ISSN 2320 - 2602

Volume 2, No.9, August 2013

International Journal of Advances in Computer Science and Technology

Available Online at http://warse.org/pdfs/2013/ijacst06292013.pdf

Cepstral Analysis of Assamese Vowel Phonemes

Tapashi KashyapDas[1], P.H.Talukdar[2]

[1] Department of Instrumentation and USIC Gauhati University, Assam, India, tapashi.kashyap@gmail.com
[2]Department of Instrumentation and USIC Gauhati University, Assam, India, phtassam@yahoo.com

ABSTRACT

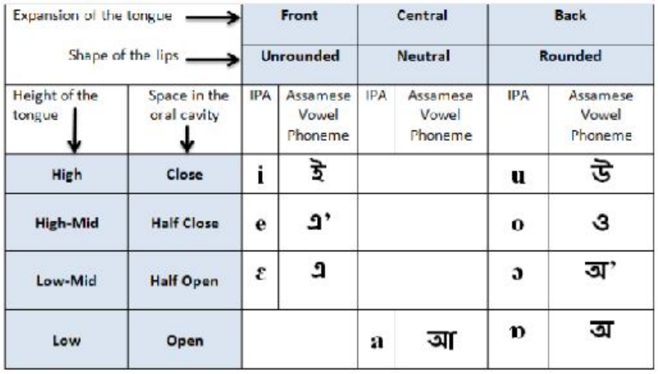
Assamese is an Indo-Aryan family of languages, mainly spoken in the North-Eastern part of India and possesses its own unique identity. We carry out the cepstral analysis of Assamese vowel phonemes which is useful for accurate classification and identification of originality and gender of Assamese speaking informants. We create a small database for the Assamese vowels, spoken in isolation by 20 speakers with equal numbers of male and female speaker. Each spoken phoneme is repeated 10 times by each speaker. Thus, our database consists of 1600 phonemes. We use the linear prediction coding technique to compute the linear prediction cepstrum coefficients (LPCC), the weighted LPCC and the delta weighted LPCC for all the Assamese vowel phonemes of our database. These features are useful for Assamese speech recognition.

Key words: Assamese vowel phonemes, Cepstral analysis, LPC, LPCC, Weighted LPCC, Delta weighted LPCC.

1. INTRODUCTION

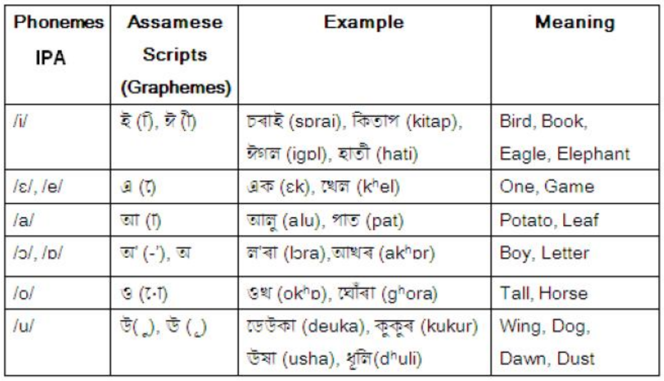
The Assamese (or Asamiya) (IPA:ɔxɔmija) is a major language in the north-eastern part of India whose origin root back to the Indo-European family of languages [1, 2, 3]. The phonetic character set of Assamese which was derived from Sanskrit, possesses its own unique identity. There are thirty two essential phonemes in Assamese language out of which eight are vowel phonemes and twenty four are consonant phonemes [1]. Vowels are classified as front, mid, or back, corresponding to the position of the tongue hump, while consonants are basically classified depending on the touch point of tongue inside the mouth as kanthhawya (velar), talawya (palatal), murdhanya (retroflex), dantawya (dental), and aushthawya (labial). In Table 1, we depict the arrangements of Assamese vowels according to the above classification. It should be clearly stressed that many international language like Finnish, Japanese, Turkish, etc. are phoneme based, i.e., phonemes are the same as letters (graphemes). In contrast, Assamese scripts, derived from the same source as that of the Devanagari scripts, consist of thirty-nine consonant and eleven vowel symbols [1, 4] which are arranged in the way the Devanagari scripts are arranged in a well-structured scientific manner based on phonetic principles. The written symbols in Assamese vowel scripts and their corresponding vowel phonemes are presented in Table 2. It is obvious from these tables that single phoneme may corresponds to more than two or three graphemes. Furthermore, there are possibilities that a grapheme may correspond to multiple phonemes.

