Int. J. Open Problems Compt. Math., Vol. 3, No. 5, December 2010
ISSN 1998-6262;
www.i-csrs.org

Nonlinearity and Cepstral/Mel Cepstral Measure of the Spectral Characteristics of Assamese and Bodo Phonemes

Sanjib Kr. Kalita

Department of Computer Science, Gauhati University:Kokrajhar Campus, Kokrajhar – 783370
e-mail: sanjib959@rediffmail.com

Abstract

Through this paper it is proposed to study the different phonetic characteristics of the Assamese and Bodo phonemes, studying their spectral characteristics. In the present study we have been considered six vowels of Bodo language and the eight vowels of the Assamese language. The speeches are recorded from both male and female informants(10 informants of each kind) from both the linguistic groups. In the present study, the LPC(Linear Predictive Coding) have been used to study the cepstral features of both Assamese & Bodo phonemes. The same data set have also been used to study the non-linear characteristics and the mel-cepstral features of both the languages, as discussed in section (1.1) & (1.7) respectively.

Keywords: LPC, MFCC, Non-linearity.

1 Introduction

The North-Eastern(NE) region of India is a land of mysterious mixture of varieties of languages and cultural heritages. This region can readily be considered as miniature India. Each of the seven states of this region has a large number of different linguistic groups. These groups, from different linguistic domains, remained confined within a small and limited logical domain for decades together. This causes a big social barrier, against the social homogeneity and national integrity, among these ethnic groups. It seems to fuel a feeling of isolation and exploitation among them. The linguistic scenario of the NE region is a mixture of Indo-Aryan and Sino-Tibetan family of languages. They are then branches into several forms, linguistically. The tomographical characteristics, geographical location, long exploitation by orthodox rulers and socio economic conditions are the prime factors of these branching.

In the state of Assam, which is the largest state of NE region, two major linguistic groups are – the Assamese community and the Bodo community. The mother tongue of the two communities are – “Asamiya” spoken by the Assamese people and the “Boro” spoken by the Bodo people. The Bodos are mostly bilingual. People from other ethnic groups mostly uses Assamese as a lingua franca during conversation. Hindi language is also a lingua franca for non-Bodos & non-Assamese people.

1.1 Non-linear spectral characteristics of assamese phonemes

The word ‘non-linear’ is used, almost, in all fields of science today. In different fields we have different non-linear theories, non-linear electronic devices etc. In voice science, the vocal fold vibration is described as a non-linear dynamic system [13]. There are several non-linearities that are involved in the process of vocal fold vibration [14] such as –

1. Non-linear stress-strain characteristics of vocal fold tissue.
2. Strong restoring forces at collision of the fold.
3. Highly non-linear dependence of the airflow on glottal area.

Speech is a non-linear phenomenon [15,16]. The information provided by the cepstral measures of vowel recognition through LPC (linear predictive coding) analysis is not enough to ascertain the degree of non-linearity present in the pronunciation of vowels. The fitting of polynomial, using matrix method, is a faster method to study the non-linearity of the vowels. In the present study, the analysis & synthesis of Assamese vowels are made studying their cepstral features and formant characteristics through Linear Predictive Coding (LPC) [1].

To study the degree of non-linearity, the formant frequencies are subjected to fit a polynomial of degree p, as described by equation (1.1)

$$Y=b_0+b_1x+b_2x_2+ \ldots\ldots\ldots + b_px_p \qquad (1.1)$$

Where b_0, b_1, b_2 ………. b_p are coefficients to be determined with the help of the following matrix method.

The equation (1.1) is represented in the matrix form as given by (1.2)

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ : \\ : \\ : \\ b_p \end{bmatrix} = \begin{bmatrix} n & \Sigma x & \Sigma x^2 & \Sigma x^3 & \ldots \Sigma x^p \\ \Sigma x & \Sigma x^2 & \Sigma x^3 & \Sigma x^4 & \ldots \Sigma x^{p+1} \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 & \Sigma x^5 & \ldots \Sigma x^{p+2} \\ : & : & : & : & \ldots : \\ : & : & : & : & \ldots : \\ : & : & : & : & \ldots : \\ \Sigma x^p & \Sigma x^{p+1} & \Sigma x^{p+2} & \Sigma x^{p+3} & \ldots \Sigma x^{2p} \end{bmatrix} \begin{bmatrix} \Sigma x \\ \Sigma xy \\ \Sigma x^2 y \\ : \\ : \\ : \\ \Sigma x^p y \end{bmatrix}$$

Following this matrix method, the values of b_0, b_1, b_2, b_3 and b_4 etc. are estimated. The range of variation of the values of these coefficients for **male and female as** obtained in the present investigation, are given in Table 1.1 and Table 1.2 respectively. From these tables, it is seen that the range of variation of the values of the coefficients b_4 lies between -0.031 and 0.096 i.e. $-0.031<b_4<0.096$ (for male) and -0.029 and 0.038 i.e. $-0.029<b_4<0.038$ (for female). As b_4 is very very small, so the x^4 term of the polynomial could be neglected. Thus, the equation for the representation of the formant frequency and amplitude is non-linear with the degree of non-linearity being three.

