International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.05, May 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Feature Extraction Methods based on Linear Predictive Coding and Mel Frequency Cepstral Coefficients for Recognizing Spoken Words in Assamese Language

Dr. Mousmita Devi

Department of Computer Science, Handique Girls' College, India

ABSTRACT

In order to obtain or designing an intelligent and accurate system for the automatic recognition of speech, feature extraction process is considered as the key and most important phase. There are different speech feature extraction techniques available, but the most powerful and dominant techniques are considered as Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC). These techniques are based on spectral analysis. In this paper, an attempt has been made to investigate various spectral properties of Assamese speech at the word level with the help of these two spectral feature extraction methods. Algorithm to find out the feature vectors are discussed in this paper. The implementation results obtained are analyzed for vowels as well as different types of word structures. Paper is concluded with applying the k-mean clustering to the features extracted.

Key words: Assamese word, LPC, MFCC, Spectral features, Speech Recognition.

I.INTODUCTION

It is difficult for a machine to differentiate between different kinds of sounds as human beings perceive it. For example, a particular word is uttered by a number of speaker, the sound waves produced will be different due to the speech variations present in each individuals. Human beings are capable of recognizing this word because these sound waves have some common features that they can able to differentiate them. But for a machine to differentiate different sound waves, important features of the speech signals have to be extracted by some feature extraction techniques. The principal objective of front end processing in speech recognition is to bring a projection of the speech signal to a feature vector space [2]. And with the help of this feature vector space, for further processing, some relevant and important information from the speech signals can be easily extracted. Spectral analysis of speech signals basically involves digital filtering techniques to remove the additive noise and it also emphasizes important frequency components of interest [3]. Though a speech signal is non-stationary in nature, it is assumed to be stationary or static during a short period of time [1]. So the speech signal is divided into a number of frames and spectral analysis is done on these frame based segments [4]. In this study, each speech signal having sampling frequency 16000 Hz is blocked into frames of 256 samples, and consecutive frames are spaced 16 samples apart. Each frame is then multiplied by 8 sample Hamming window. Using the

Levinson-Durbin algorithm and autocorrelation analysis on each frame, an LPC analysis of order 10 is performed which allows us to estimate the LPC coefficients. After that I performed the cepstral analysis and converts the LPC coefficients into the cepstral coefficients (LPCC). From each frame of a speech signal, first 20 LPCC have been extracted. Similarly I perform the feature extraction of Assamese word speech Using MFCC. I employ the well-known K-means algorithm which constitutes clusters of the non-static feature structure extracted from the speech sample. K-Means function is used to group the speech feature parameters whose dimension has been reduced**.** From the resulting cluster centers (centroids), I have found the number of clusters as k=10 and k=16, performing 50 runs for each value of k. We conclude by noting that in this study, LPC and MFCC are considered as features which are further used to recognize Assamese word using Neural Network model.

II. IMPLEMENTATION OF FEATURE EXTRACTION SYSTEM USING LPC

LPC is known as a digital method used for encoding an analog signal. It is basically based on the mathematical approximation of the vocal tract, which in turn models the vocal tract as an Infinite Impulse Response (IIR) system to produce the speech signal.

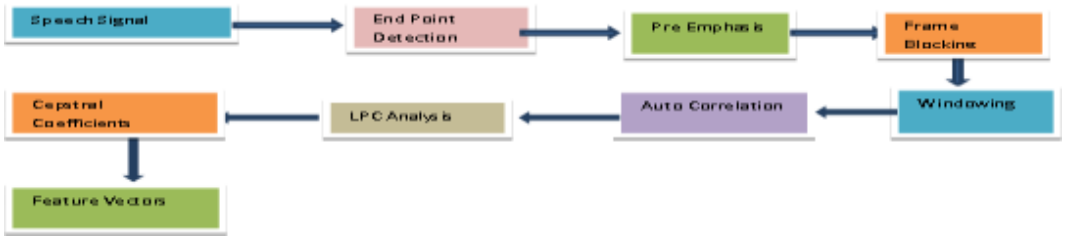

Figure 1: The block diagram of LPC processor.