Table 1: Classification of Assamese Vowels and their IPA representations



Expansion of the tongue →		Front		Central		Back	
Shape of the lips →		Unrounded		Neutral		Rounded	
Height of the tongue ↓	Space in the oral cavity ↓	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme
High	Close	i	ই			u	উ
High-Mid	Half Close	e	এ'			o	ও
Low-Mid	Half Open	ɛ	এ			ɔ	অ'
Low	Open			a	আ	ɒ	অ

Table 2: Assamese Vowel Phonemes



Phonemes IPA	Assamese Scripts (Graphemes)	Example	Meaning
/i/	ই (ি), ঈ (ী)	চৰাই (sɒrai), কিতাপ (kitap), ঈগল (igɒl), হাতী (hati)	Bird, Book, Eagle, Elephant
/ɛ/, /e/	এ (ে)	এক (ɛk), খেল (kʰel)	One, Game
/a/	আ (া)	আলু (alu), পাত (pat)	Potato, Leaf
/ɔ/, /ɒ/	অ' (-'), অ	ল'ৰা (lɔra),আখৰ (akʰɒr)	Boy, Letter
/o/	ও (ো)	ওখ (okʰɒ), ঘোঁৰা (gʰora)	Tall, Horse
/u/	উ(ু), ঊ (ূ)	ডেউকা (deuka), কুকুৰ (kukur) ঊষা (usha), ধুলি(dʰuli)	Wing, Dog, Dawn, Dust

Linear Predictive Coding (LPC), formulated in the late 1960s and early 1970s, is one of the most important and frequently used techniques in speech processing, particularly, to model the vocal tract [5, 6, 7, 8]. The term “linear prediction” itself signifies that the prediction of the output of a linear system based on its input x_n and previous outputss_{n-1}, s_{n-2}, . . . , s_{n-p}, given by

$$\hat{s}_n = \sum_{k=1}^{p} a_k s_{n-k} + \sum_{k=0}^{N} b_k x_{n-k}. \qquad (1)$$

Here $\hat{s}_n$ refers to the estimate or prediction of s_n. The idea is that once we know the input x_n and the output s_n, it would be possible to predict the behavior of the unknown system (vocal tract). The output s_n is delayed so that we cannot use the real output. The problem is now to determine the constants a_k and b_k in such a way that $\hat{s}_n$ approximates the real output s_n as accurately as possible.

According to z-transform, if S (z) is the z-transform of the signal s_n in time domain, the vocal tract transfer function of H (z) is the ratio of output S (z) and input X (z), and is modeled by an all-pole filter

$$H(z) = \frac{S(z)}{X(z)} = \frac{b_0}{1 - \sum_{k-1}^{p} a_k z^{-k}}, \qquad (2)$$

where b_0 is the gain factor, a_k are the linear prediction coefficients, and p is the number of poles. The cepstrum C (z) can be obtained from the transfer function H (z) as

$$\ln H(z) = C(z) = \sum_{n=1}^{\infty} c_n z^{-n}, \qquad (3)$$

where

$$c_n = \frac{1}{2\pi i} \int_C \ln H(z) z^{n-1} dz \qquad (4)$$

in which C is the closed contour about the origin and lies within the region of convergence. Based on the all-pole model for H (z), given in Eq. (2), a recursion relation between the linear predictive cepstral coefficients (LPCC), c_k, and the prediction coefficients a_k can be found out. Taking derivative of Eq. (4), one can find out the recursion as follows:

$$\begin{aligned} c_1 &= a_1, \\ c_n &= a_n + \sum_{k-1}^{n-1} \frac{k}{n} c_k \, a_{n-k}, \quad 1 \leqslant n \leqslant p \\ c_n &- \sum_{k-1}^{n-1} \frac{k}{n} c_k \, a_{n-k}, \quad n > p, \end{aligned} \qquad (5)$$

where $a_1, \dots, a_p$ are the LPC coefficients of order p and c_n (n = 1, . . . , p) are the corresponding first p values of the cepstrum.