Similarly, the degree of non-linearity of formant frequency and cepstral coefficient can be obtained using the same set of matrix equations

1.1.1 Selection of Informants

Informants are selected from different age groups. The selection of informants is done on the basis of their audition performances for assessing correct pronunciation of vowels and nasal sounds. The audition performance is evaluated by an acoustic phonetic expert. In the present study 10 informants (5 male and 5 female) from each of the Assamese and Bodo language are selected.

1.2 Non-Linear Spectral Characteristics of Bodo Phonemes

To study the degree of non-linearity of Bodo vowels, we used these formant frequencies and the coefficients of the polynomial equation (1.1) are estimated following same techniques as done in case of Assamese vowels. The range of variation of coefficients b_o, b_1, b_2, b_3, and b_4, for both male and female informants, are given in the Table 1.3 and Table 1.4

From the Table 1.3 and Table 1.4, it is seen that the range of variation of the coefficient b_4 lies between $-1.27<b_4<0.00005$ (for male) and $-0.00006<b_4<0.0089$(for female). Thus, the x^4 term of the polynomial can be neglected, as b_4 is very very small. Hence, the equation representing the formant frequency characteristics of Bodo vowels is nonlinear in nature with a degree of nonlinearity three.

1.3 Non-linear characteristics of assamese phonemes

To study the degree of non-linearity of Assamese vowels, we used the cepstral coefficients and the coefficients of the polynomial equation (1.1). The range of variation of coefficients b_o, b_1, b_2, b_3, and b_4, for both the male and female informants, are given in the Table 1.5 and Table 1.6

From the Table 1.5 and Table 1.6, it is seen that the range of variation of the coefficient b_4 lies between $-0.00003<b_4<0.00245$ (for male) and $-0.000006<b_4<0.000007$ (for female) which is insignificant. Thus, the x^4 term of the polynomial can be neglected. Hence, the equation representing the cepstral coefficient characteristics of Assamese vowels are nonlinear in nature with a **degree of nonlinearity three**.

1.4 Non-linear characteristics of bodo phonemes

Like Assamese vowels to study the degree of non-linearity of Bodo vowels, we used the cepstral coefficients and the coefficients of the polynomial equation (1.1). The range of variation of coefficients b_o, b_1, b_2, b_3, and b_4, for both the male and female informants, are given in the Table 1.7 and Table 1.8

From the Table 1.7 and Table 1.8, it is seen that the range of variation of the coefficient b_4 lies between $-0.000008<b_4<0.00006$ (for male) and $-0.000007<b_4<0.000011$ (for female) which is insignificant. Thus, the x^4 term of the polynomial can be neglected. Hence, the equation representing the cepstral coefficient characteristics of Bodo vowels are nonlinear in nature with a **degree of nonlinearity three**.

1.5 Linear predictive coding (LPC)

The basic idea behind the LPC model is that – a given speech signal, at the time n, s[n], can be represented, approximately, as a linear combination of the past p speech samples [1] such that

$$S(n) \approx a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p s(n-p) \qquad (1.3)$$

The equation can be further transformed by including an excitation term Gu(n) as:

$$S(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n) \tag{1.4}$$

Where a_i (i =1,2,3,……., p) are coefficients and assumed to be constant over the speech analysis frame. u(n) is the normalized excitation and G is the gain of excitation. The mean short time predictor error per frame is given as [2]:

$$E_n = \sum_m e_n^2(m) = \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2 \tag{1.5}$$

Where $s_n(m)$ is the segment of speech selected in the neighborhood of a sample

$s_n(m) = s(m+n)$. The value of the coefficients a_k , that minimizes the error rate E_n , can be obtained, considering $\frac{dE_n}{da_i} = 0$ where i=1,2,3, ………. ,p, as given in equation (1.4),

$$\sum_m S_n(m-i) S_n(m) = \sum_{k=1}^{p} a_k^{/} \sum_m s_n(m-i) s_n(m-k), \quad 1 \le l \le p \tag{1.6}$$

where $a_k^{/}$ are the values of a_k that minimizes E_n .

Defining, $\Phi_n(i,k) = \sum_m s_n(m-i) s_n(m-k)$, the equation (1.6) can be rewritten as :

$$\sum_{k=1}^{p} a_k \Phi_n(i,k) = \Phi_n(i,0), \; i = 1,2,3,......p \tag{1.7}$$

This is a system of p equations with p variables. The equations can be solved to find a_k coefficients for the segment $s_n(m)$. Thus, E_n can be represented as

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^{p} a_k \sum_m s_n(m) s_n(m-k) \tag{1.8}$$

and in compact form, E_n can be further reduced to the form

$$E_n = \Phi_n(0,0) - \sum_{k=1}^{p} a_k \Phi_n(0,k) \qquad (1.9)$$

Now, the values $\Phi_n(i,k)$ have to be obtained for $1 \le i \le p$ and $1 \le k \le p$ and the a_k coefficients are obtained by solving equation (1.7). The solution of the equation (1.9) is obtained following the autocorrelation method [2].