In the present study, the values of the parameters are as given in **Table 1.**

Table 1: Values of LPC analysis parameters

PARAMETERS	**NOTATIONS**	**VALUES**
Sampling Frequency	Fs	16KHz
Number of samples in each frame	N	256
Number of samples shift between frames	M	128
Order of LPC analysis	P	10
Dimension of LPC based feature vectors	Q	20
Number of frames over which cepstral time derivatives are computed	K	2

After passing through all the steps, it yields the first 20 cepstral coefficients together with their corresponding graphs for all 32 frames. Physically, these coefficients are reflecting the difference of the biological structure of human vocal tract. The sampled speech signals, having sampling frequency 16000 Hz, are blocked into 32 frames with each frame containing 256 samples.

We have seen from the **Figure 2.1, Figure 2.2, Figure 2.3, Figure 2.4, Figure 2.5** and **Figure 2.6** that the value of the cepstral coefficients are more significant for 12th frame out of all 32 frames. In **Figure 2.1, Figure 2.2, Figure 2.3, Figure 2.4, Figure 2.5** and **Figure 2.6** it can be nicely observed that from 13th frames onwards the values are negative or tends to zero. This implies least significant bits are available from 13th frame onwards. This is clearly indicates that the speech features are more prominent for 12th frame and hence they are considered sufficient as features for further processing. In this present study several word patterns including (CV, CVC, VC etc) are taken and corresponding cepstral coefficients are evaluated. This can be realized by making a comparison of cepstral coefficient graphs for the first 12th frame of each speech signal for male and female informants each. The preliminary step in Spectral analysis of speech signal is to create a vowels database since for recognition of vowel sounds are most important in any language. For creating a most successful speech recognition system, accurate Spectral analysis of vowel sounds acts as the backbone. Below, we produce such comparison graphs for /a/(আ), /i/ (ই) and /u/ (উ) from vowel database and words like / nɔ / (**ন**), /azi/ (**আজি**) and /nak/ (**নাক**) bearing structures like CV, VCV and CVC respectively from word database spoken by a female and a male informant. In **Figure 2.7, Figure 2.8, Figure 2.9, Figure 2.10, Figure 2.11** and **Figure 2.12,** the cepstral coefficient graphs for the 12th frame of the Assamese vowel /a/(আ), /i/ (ই) and /u/ (উ) and words like / nɔ / (**ন**), /azi/ (**আজি**), /nak/ (**নাক**) for a female and a male speaker respectively have been shown. We have seen that there are distinct differences in the cepstral coefficient of male and female informants in the graphs. This provides an alternative and perhaps more technical way to identify sex of Assamese speaker.

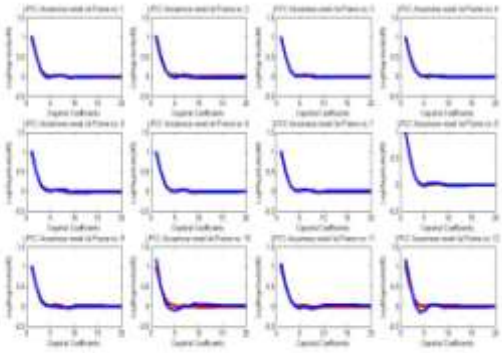

Figure 2.1: Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /a/ (আ) for a female and a male speaker.

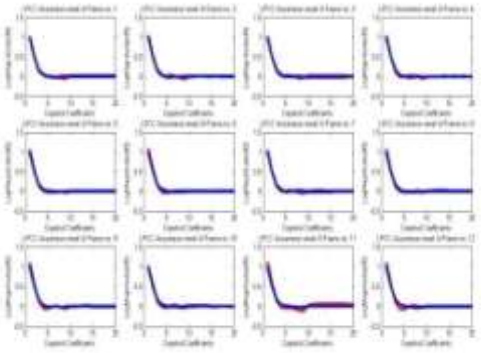

Figure 2.2: Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /i/ (ই) for a female and a male speaker.