Since LPC based speech features have long been effectively used in speech recognition of various international languages [6, 7, 9, 10], here we take such an approach to analyze the cepstral features of the Assamese vowel phonemes based on LPC technique. The rest of the paper is organized as follows. In Sec. II, we present the feature extraction method. In Sec. III, we present our results with an inclusive conclusion highlighting some related emerging questions that may lead to further research along the same line.

2. LPC BASED FEATURES EXTRACTION

Since, there are a large number of thinly populated ethnic communities with their distinct languages and dialects in the north-east part of India, particularly in Assam, we choose the speakers from a wide area of Assam in order to keep the purity of the Assamese language. We create a database for eight Assamese clean vowels spoken by 20 speakers with equal numbers of male and female speaker. All speakers are well educated, speak Assamese as their native language, and their ages ranged from 25 to 35 years. Each speaker uttered all the eight Assamese vowel phonemes in noise free environment. Each spoken phoneme is repeated 10 times by each speaker. Thus, our database consists of 1600 phonemes.

Signals received by high quality index head microphone are recorded randomly from the speakers in closed-room noise-free environment using Cool Edit Pro 2.0 and Wave surfer software. The recorded analog speech signal is then digitized at a sampling rate of 8000Hz and quantized by 16 bit analog to digital converter in order for subsequent processing. This is made by using the software PRAAT in the Mono, 16 bit PCM wav format with 8KHz sampling rate. Each digitized signal is then blocked into 32 frames of data in which each frame contains 250 samples, and consecutive frames are spaced 30 samples apart. Each frame is processed via a pre-emphasizing filter defined as

$$s'_n = s_n - 0.96 \times s_{n-1}, \quad n = 1, 2, \dots 249$$

Where s_n is the nth sample of the frame s and$s'_0 = s_0$. Each preemphasized frame is then multiplied by an 80 sample Hamming window. An LPC analysis of order 10 is performed in each frame which allow us to estimate the LPC coefficients. From the LPC coefficients, a_m, we extract the first 20 LPC cepstral coefficients (LPCC), c_m, for each frame using the recursion given in Eq. (5).

We have seen that the first 20 LPCC values are more significant for 16th frame out of all 32 frames. This can be realized by making a comparison of LPCC graphs for 8th, 16th, and 24th frame of each speech signal. In Figure 2 and Figure 3, we have shown such graphs for the vowels /ɒ/ and /a/ spoken by a male speaker. This clearly indicates that the speech features are more prominent for 16th frame and hence they are sufficient to consider as features for further processing. In Figure 4 –11, we show the cepstral coefficient graphs for the 16th frame of the Assamese vowels for a female and a male speaker respectively. We have seen that there are distinct differences in the cepstral coefficient graphs of male and female informants for the vowels /a/, /ɛ/, and /u/. This provides a technically sound way to identify gender of Assamese speaker.

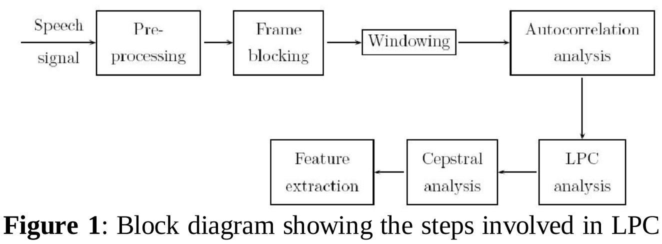


Figure 1: Block diagram showing the steps involved in LPC based feature extraction.