Table 1.1 *: Range of variation of the coefficients b_0, b_1, b_2, b_3, b_4 of the polynomial fitting for formant frequencies of Assamese vowels corresponding to male informants.*

Vowels	Coefficients				
	b_0	b_1	b_2	b_3	b_4
/a/	$0.649<b_0<2.328$	$-3.819<b_1<-0.825$	$0.460<b_2<2.242$	$-0.453<b_3<-0.035$	$-0.0062<b_4<0.031$
/aa/	$1.10<b_0<6.526$	$-11.625<b_1<1.392$	$0.694<b_2<6.841$	$-1.408<b_3<-0.177$	$0.0002<b_4<0.096$
/e/	$-2.689<b_0<-1.804$	$1.497<b_1<4.350$	$-1.902<b_2<-0.274$	$0.029<b_3<0.418$	$-0.031<b_4<-0.0004$
/eai/	$-2.529<b_0<-1.955$	$2.682<b_1<4.074$	$-1.759<b_2<-0.658$	$0.061<b_3<0.388$	$-0.029<b_4<0.0002$
/ea/	$-1.542<b_0<1.133$	$-1.868<b_1<2.711$	$-1.080<b_2<1.088$	$-0.188<b_3<0.247$	$-0.019<b_4<0.011$
/o/	$2.675<b_0<2.566$	$-4.866<b_1<-0.294$	$0.236<b_2<2.984$	$-0.619<b_3<-0.017$	$-0.0002<b_4<0.043$
/u/	$1.083<b_0<2.075$	$-3.592<b_1<-1.745$	$0.245<b_2<2.103$	$-0.424<b_3<-0.081$	$-0.002<b_4<0.029$
/ou/	$-1.870<b_0<2.233$	$-3.801<b_1<2.698$	$-0.683<b_2<2.236$	$-0.452<b_3<0.065$	$-0.0004<b_4<0.031$

Table 1.2 *: Range of variation of the coefficients b_0, b_1, b_2, b_3, b_4 of the polynomial fitting for formant frequencies of Assamese vowels corresponding to female informants.*

Vowels	Coefficients				
	b0	b1	b2	b3	b4
/a/	0.345 < b0 < 1.075	-1.282 < b1 < -0.211	0.170 < b2 < 0.649	-0.057 < b3 < 0.012	-0.004 < b4 < -0.006
/aa/	0.416 < b0 < 1.100	-1.184 < b1 < -0.259	0.333 < b2 < 1.035	-0.244 < b3 < -0.059	0.004 < b4 < 0.019
/e/	-1.920 < b0 < -1.608	2.225 < b1 < 2.689	-0.665 < b2 < 0.495	0.042 < b3 < 0.056	0.0006 < b4 < 0.0008
/eai/	-4.116 < b0 < 1.344	-3.001 < b1 < 2.621	-0.598 < b2 < 2.307	-0.477 < b3 < 0.040	0.002 < b4 < 0.031
/ea/	-4.117 < b0 < -2.705	4.889 < b1 < 6.509	-2.652 < b2 < -2.161	0.444 < b3 < 0.498	-0.031 < b4 < 0.033
/o/	-1.016 < b0 < 0.483	-0.801 < b1 < 2.014	-1.089 < b2 < 0.519	-0.046 < b3 < 0.295	-0.025 < b4 < -0.0004
/u/	1.583 < b0 < 2.500	-4.287 < b1 < -2.816	1.840 < b2 < 2.549	-0.526 < b3 < 0.402	0.030 < b4 < 0.038
/ou/	1.895 < b0 < 2.000	-3.377 < b1 < -3.193	1.936 < b2 < 2.037	-0.414 < b3 < 0.392	0.026 < b4 < 0.029

Table 1.3 : *Range of variation of b_o, b_1, b_2, b_3, and b_4 corresponding to Bodo male informants*

vowels	bo	b1	b2	b3	b4
/a/	-0.49 < bo < 26.16	-32.87 < b1 < 1.02	-1.01 < b2 < 14.94	-2.73 < b3 < .0095	-0.002 < b4 < 0.18
/e/	-0.91 < bo < 4.26	-6.78 < b1 < 2.36	-1.25 < b2 < 4.09	-0.83 < b3 < 0.29	-0.02 < b4 < 0.06
/i/	-3.68 < bo < 1.69	-2.13 < b1 < 5.60	-1.89 < b2 < 1.16	-0.20 < b3 < 0.31	-0.02 < b4 < 0.00005
/o/	0.77 < bo < 3.93	-7.07 < b1 < -1.32	0.83 < b2 < 4.35	-6.87 < b3 < -0.07	-0.0002 < b4 < 0.06
/u/	-65.82 < bo < 4.01	-7.28 < b1 < 12.31	-71.96 < b2 < 4.26	-0.86 < b3 < 16.55	-1.27 < b4 < 0.058

/w/	-0.66<bo<0.79	-1.57<b1<1.41	-0.37<b2<1.56	-6.25<b3<0.08	-0.008<b4<0.02

***Table 1.4 :** Range of variation of b_o, b_1, b_2, b_3, and b_4 corresponding to Bodo female informants*

vowels	bo	b1	b2	b3	b4
/a/	0.33<bo<5.84	-0.90<b1<0.98	-0.03<b2<5.82	-1.19<b3<0.06	-0.007<b4<0.08
/e/	0.97<bo<2.28	-2.61<b1<-0.94	0.53<b2<1.35	-0.20<b3<-0.05	-0.00006<b4<0.0089
/i/	-2.23<bo<4.67	-8.43<b1<3.16	-0.85<b2<4.08	-0.95<b3<0.09	-0.0016<b4<0.063
/o/	0.59<bo<2.92	-4.98<b1<-0.96	0.59<b2<3.03	-0.61<b3<-0.06	0.0008<b4<0.04
/u/	-1.13<bo<2.51	-4.80<b1<2.46	-0.35<b2<3.07	0.63<b3<0.05	-0.0025<b4<0.04
/w/	-0.02<bo<2.43	-3.60<b1<0.16	-0.14<b2<2.05	-0.36<b3<0.0003	-0.0016<b4<0.02