International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.05, May 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

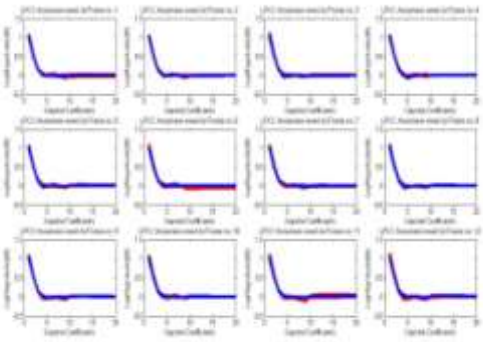

Figure 2.3: Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /u/ (উ) for a female and a male speaker.

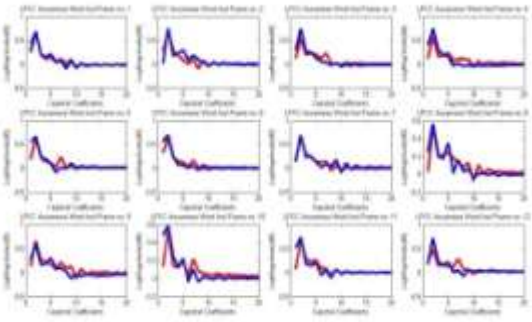

Figure 2.4: Cepstral Coefficients extracted from the first 12th frame of Assamese word / nɔ / (ন) "Nine", for a female and a male speaker.

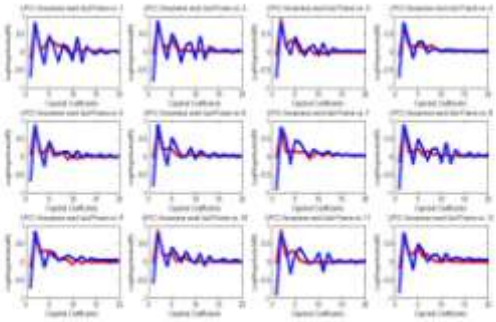

Figure 2.5: Cepstral Coefficients extracted from the first 12th frame of Assamese word / azi/ (আজি) "**Today**" **for a female and a male speaker.**

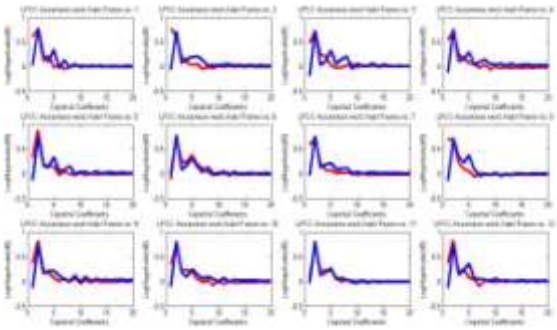

Figure 2.6: Cepstral Coefficients extracted from the first 12th frame of Assamese word /nak/ (নাক) "Nose" for a female and a male speaker.

International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.05, May 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

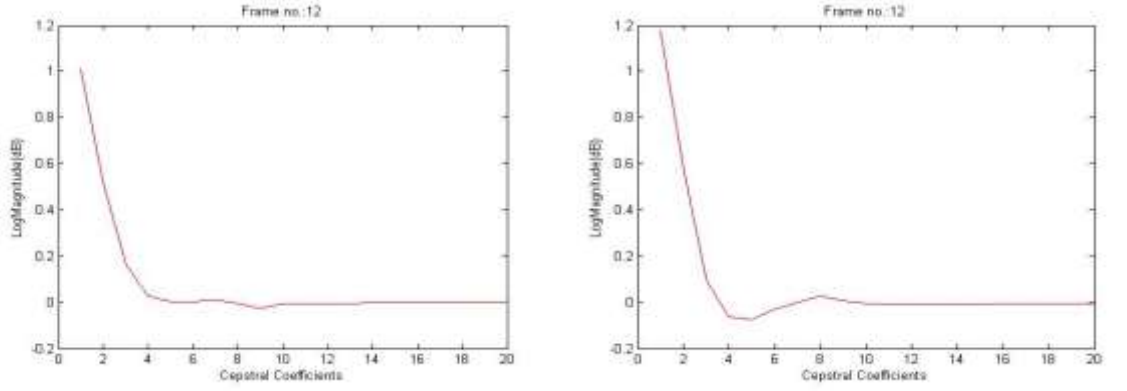

Figure 2.7: Cepstral Coefficients extracted from the 12th frame of Assamese vowel /a/ (আ) for a male and a female speaker.