However, we have seen that the LPCC values calculated above have numerically smaller values and provide less contribution. Further, various research suggest that the low order cepstral coefficients are sensitive to overall spectral slope while the high order cepstral coefficients are sensitive to noise [7, 11, 12]. These cepstral coefficients are, therefore, not suitable to consider as input features for a statistical speech recognizer. It is therefore necessary to weight the cepstral coefficients by a tapered window so as to minimize these sensitivities by providing most weight to the middle coefficients. In the present approach, as suggested by Rabiner [7], the cepstral coefficients are weighted by a window function tapers at both ends given by

$$w_c(n) - 1 - \frac{Q}{2}\sin\frac{\pi n}{Q}, 1 \leqslant n \leqslant Q, \qquad (6)$$

where Q is the duration of the cepstral window weighting. Fig. 12 shows such a cepstral window function used in our purpose for Q = 20. This yields the corresponding weighted cepstral coefficients (WLPCC), $c_w(n)$, as

$$c_w(n) = c_n\, w_c(n). \qquad (7)$$

In addition to weighted cepstral coefficients, we also calculate the delta weighted LPCC

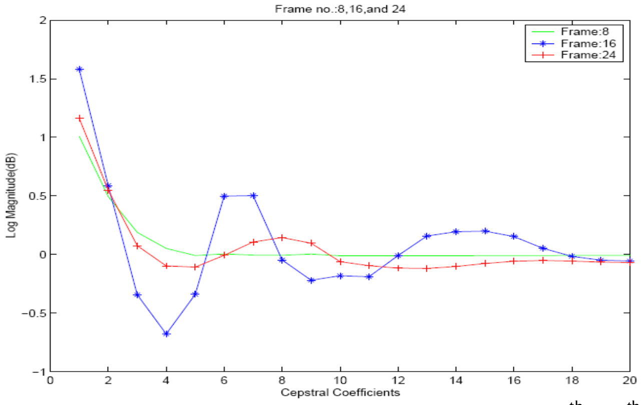


Figure 2: Cepstral coefficients extracted from the 8th, 16th, and 24th frame of the Assamese vowel /ɒ/ for a male speaker.

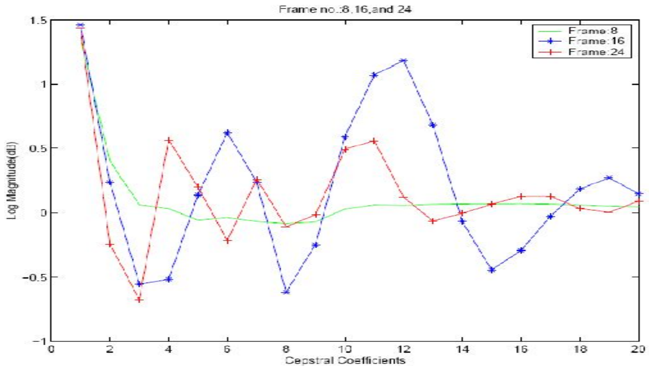


Figure 3: Cepstral coefficients extracted from the 8th, 16th, and 24th frame of the Assamese vowel /a/ for a male speaker.

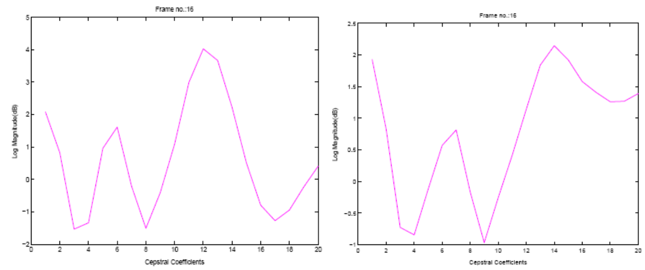


Figure 4: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /ɒ/for a female and a male speaker respectively.

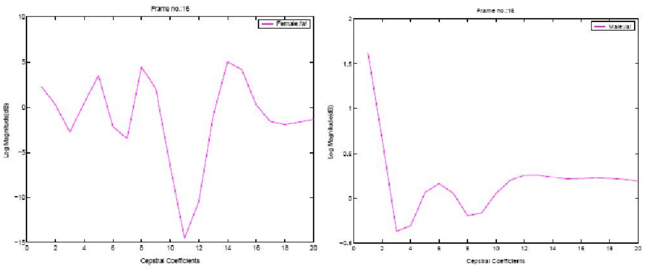


Figure 5: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /a/for a female and a male speaker.