***Table 1.5 :** Range of variation of the coefficients b_o, b_1, b_2, b_3, and b_4 of the polynomial fitting for cepstral coefficient of Assamese vowels (for male)*

vowels	bo	b1	b2	b3	b4
/a/	0.0018<b0<0.0046	-0.0038<b1<-0.0009	0.00006<b2<0.00089	-0.00007<b3<0.00001	-0.00003<b4<0.000002
/aa/	0.0018<b0<0.009	-0.0011<b1<-0.0009	0.00016<b2<0.00023	-0.00002<b3<-0.00001	0.0000006<b4<0.0000008
/e/	-0.0014<b0<0.0019	-0.0011<b1<0.0024	-0.00073<b2<0.00022	-0.00002<b3<0.00009	-0.000004<b4<0.0000007
/u/	0.0018<b0<0.0057	-0.003<b1<0.0045	0.00005<b2<0.00051	-0.00009<b3<0.00003	-0.000002<b4<0.000003
/ea/	0.0018<b0<0.0025	-0.0012<b1<-0.001	0.00011<b2<0.0002	-0.00002<b3<0.00002	-0.0000008<b4<0.0000006
/eai/	-0.0045<b0<0.0018	-0.0010<b1<0.0048	-0.0014<b2<0.00021	-0.00002<b3<0.00015	-0.000006<b4<0.0000006
/o/	0.0018<b0<0.0044	-0.003<b1<-0.001	0.00015<b2<0.00054	-0.0002<b3<0.000008	-0.000009<b4<0.0000012
/ou/	0.0018<b0<0.643	-0.894<b1<-0.001	0.00021<b2<0.0036	-0.00005<b3<0.052	0.0000006<b4<0.00246

***Table 1.6 :** Range of variation of the coefficients b_o, b_1, b_2, b_3, and b_4 of the polynomial fitting for cepstral coefficient of Assamese vowels (for female)*

vowels	b_o	b_1	b_2	b_3	b_4
/a/	$0.0028<b_0<0.0040$	$-0.0037<b_1<-0.0011$	$0.0011<b_2<0.0022$	$-0.0001<b_3<-0.00002$	$0.0000007<b_4<0.000005$
/aa/	$0.0018<b_0<0.0043$	$-0.0042<b_1<-0.0010$	$0.0013<b_2<0.0021$	$-0.0002<b_3<-0.00002$	$0.0000007<b_4<0.000007$
/e/	$0.0011<b_0<0.0018$	$-0.0011<b_1<0.0026$	$-0.0010<b_2<0.0022$	$-0.00002<b_3<0.00013$	$-0.000006<b_4<0.0000006$
/u/	$0.00183<b_0<0.00188$	$-0.0011<b_1<-0.0010$	$0.00020<b_2<0.00023$	$-0.00002<b_3<-0.00001$	$0.0000007<b_4<0.0000006$
/ea/	$0.0018<b_0<0.0047$	$-0.0011<b_1<0.0035$	$0.0002<b_2<0.00075$	$-0.00007<b_3<0.00006$	$0.0000005<b_4<0.000001$
/eai/	$0.0010<b_0<0.0019$	$-0.0011<b_1<0.0025$	$-0.0009<b_2<0.0022$	$-0.00002<b_3<0.00012$	$-0.000005<b_4<0.0000007$
/o/	$0.0018<b_0<0.0045$	$-0.0035<b_1<-0.0010$	$0.0002<b_2<0.0008$	$-0.00008<b_3<-0.00002$	$0.0000006<b_4<0.000002$
/ou/	$0.0018<b_0<0.0061$	$-0.0055<b_1<-0.0010$	$0.00021<b_2<0.0015$	$-0.0007<b_3<-0.00002$	$0.0000006<b_4<0.000006$

Table 1.7 : Range of variation of the coefficients b_o, b_1, b_2, b_3, and b_4 of the polynomial fitting for cepstral coefficient of Bodo vowels (for male)

vowels	b_o	b_1	b_2	b_3	b_4
/a/	$-0.015<b_0<0.017$	$-0.013<b_1<0.023$	$-0.0061<b_2<0.00032$	$-0.00003<b_3<0.00004$	$-0.000003<b_4<0.000008$
/e/	$-0.017<b_0<0.0061$	$-0.004<b_1<0.005$	$0.0004<b_2<0.00052$	$-0.00011<b_3<0.00006$	$-0.000004<b_4<0.000005$
/i/	$-0.021<b_0<0.0023$	$-0.014<b_1<-0.008$	$0.00011<b_2<0.0002$	$-0.00002<b_3<0.00002$	$-0.000008<b_4<0.000006$
/o/	$-0.0045<b_0<0.0023$	$-0.0011<b_1<0.0052$	$-0.0016<b_2<0.00023$	$-0.00003<b_3<0.00017$	$-0.000006<b_4<0.000006$
/u/	$0.0021<b_0<0.0046$	$-0.006<b_1<-0.002$	$0.00017<b_2<0.00057$	$-0.0004<b_3<0.000005$	$-0.000005<b_4<0.000007$