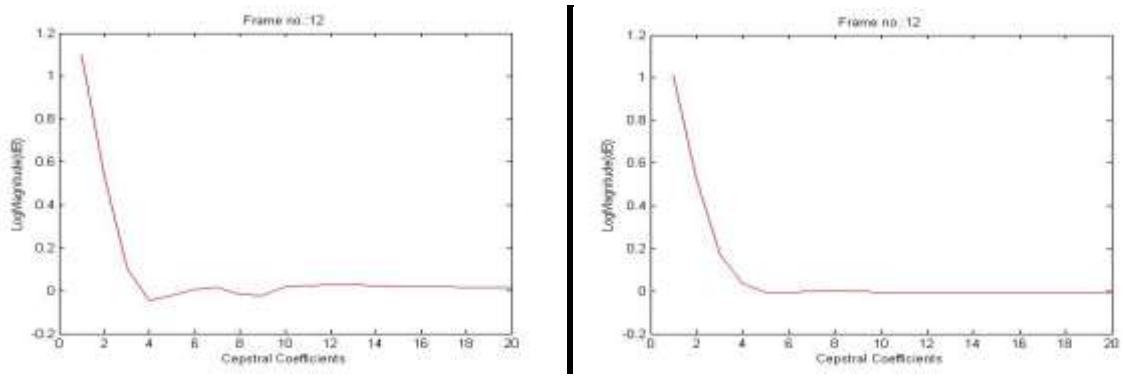

Figure 2.8: Cepstral Coefficients extracted from the 12th frame of Assamese vowel /i/ (ই) for a male and a female speaker.

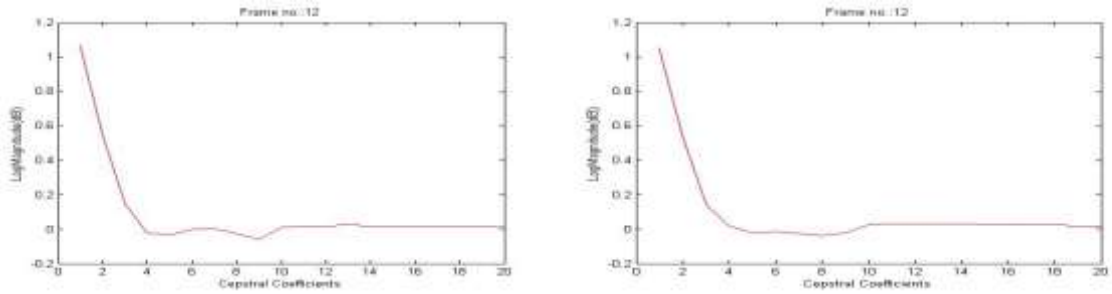

Figure 2.9: Cepstral Coefficients extracted from the 12th frame of Assamese vowel /u/ (উ) for a male and a female speaker.

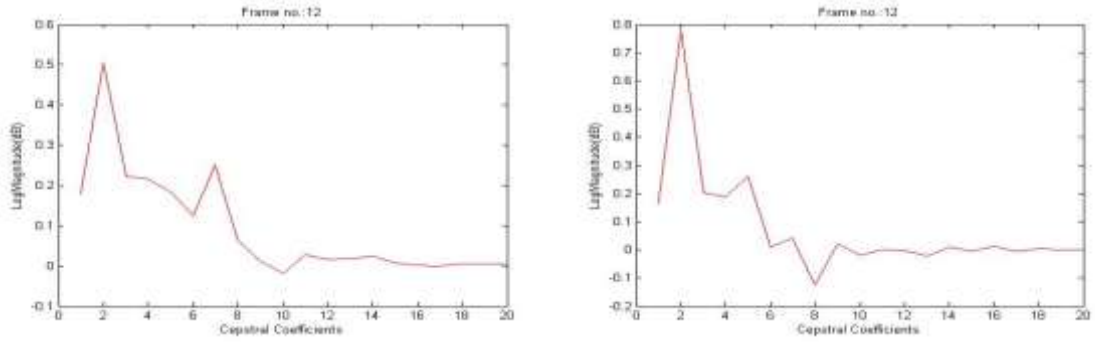

Figure 2.10: Cepstral Coefficients extracted from the 12th frame of Assamese word / nɔ / (ন) "Nine", for a male and a female speaker.