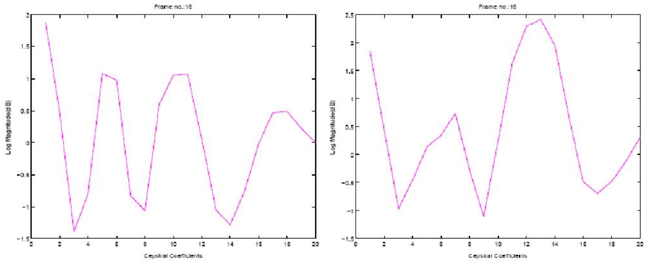


Figure 6: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /ɔ/for a female and a male speaker.

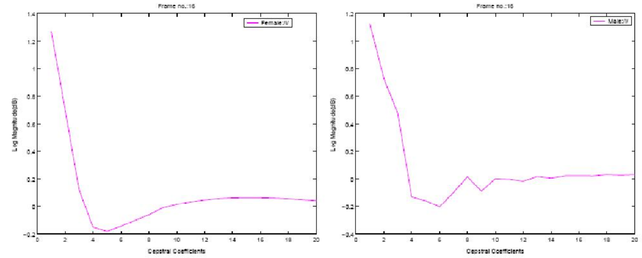


Figure 7: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /i/for a female and a male speaker.

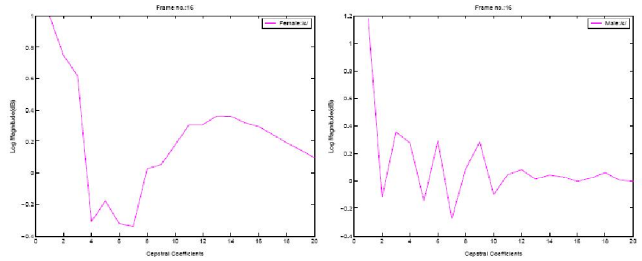


Figure 8: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /ɛ/ for a female and a male speaker.

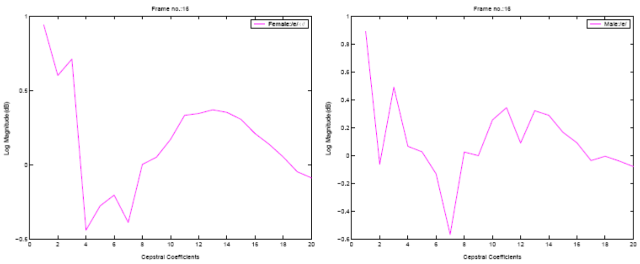


Figure 9: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /e/for a female and a male speaker.

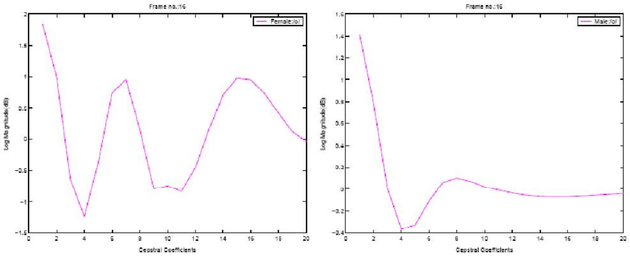


Figure 10: Cepstral coefficients extracted from the 16th frame of the Assamese vowels /o/for a female and a male speaker.

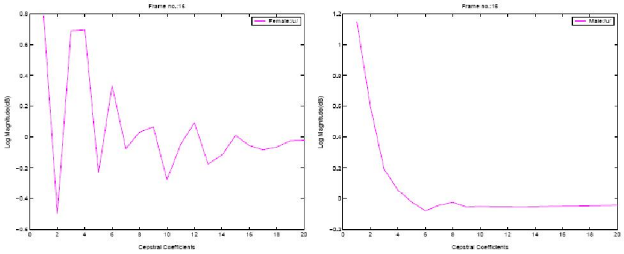


Figure 11: Cepstral coefficients extracted from the 16th frame of the Assamese vowel /u/for a female and a male speaker.