/w/	$-0.0014<b_0<0.0043$	$-0.894<b_1<-0.0001$	$0.00021<b_2<0.00029$	$-0.00005<b_3<0.00032$	$-0.000008<b_4<0.00006$

Table 1.8 : Range of variation of the coefficients b_o, b_1, b_2, b_3, and b_4 of the polynomial fitting for cepstral coefficient of Bodo vowels (for female)

vowels	b_o	b_1	b_2	b_3	b_4
/a/	$0.0023<b_0<0.0028$	$-0.0012<b_1<0.0023$	$-0.0014<b_2<0.0026$	$-0.0003<b_3<0.00016$	$-0.000007<b_4<0.000009$
/e/	$0.0017<b_0<0.0018$	$-0.0009<b_1<0.0013$	$0.00019<b_2<0.00031$	$-0.00003<b_3<0.00006$	$0.000002<b_4<0.000005$
/i/	$0.0013<b_0<0.0045$	$-0.0021<b_1<0.0041$	$0.0005<b_2<0.00067$	$-0.00003<b_3<0.00008$	$0.000004<b_4<0.000008$
/o/	$0.0012<b_0<0.0021$	$-0.0012<b_1<0.0023$	$-0.0006<b_2<0.0019$	$-0.00002<b_3<0.00015$	$-0.000003<b_4<0.000011$
/u/	$0.0015<b_0<0.0039$	$-0.0029<b_1<-0.0013$	$0.0005<b_2<0.0018$	$-0.00004<b_3<0.00012$	$0.000002<b_4<0.000007$
/w/	$0.0023<b_0<0.0059$	$-0.0049<b_1<-0.0012$	$0.00018<b_2<0.0016$	$-0.0002<b_3<0.00002$	$0.0000003<b_4<0.000005$

Instead of using directly the LPC coefficients as feature vectors, the cepstral coefficients, based on LPC analysis, are usually used because of their superior recognition capabilities [3]. The LPC based cepstral coefficients are described as -

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k}, 1 \le k \le p \qquad (1.10)$$

$$c_m = \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k}, \qquad m > p \qquad (1.11)$$

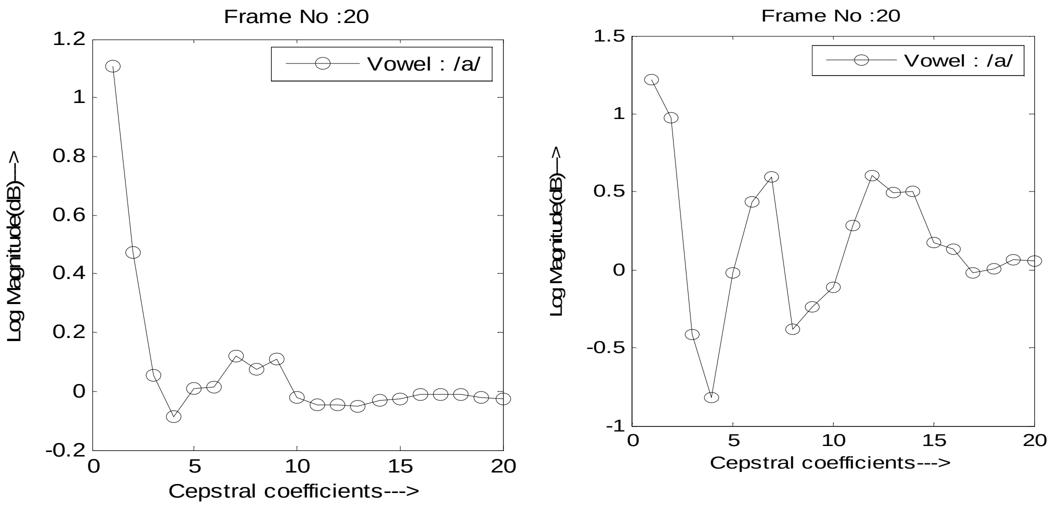
Where, c_k is an LPC based cepstral coefficient. In the present investigation, the following typical values are used –

N=240, P=10, Q=12, K=2 and G=0.316 [4].

In the present study, each digitized voice or word uttered, is divided or blocked into 32 frames of duration 31.25msec. Each frame contains 250 samples and for each frame 20 cepstral coefficients have been calculated. Approximately 12, samples are averaged to obtain one coefficient. In the present study, the 20th frame of all utterances of male and female informants have been considered to verify speaker's identity.

