tilde{x}(n) = x(n)w(n), 0 \le n \le N-1$$

A typical window used is the Hamming window, which has the form

$$w(n) = 0.54 - 0.46\cos(\frac{2n}{N-1}), 0 \le n \le N-1$$

The value of the analysis frame length N must be long enough so that tapering effects of the window do not seriously affect the result.

E. Autocorrelation analysis

Auto correlation analysis is used to find fundamental frequency or a pitch of the signal[6].. Each frame of windowed signal is next auto correlated to give

$$R(n) = \sum_{x=0}^{N-1-m} \tilde{x}(n)\tilde{x}(n+m), m = 0,1,2,....., p$$

Where, the highest autocorrelation value, P is the order of the LPC analysis. The selection of p depends primarily on the sampling rate.

F. LPC analysis

The next processing step is the LPC analysis which converts each frame of autocorrelation coefficients R into the LPC parameters.This method of converting autocorrelation coefficients to LPC coefficients is known as Durbin's method. Levinson-Durbin recursive algorithm is used for LPC analysis.

$$E_0 = R(0)$$

$$k_i = [R(i) - \sum_{j-1}^{i-1} a_j^{i-1} R(i-j)] / E_{i-1}, 1 \le i \le p$$

$$a_i^i = k_i$$

$$a_i^i = a_i^{i-1} - k_i a_{i-j,}^{i-1} 1 \le j \le i-1$$

$$E_i = (1-k_i^2)E_{i-1}$$

The above set of equations is solved recursively for i = 1,2, p, where p is the order of the LPC analysis. The k_i are the reflection or PARCOR coefficients. The a_j are the LPC coefficients.The final solution for the LPC coefficients is given as

$$a_j = a_j^{(p)}, 1 \le j \le p$$

G. Cepstrum analysis

LPC cepstral coefficient, is a very important LPC parameter set, which is derived directly from the LPC coefficient set. The recursion used is