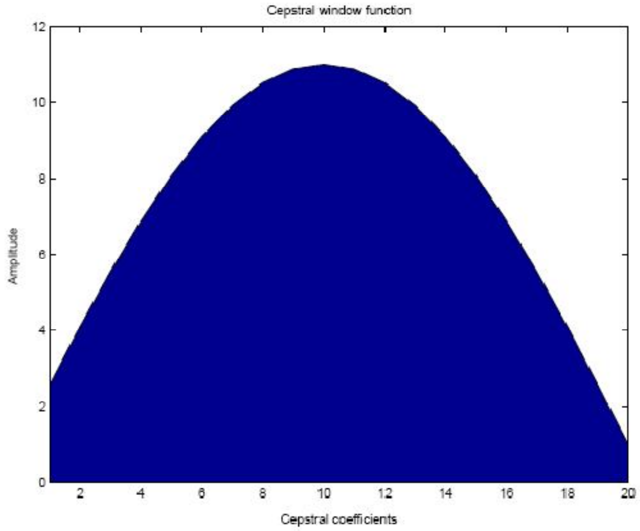


Figure 12: The cepstral window function given by Eq. (6). which is defined as

$$\Delta c_w(m,n)=\frac{1}{2}\left[c_w(m+1,n)-c_w(m-1,n)\right], \qquad 1\leqslant n\leqslant Q, \tag{8}$$

where m indicates the number of frames [7, 9]. For each phoneme sound s, a frame based analysis is performed by writing a Matlab program where we define a function

function x=hmmfeatures(s)

This yields the features, namely, the first 20 weighted LPCC, $c_w(n)$, and delta weighted LPCC (delta WLPCC), $\Delta c_w(n)$. In Fig. 13 we have shown the graphical plots of these features for the vowel /a/ spoken by a female informant.

In Fig. 14 and Fig. 15, we have only shown the LPCC, WLPCC, and the delta WLPCC graphs for the 16th frame of vowels /ɒ/ and /a/ spoken by a male and a female Assamese speaker. These graphs clearly indicates that the WLPCC as well as delta WLPCC values are significantly higher compared to their corresponding LPCC values and hence more feasible to consider as relevant features for subsequent application.

3. RESULTS AND CONCLUSIONS

LPCC is a very compact representation of the spectral envelop of speech signal. This representation of speech signal is highly beneficial in comparison with the other techniques such as discrete Fourier transform (DFT) or fast Fourier transform (FFT) in the sense that the number of coefficients in the output is based on the order of the FFT while the LPCC method depends only on the order of the pole p of the all-pole model. Physically, the LPCC coefficients reflect the difference of the biological structure of human vocal tract. This is evident from the distinct differences in the cepstral coefficient graphs of male and female informants for the vowels /a/, /ɛ/, and /u/. Thus, the LPCC method technically provides a reliable way to identify gender of Assamese speaker. Our cepstral analysis of vowel phonemes suggested that by looking at the cepstral coefficient graphs for the vowels/a/, /ɛ/, and /u/, one can readily distinguish the male and female informants.

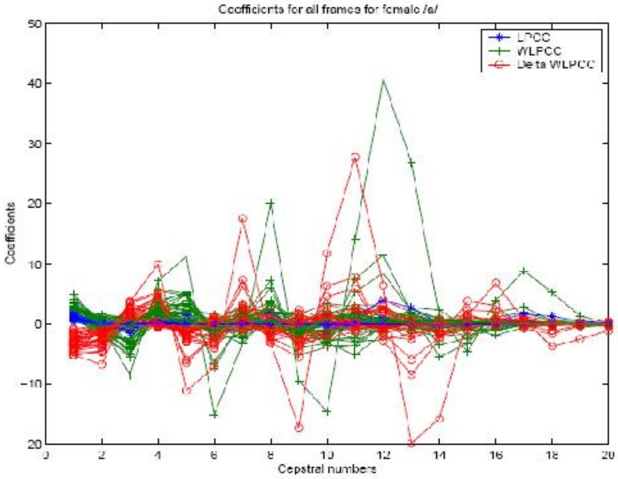
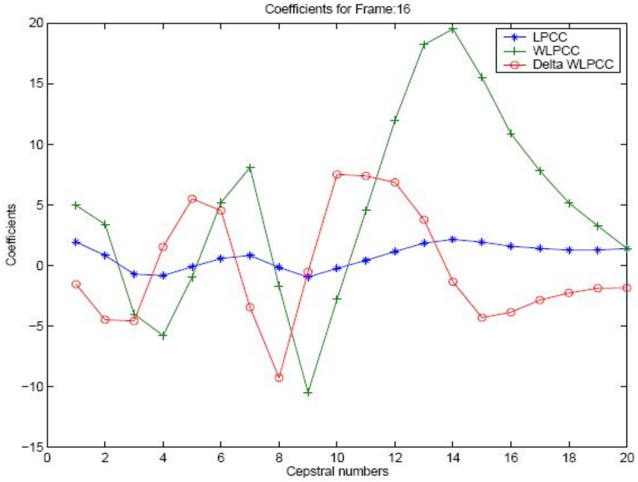