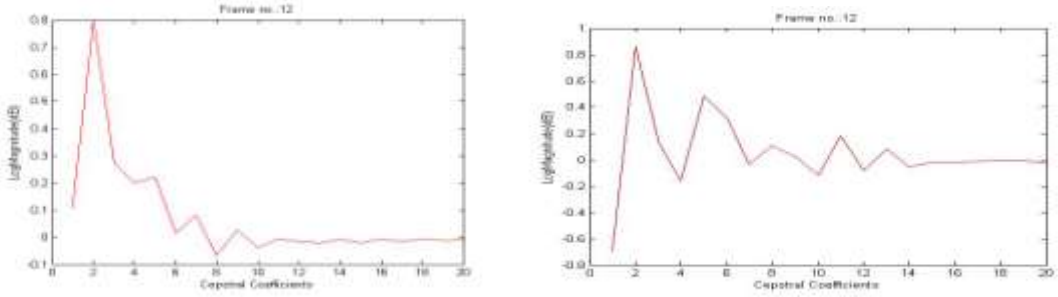

Figure 2.11: Cepstral Coefficients extracted from the 12th frame of Assamese word /azi/ (আজি) "Today", for a male and a female speaker.

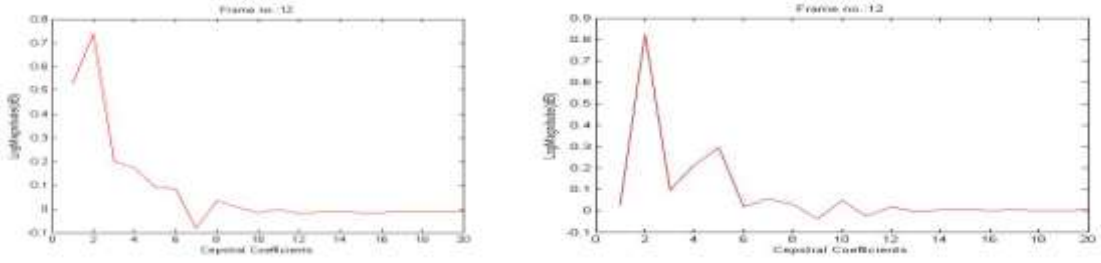

Figure 2.12: Cepstral Coefficients extracted from the 12th frame of Assamese word /nak/ (নাক) "Nose" for a male and a female speaker.

III. IMPLEMENTATION OF FEATURE EXTRACTION SYSTEM USING MFCC

MFCC is another one of the most popular spectral feature extraction technique which is a special case of homomorphic signal processing. In the speech processing technology, the Mel-Frequency Capstrum (MFC), is referred as the discrete cosine transform (DCT) of the log filter bank amplitude. MFCC is based on the known

variation of the human ear's critical bandwidths with the frequencies. The speech signal is expressed in the Mel frequency scale for determining the phonetically important characteristics of speech.

The flowchart to determine the MFCC is as shown below in **Figure 3.**

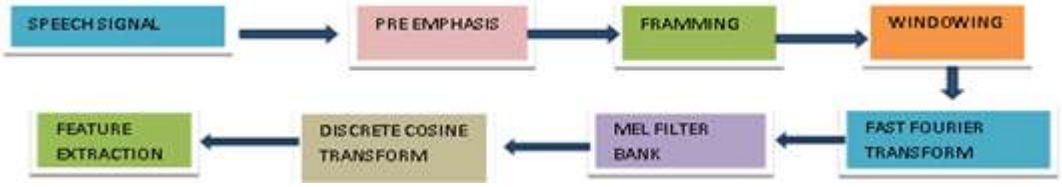

Figure 3: MFCC flowchart

The implementation procedure using MFCC also involves different computational steps like LPC. It yields the first 20 coefficients together with their corresponding graphs for the first 12 frames. The sampled speech signals, having sampling frequency 16000 Hz, are blocked into first 12 frames with each frame containing 256 samples.