$$c_j = a_j + \sum_{k=1}^{j-1} (\frac{k}{j}).c_k.a_{j-k}; 1 \le j \le p$$

$$c_j = \sum_{k=j-p}^{j-1} (\frac{k}{j}).c_{k.}a_{j-k}; j > p$$

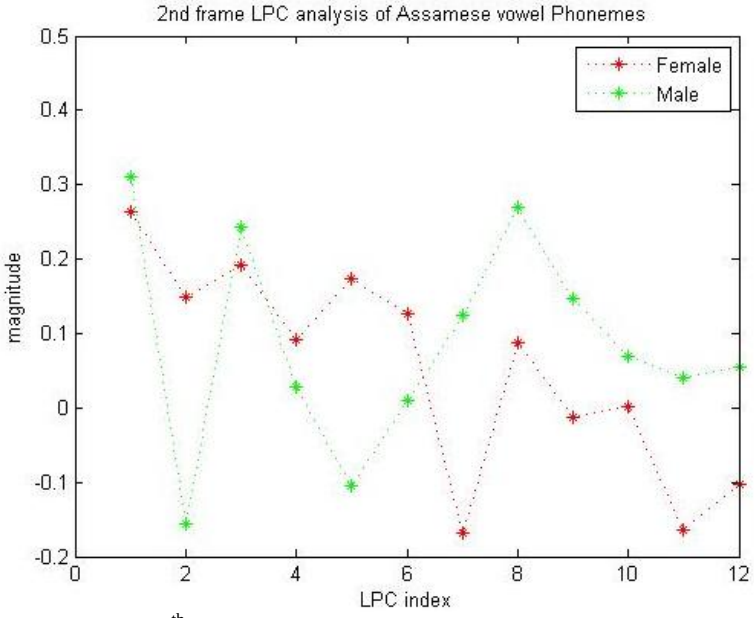

Fig2: 12th LPC analysis of vowel phoneme /i/ at frame2

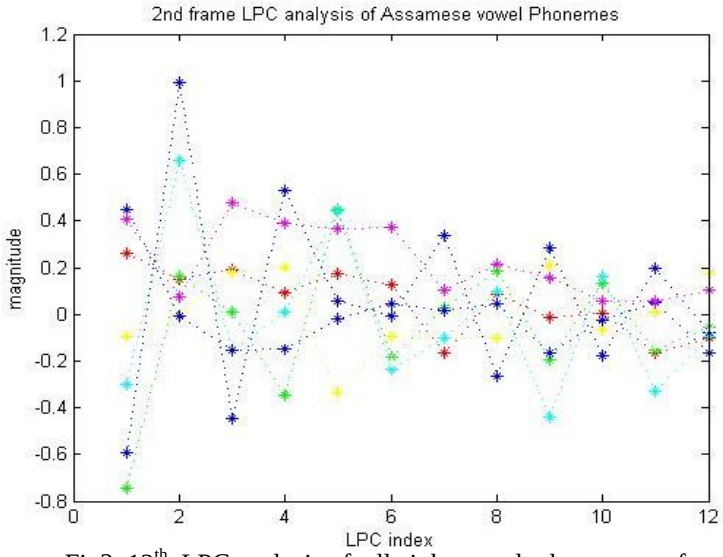

Fig3: 12th LPC analysis of all eight vowel phonemes at frame2

III. MFCC

The mel-frequency cepstral coefficients (MFCCs) introduced by Davis and Mermelstein is perhaps the most popular and common feature for SR systems. This may be attributed because MFCCs models the human auditory perception with regard to frequencies which in return can represent sound better. Figure2 shows the block diagram of the MFCCs [11,12]

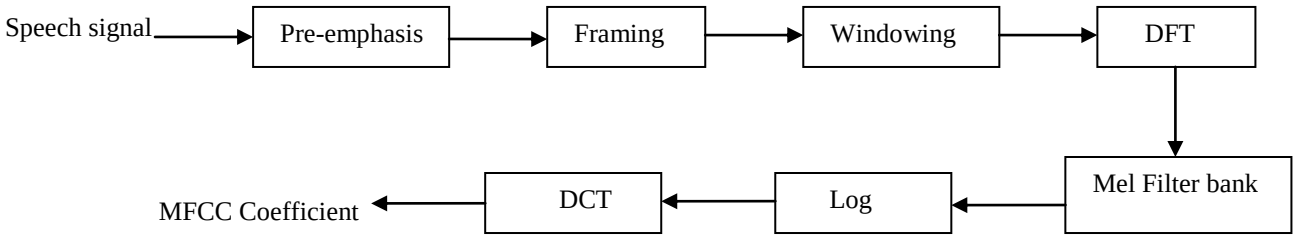

Fig4: Block diagram of MFCC

To obtain the MFCCs of a speech signal, the signal is first subjected to pre-emphasis filtering with the following finite impulse response(FIR)filter given by as:

$$H_{pre}(z) = \sum_{k=0}^{N} a_{pre}(k) z^{-k}$$

Its corresponding Z-transform:

$$H_{pre}(z) = 1 + a_{pre} z^{-1}$$

The value of the coefficient a_{pre} usually takes the value between -1.0 to -0.4. However , in speech recognition systems values that are almost near to -0.1 are usually used. The speech is processed on a frame-by-frame basis in what is called framing. Normally , a frame size of 20ms to 30ms is used and Windowing of these frames are done to compensate discontinuities within the speech signal as a result of segmentation and overlapped frames. A hamming window is used by equation;

$$w(n) = 0.54 + 0.46\cos(\frac{2\pi t}{T})$$

Windowing means multiplying the window function w(n) with the framed speech signals(n) to obtain the windowed speech signal $s_{0w}(n)$;

$$s_{0w}(n)=s(n)w(n)$$

The discrete Fourier transform(DFT) of the windowed speech signal is then computed by the following equations:

$$\hat{S}_{0w}(k) = \sum_{n=0}^{N-1} s_{0w}(n) e^{\frac{-j2\pi kn}{N}}$$

The mel-filterbank is a triangular bandpass filter which is equally spaced around the Mel-Scale. A Mel is a unit of perceived pitch or frequency of a tone. The mapping between real frequency (hz) and Mel-frequency is given by the following equations as[5,6,7]:

$$f_{mel} = 2595\log(1+\frac{f}{700})$$

The power spectrum from the DFT step is then binned by correlating it with each triangular filter in order to reflect the frequency resolution of the human ear. Binning means multiplying the power spectrum coefficients with the triangular filter gain or coefficients and summing the resultant values to obtain the mel-spectral coefficients as in equation:

$$G(k) = \sum_{n=0}^{\frac{N}{2}} \eta_{kn} . [\hat{s}_{0w}(k)]^2$$

Where η_{kn} is the triangular filter coefficients, k=0,1,2,…..k-1, n=0,1,2,…..N/2 and G(k) is the mel-spectral coefficients.