Figure 13: LPCC, WLPCC, and Delta WLPCC plots for all the frames of vowel /a/spoken by a female informant.

We have calculated the weighted LPCC and the delta weighted LPCC of each speech signal of our database. The graphical plot of these speech features indicates that the WLPCC as well as delta WLPCC values are significantly higher compared to their corresponding LPCC values. Thus, these features are more feasible than the LPCC for speech processing application. For statistical speech recognition, the WLPCC and delta WLPCC are together considered as feature vectors and stored as a 64 × 20 matrix $x =[c_w(n);\Delta c_w(n)]$. Each column of the matrix x forms the template vectors. Thus, from each speech signal one can have a 64 × 20 matrix comprising 20 template vectors. These LPC based features are potentially more feasible as can be seen from their applications in speech recognition. It can be noted that using WLPCC and delta WLPCC as features, Gao et al. [13] obtain near about 90% recognition accuracy for musical signal with 44.1 KHz sampling rate using hidden Markov model. Recently, Bhattacharjee [14] have made a comparative study on the multilayer perception based Assamese vowel phoneme recognition using LPCC as well as Mel frequency cepstral coefficients (MFCC) features. This study has been carried out both in quiet environment as well as at different level of noisy environment. It is observed that the performance of LPCC based system degrades more rapidly compared to MFCC based system under different environmental noise condition. However, under same environmental condition, when different set of speaker is used for training and testing the MLP based recognition, LPCC feature vector gives a recognition accuracy of 94.23% whereas for MFCC the recognition accuracy is 89.14%. This results support the fact that, in case of speaker independent recognition, LPCC based features is more viable than that of MFCC features.

We would like to conclude this paper by noting that the cepstral features of Assamese vowel phonemes that we have extracted here can effectively be used as input feature vectors to any statistical phoneme recognizer which can recognize the Assamese phonemes based on statistical pattern matching techniques. We leave it as a further work in near future.



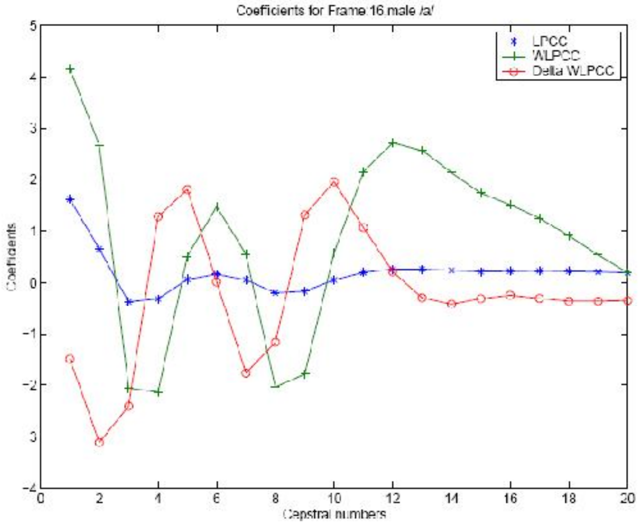
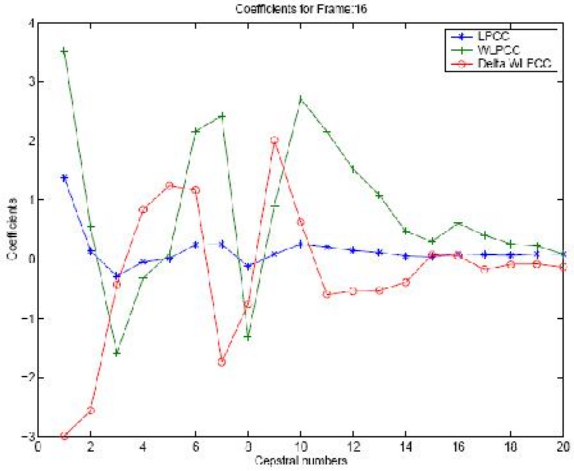