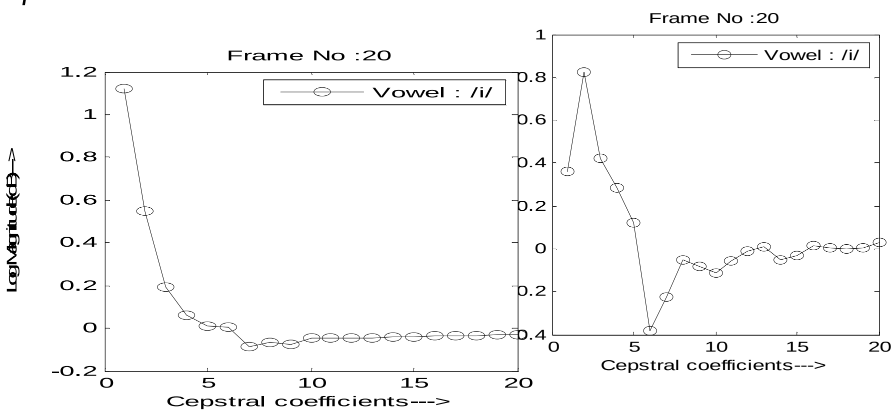
1.6 Cepstral measure of assamese and bodo vowels

The spectral characteristics of Assamese vowel /a/, corresponding to male and female informants, have been depicted in Fig. 1.1(a) and Fig 1.1(b) corresponding to the 20th frame.



(a) *(b)*

Fig. 1.1 : Cepstral characteristics of Assamese vowel /a/ corresponding to 20th speech frame for male & female informants.



(a) *(b)*

Fig 1.2 : Cepstral characteristics of Bodo vowel /i/ corresponding to 20th speech frame for male & female informants.

The spectral characteristics of Bodo vowel /i/, corresponding to male and female informants, have been depicted in Fig. 1.2(a) & (b) corresponding to the 20th frame.

1.7 Measure of mel frequency cepstral coefficients (mfcc) for assamese and bodo phonemes

The effectiveness of the speech recognition or speaker verification depends mainly on the accuracy of discrimination of speaker models, developed from speech features. The features extracted and used for the recognition process must posses high discriminative power. In the earlier method of feature extraction, preferably, used the Linear Predictive Coding(LPC) [1, 6, 8]. **Mel Frequency Cepstral Coefficient(MFCC)** analysis has been widely used in signal processing in general and speech processing in particular. This coefficient has a great success in speech recognition application [6, 7]. It is derived from the Fourier Transform of the audio clip. In this technique, the frequency bands are positioned logarithmically, whereas, in the Fourier Transform, the frequency bands are not positioned logarithmically. As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than in any other system [6]. These coefficients allow better processing of data. In the Mel Frequency Cepstral Coefficients, the calculation of the Mel Cepstrum is same as the real Cepstrum, except the Mel Cepstrum's frequency scale is warpped to keep up a correspondence to the Mel scale, proposed by Stevens[12]. The Mel scale is mainly based on the study of observing the pitch or fundamental frequency, perceived by the human. Fig.1.3 represents the algorithms involved while calculating Mel Frequency Cepstral Coefficients (MFCC).

The voice signals corresponding to the eight Assamese vowels have been recorded directly in computer through microphone. The utterances corresponding to each vowel, is recorded for the period of one second. The speech signal is then sampled at a sampling rate of 8KHz with 16-bit resolution. The spectra so obtained is analysied with the help of MFCC. The basic idea behind this model is that - cepstrum is the Fourier Transform of the log with unwrapped phase of the Fourier Transform. Mathematically, it is represented as [10, 11] –

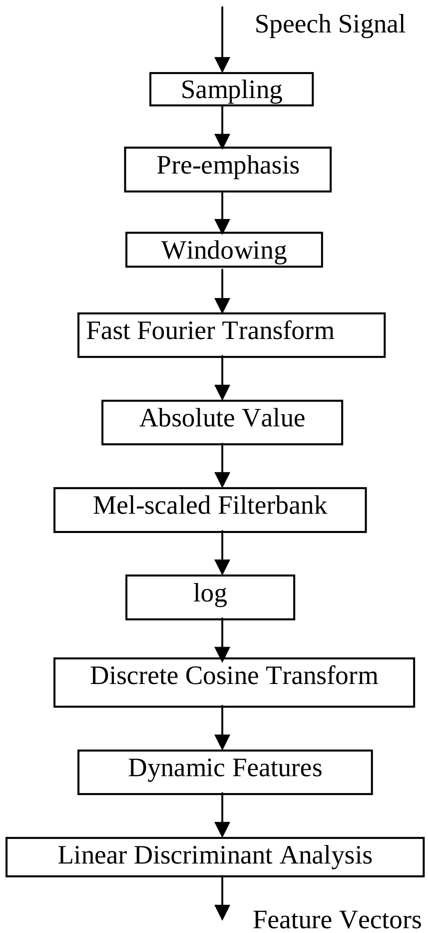


Fig. 1.3 *:Algorithm for calculating Mel Frequency Cepstral Coefficients*

$$\text{Cepstrum of signal} = FT(\log(FT(\text{the signal}))+j2\pi m) \quad (1.12)$$

Where, m is the integer, required to properly unwrap the angle or imaginary part of the complex log function. Algorithmically, it can be represented as– *Signal - FT - log - phase unwrapping - FT - Cepstrum.*The Mel scale is normally a linear mapping below 1000 Hz and logarithmically, spaced above 1000 Hz. Equation (1.13) shows the mapping of normal frequency into the Mel frequency [10].