After that, the log of the mel-spectral coefficients G(k), is taken. This step is to level unwanted ripples in the spectrum and done the following equation;

$$m_k = \log G(k)$$

Finally , DCT is applied to the log mel-cepstrum m_k as in equation to obtain the Mel-frequency Cepstral Coefficients(MFCC) c_i of the ith frame:

$$c_i = \sqrt{\frac{2}{N}} \sum_{k=1}^{N} m_k \cos(\frac{\pi i}{N}(k-0.5))$$

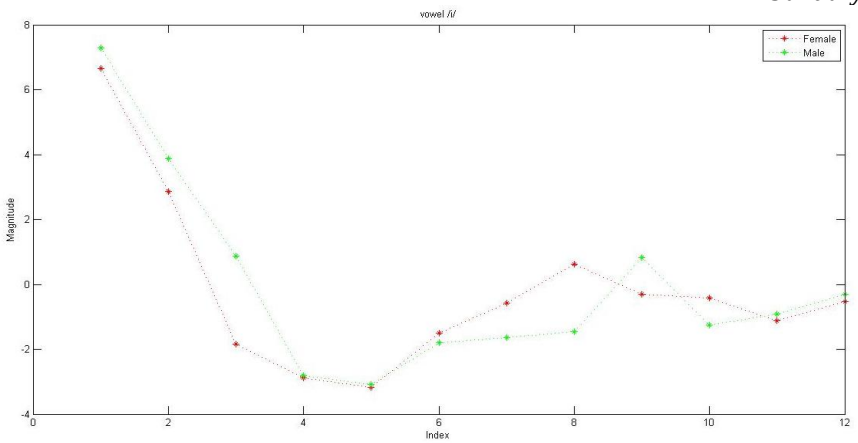

Fig5: The 12 MFCC coefficients of 8th frame of vowel phoneme /i/

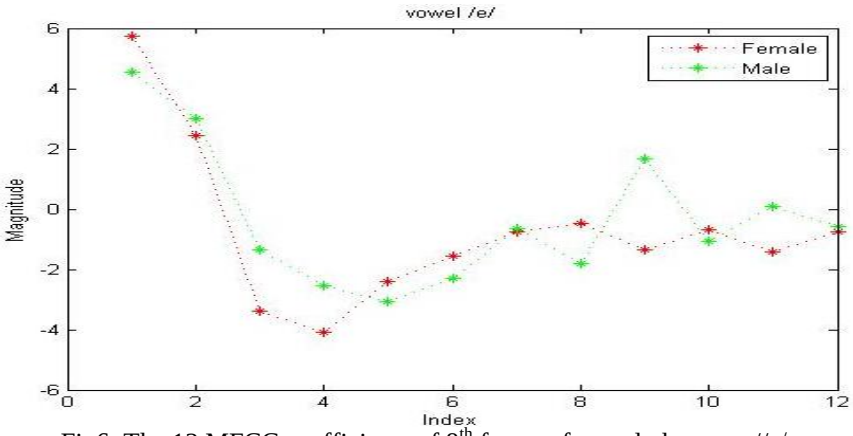

Fig6: The 12 MFCC coefficients of 8th frame of vowel phoneme //ε/

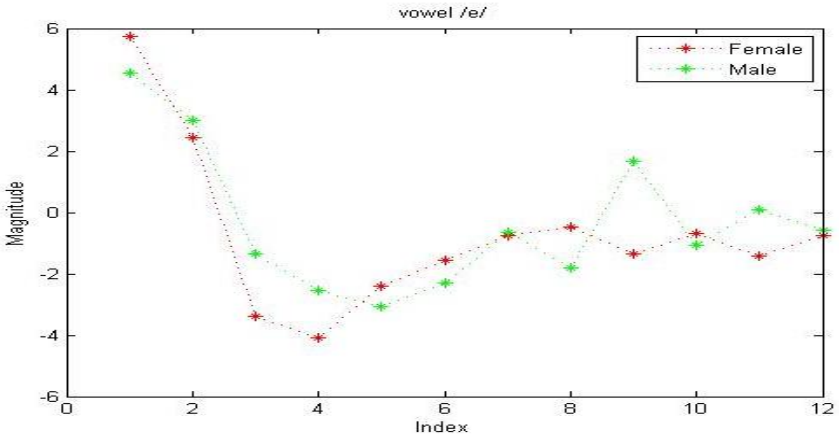

Fig7: The 12 MFCC coefficients of 8th frame of vowel phoneme /e/

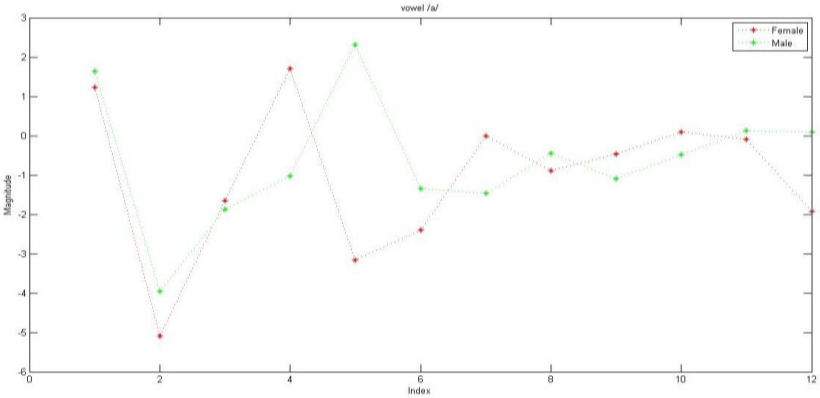

Fig8: The 12 MFCC coefficients of 8th frame of vowel phoneme /a/

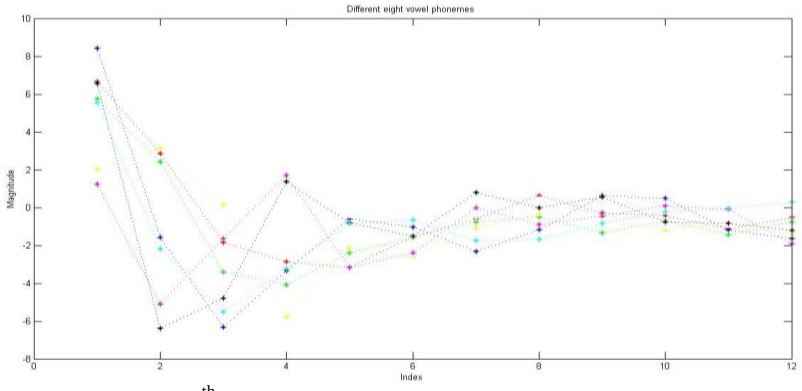

Fig9: The 12 MFCC coefficients of 8th frame of eight all Assamese vowel phoneme uttered by a female Speaker

IV. RESULT AND CONCLUSIONS

From the experimental results, the feature parameters LPC and MFCC can identify and recognise the speech signal. Result shows that the estimation of LPC and MFCC reflect effectively the difference in different speakers of different Assamese vowel phonemes. This will be very helpful to design a automatic Assamese Speech and Speaker Recognition System.

REFERENCES

[1] Banikanta Kakati, *Assamese, its Formation and Development*,5th ed.,Guwahati, India, LBS Publications,2007.