Figure 14: LPCC, WLPCC, and Delta WLPCC graphs for the 16^{th} frame of vowel /ɒ/and /a/ spoken by a male.



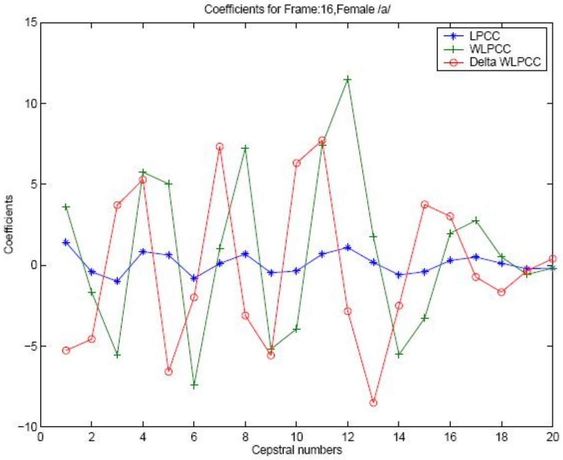


Figure 15: LPCC, WLPCC, and Delta WLPCC graphs for the 16th frame of vowel /ɒ/and /a/ spoken by a female.

REFERENCES

1. Banikanta Kakati, ***Assamese, its Formation and Development,*** 5th edition, Guwahati, India, LBS Publications, 2007.

2. G. A. Grierson, ***Linguistic survey of India,*** Vol. 3, 1903.
3. S. K. Chatterji, **The name Assam-Ahom,** *J. Asiatic Soc.*, Vol. 22, pp. 147-153, 1956.
4. Hem Baruah, ***The Red River & the Blue Hill***, Guwahati, India, 1954, Lawyer's Book Stall.
5. J. L. Flanagan, ***Speech Analysis, Synthesis, and Perception***, 2ndedition, New York, 1972, Springer-Verlag.
6. J. R. Deller, J. H. L. Hansen, and J. G. Proakis, ***Discrete-Time Processing of Speech Signals***, New York , 2000, IEEE Press.
7. L. R. Rabiner and B. H. Juang, ***Fundamentals of Speech Recognition,*** Englewood Cliffs, New Jersey, 1993, Prentice-Hall.
8. L. R. Rabiner and R. Schafer, ***Digital Processing of Speech Signals***, Englewood Cliffs, NJ, 1979, Prentice-Hall.
9. B. Gold and N. Morgan, ***Speech and Audio Processing: Processing and Perception of Speech and Music***, New York, 2000, John Wiley & Sons.
10. D. O'shaughnessy, **Interacting with computers by voice: Automatic speech recognition and synthesis**, *Proc. of the IEEE*, Vol. 91, pp. 1272-1305, 2003.
11. G. K. Vallabha and B. Tuller, **Systematic errors in the formant analysis of steady-state vowels**, *Speech Communication*, Vol. 38, pp. 141, 2002.
12. F. Jelinek, ***Statistical Methods for Speech Recognition***, Cambridge, 1998, The MIT Press.
13. S. Gao et al., **A hidden Markov model based approach to music segmentation and identification**, *in Proc. 2003, ICICS-PCM*, IEEE.
14. U. Bhattacharjee, **A comparative study of LPCC and MFCC features for the recognition of Assamese phonemes**, *Int. J. Eng. Res. Tech.* Vol. 2, pp. 1-6, 2013.