$$m=1127.01048\log_e(1+f/700) \quad (1.13)$$

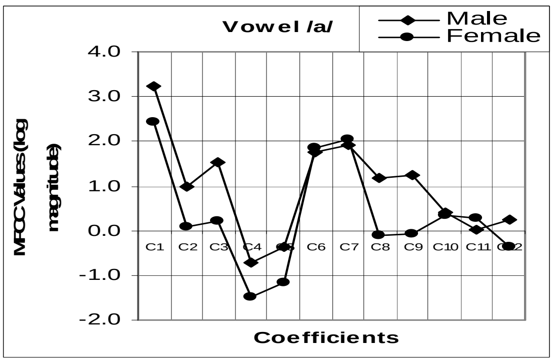
$$f=700(e^{m/1127.01084}-1) \quad (1.14)$$

The equation (1.14) is the inverse of equation (1.13), giving the original frequency.

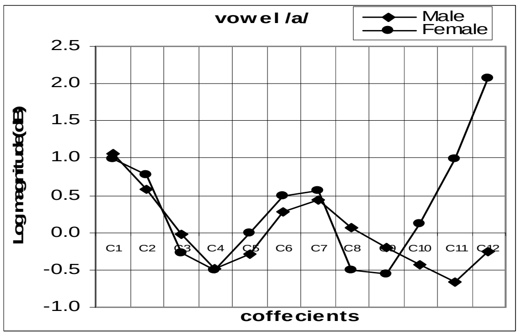
1.8 Algorithm for calculating mfcc

The following steps have been adapted for calculating the MFCC in the present study.

[i] The recorded speech signal is digitized at 8KHz sampling rate.
[ii] The digitized speech is then blocked into 32 frames and each frame containing 256 samples.
[iii]For each frame, 12 MFCC have been calculated(C1,C2, ……C12).
[iv]Frame nos 4,8,12,16,20,24,28, and 32 have been considered for the present analysis, because a spectral variation of speech features seems almost invariant (constant) within a span of 4 frames.
[v] The maximum and minimum MFCC were derived using the techniques as discussed in the text.
[vi]Matlab function 'melcepst' is used to calculate the mel cepstrum corresponding to each selected frame of the speech signal.

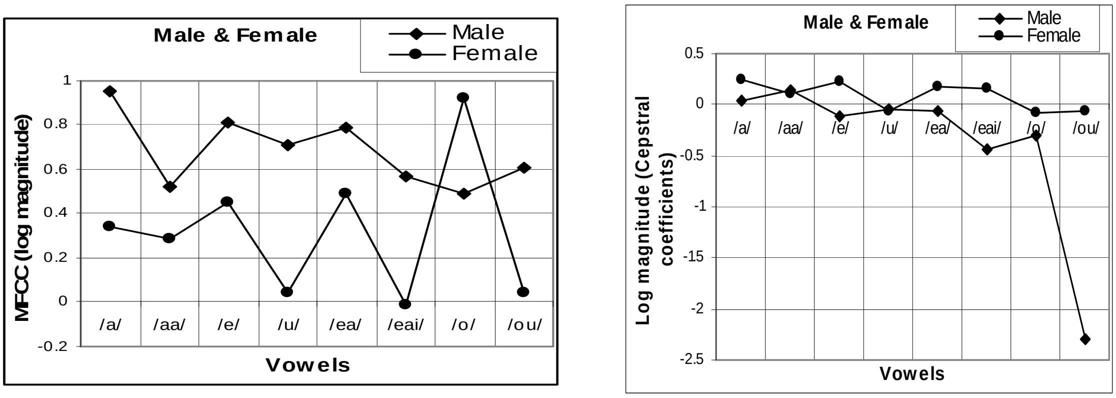


(a) MFCC plot



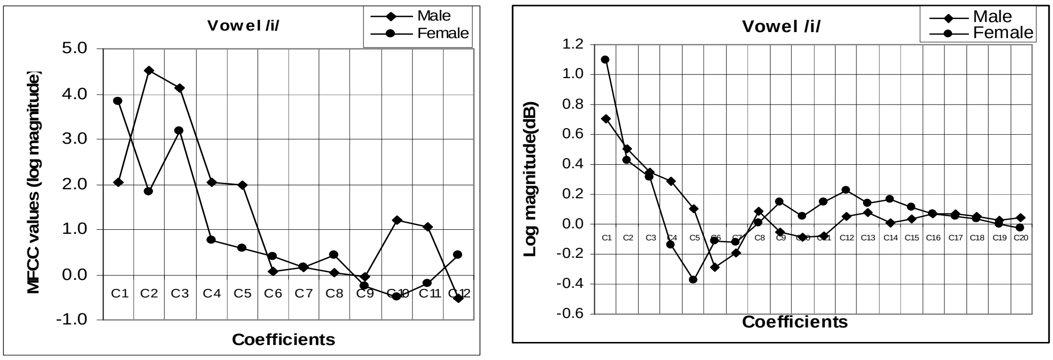
(b) Cepstral frequency plot

***Fig 1.4** : Range of variation of coefficients corresponding to Assamese male & Female considering MFCC & Linear cepstral coefficients.*