This yields the first 20 coefficients together with their corresponding graphs for the first 12 frames for a male and a female speaker. **Figure 3.1, Figure 3.2, Figure 3.3, Figure 3.4, Figure 3.5** and **Figure 3.6** shows that the value of the coefficients for are more significant for first 12^{th} frame for male as well as female informants. From the all figures it can be nicely observed that from 13^{th} frames onwards the values are negative or tends to zero. This implies list significant bits are available from 13th frame onwards. Since mfcc correlates to the VOT of a speaker which in turn reflects the speech sounds that has been uttered. From the coefficients values of 12^{th} frame incorporates the most significant bits as per the significance of the word uttered. In this present study several word patterns including (CV, CVC and VC etc) are taken and corresponding mfcc are evaluated. This can be realized by making a comparison of mfc coefficient graphs for first 12^{th} frame of each speech signal. Below, we produce such comparison graphs for /a/(আ), /i/ (ই) and /u/ (উ) from vowel database and words like / nɔ / (**ন**), /azi/ (**আজি**) and /nak/ (**নাক**) bearing structures like CV, VC and CVC respectively from word database spoken by a female and a male informant. This is clearly indicates that the speech features are more prominent for 12^{th} frame and hence they are considered sufficient as features for further processing. Here also it have seen that there are distinct differences in the coefficient of male and female informants in the graphs. This provides an alternative and perhaps more technical way to identify gender of Assamese speaker.

International Journal of Advance Research in Science and Engineering
Volume No.07, Issue No.05, May 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

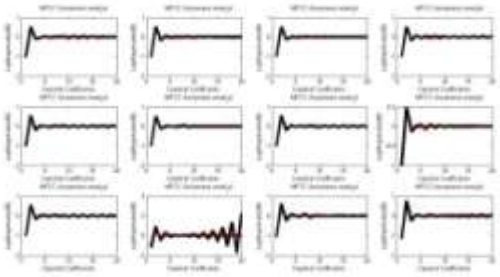

Figure 3.1: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /a/ (আ) for a female and a male speaker.

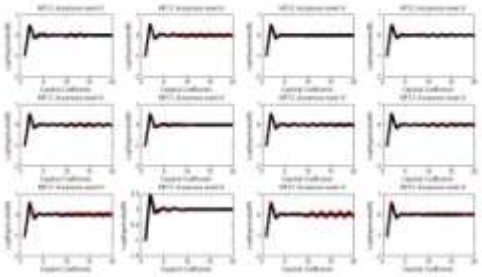

Figure 3.2: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /i/ (ই) for a female and a male speaker.

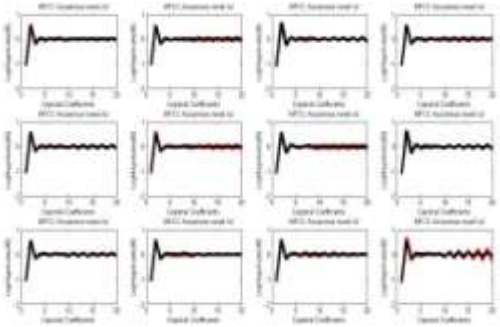

Figure 3.3: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese vowel /u/ (ঊ) for a female and a male speaker.

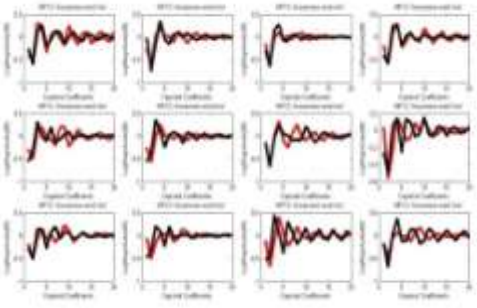

Figure 3.4: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese word / nɔ / (ন) "Nine", for a female and a male speaker.