[2] Gold and N. Morgan, *Speech and Audio Processing: Processing and Perception of Speech and Music* ,New York, 2000.
[3] G.K.Vallabha and B. Tuller, *Systematic errors in the formant analysis of steady-state vowels*, Speech Communication, Vol. 38, 2002,pp.141.
[4] L. Flanagan , *Speech Analysis, Synthesis, and Perception*, 2nd edition, New York, Springer-Verlag, 1972.
[5] L.R.Rabiner and R. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, NJ,Prentice Hall, 1979.
[6] F. Jelinek, *Statistical Methods for Speech recognition*, Cambridge, The MIT Press, 1998.
[7] T.B. Adam and Md. Salam, *Spoken English Alphabet Recognition with Mel Frequency Cepstral Coeffifients and Back Propagation Neural Networks*,IJCA, March 2012, vol. 42.
[8] T. K. Das and P.H. Talukdar, *Cepstral Analysis of Assamese Vowel Phonemes*, IJACST, Aug. 2013,vol2.
[9] Nisha V. S, M. Jayasheela, *Speaker Identification Using Combined MFCC and Phase Information*, IJARCCE, Feb. 2013, Vol.2.
[10] E.M.M and M.S.S, *LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification*, IJSPIPPR, June 2013, Vol.6.
[11] Dr. Joseph Picone, *Fundamentals of Speech Recognition: A short Course*, ISIP.
[12] A. Maesa, F. Garzia, M. Scarpiniti, R. Cusani,*Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models*, JIS, March 2012, PP 235.
[13] M. Ali Hossain,M.M.Rahman,U.K.Prodhan,M.F.Khan,*Implementation of Back-Propagation Neural Network For Isolated Bangla Speech Recognition*, IJIST,July 2013, Vol. 3.
[14] J.P.Campbell, *Speaker Recognition: A tutorial*, Proc IEEE, Sep. 1997,Vol.85,pp. 1437.