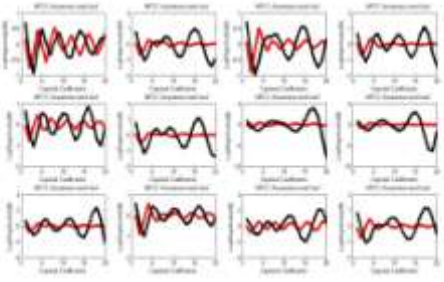

Figure 3.5: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese word / azi/

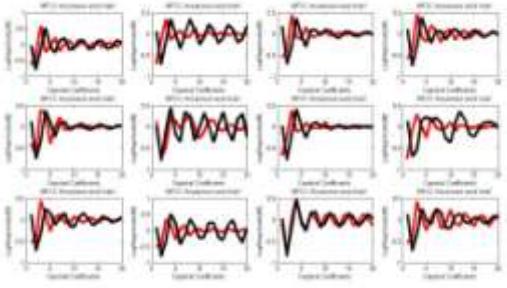

Figure 3.6: Mel Cepstral Coefficients extracted from the first 12th frame of Assamese word /nak/ (নাক)

(আজি) "Today"for a female and a male speaker. **"Nose" for a female and a male speaker.**

IV. CONCLUSION

In this paper, with technical details various aspects of feature extraction methods have been discussed. A frame based cepstral analysis is performed on both the database (vowel database and word database) which are I have taken as the speech sample. Each speech signal having sampling frequency 16000 Hz is blocked into frames of 256 samples, and consecutive frames are spaced 16 samples apart. Each frame is then multiplied by 8 sample Hamming window. Using the Levinson-Durbin algorithm and autocorrelation analysis on each frame, an LPC analysis of order 10 is performed which allows us to estimate the LPC coefficients. We then performed the cepstral analysis and converts the LPC coefficients into the cepstral coefficients (LPCC). From each frame of a speech signal, we have extracted the first 20 LPCC. From the results, it has seen that there are distinct differences in the cepstral coefficient graphs of male and female for the speech signal, which provides an alternative and perhaps technically more efficient way to identify gender of Assamese speaker. Similarly the feature extraction of Assamese word Using MFCC also has been performed. The paper is concluded by noting that in this research work, LPC and MFCC are considered as features which are further used to recognize Assamese word using Neural Network model.

REFERENCES

[1]. L. R. Rabiner, and B. H. Juang, Fundamentals of Speech Recognition,New Jersey, USA: Engle-wood Cliffs Publisher, 1993.

[2] Mousmita Devi "Spectral Analysis of Assamese Words and Their Recognition Using ANN", PhD Thesis submitted to Gauhati University, 2016.

[3] B. Plannerer. (2005). An Introduction to Speech Recognition [Online]. Available: http://www.speech-recognition.de/pdf/introSR.pdf

[4] Sonia Sunny, David Peter S., and K. Poulose Jacob, "Performance Analysis of Different Wavelet Families in Recognizing Speech, " International Journal of Engineering Trends and Technology, vol. 4, no. 4, pp. 512-517, April 2013.

[5] Sonia Sunny "A Hybrid Architecture for recognizing Speech Signals in Malayalam" Ph.D Thesis submitted to Cochin University of Science and Technology, September 2013.

[6]. Neeta Awasthy, J. P. Saini, and D. S. Chauhan, "Spectral Analysis of Speech: A New Technique," International Journal of Information and Communication Engineering, vol. 2, no. 1, pp. 19-28, Jan. 2006

[7] Qi Li, Jinsong Zheng, Augustine Tsai, and Qiru Zhou, "Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 3, pp. 146-157, Mar. 2002.

[8] Kapil Sharma, H. P. Sinha, and R. K. Aggarwal, "Comparative Study of Speech Recognition System Using Various Feature Extraction Techniques," International Journal of Information Technology and Knowledge Management, vol. 3, no. 2, pp. 695-698, Jul.-Dec. 2010.

[9] PreetiSaini, ParneetKaur, and MohitDua, "Hindi Automatic Speech Recognition Using HTK," International Journal of Engineering Trends and Technology (IJETT), vol. 4, no. 6, pp. 2223-2229, Jun. 2013.

[10] Eslam Mansour mohammed, Mohammed SharafSayed, Abdallaa Mohammed Moselhy and AbdelazizAlsayedAbdelnaiem ," LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6, No. 3, June, 2013