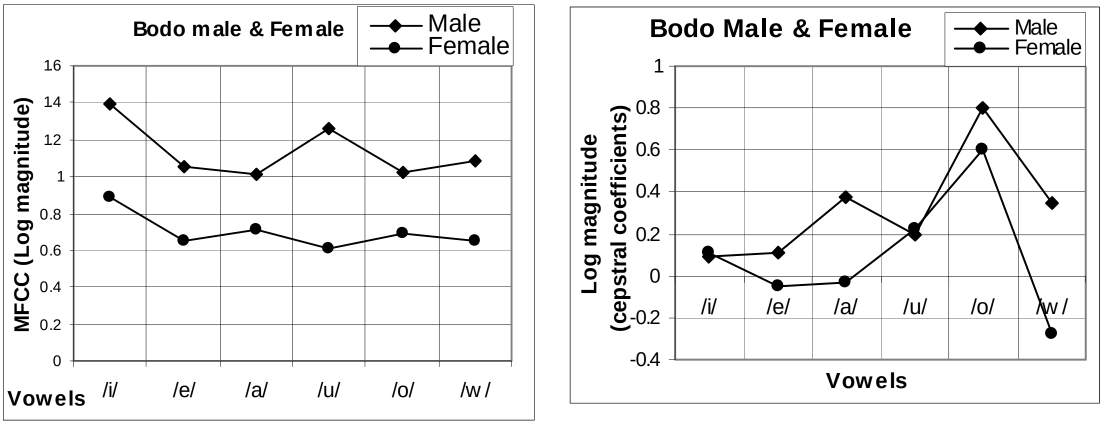
(a) Average plot of MFCC feature *(b) Average plot of cepstral*

***Fig. 1.5** Average plot of Assamese vowels corresponding to their MFCC and Linear cepstral coefficients.*



(a) MFCC plot vowel /i/ *(b) Cepstral frequency plot vowel /i/*

Fig 1.6 : Range of variation of coefficients corresponding to Bodo male & Female considering MFCC & Linear cepstral coefficients for vowel /i/



(a) Average plot of MFCC *(b) Average plot of cepstral feature*

Fig. 1.7 *Average plot of Bodo vowels corresponding to their MFCC and Linear cepstral coefficients*

1.10 : Results and discussion

1.10.1 MFCC & linear cepstral coefficients for assamese male & female

The Mel frequency cepstral coefficients and linear cepstral coefficients characteristics of eight Assamese vowels, corresponding to male and female informants, have been studied and a typical characteristics of assamese vowel /a/, is depicted in Fig. 1.4(a) and Fig 1.4(b) for male & female informants. It is seen that the cepstral frequency plot of the spectral characteristics of the assamese vowels does not reveal any clear distinction between male & female speakers. However, the MFCC plots depict a clearity of the speaker's identification as shown in Fig 1.4(a) & Fig 1.4(b). The same is further confirmed through the average MFCC plot, as shown in Fig 1.5(a) & Fig 1.5(b). Thus, it is concluded, on the basis of the present study, that MFCC is a better technique for sex verification with respect to linear cepstral measure.

1.9 MFCC & linear cepstral coefficients for bodo male & female

The Mel frequency cepstral coefficients and linear cepstral coefficients characteristics of six Bodo vowels, corresponding to male and female informants, have been studied and a typical characteristics of vowel /i/, is depicted in Fig. 1.6(a) and Fig 1.6(b) for male & female informants.

As depicts in Fig. 1.6(a), Fig 1.6(b), Fig 1.7(a) and Fig 1.7(b), it is clearly observed, that, like the Assamese phonemes, the Bodo phonemes also show a clear distinction between the spectral characteristics of Male & Female while using MFCC. It is thus concluded in the present study that MFCC measure may be a better technique for speech & speaker verification.

2. Open Problem

In this paper, we consider only vowels of two regional languages used in north eastern region of India. Speech recognition is a challenging task. To develop an artificial machine that can recognize regional languages, researchers should emphasis on word recognition, sentence recognition, word boundary detection etc.

References

[1] J. Barman, S. Kalita and P.H. Talukdar, Feature Extraction of Bodo Vowels Through LPC-Analysis, *Proceedings of Frontiers of Research on Speech and Music(FRSM),* (2004), 250-257

[2] L. Rabiner and B.H. Jwang, *Fundamentals of Speech Recognitions*, Prentice Hall, New Jersey, 1993.

[3] L. R. Rabiner, K.C. Pan and F.K. Soong, On the performance of isolated word speech recognition using vector quantization and temporal energy contours, *AT and T Bell Lab. Tech. J.*, 63(1984), 1245–1260.

[4] C.H. Lee, L.R. Rabiner, R. Pierarccini and J.G. Wilpon, *Acoustic modeling for large vocabulary speech recognition. Computer Speech and language*, 4(1990), 127-165.

[5] N. Awasthy, J.P. Saini, and D.S. Chauhan, Spectral analysis of speech : A new Technique, *International Journal of Signal Processing*, 2(1)(2006).

[6] S. Davis and P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Trans. ASSP*, 28 (1980), 357-366.

[7] Joseph, W., P., Signal modeling techniques in speech recognition, *Proceedings of the IEEE*, 81(9)(1993), 1215-1247.

[8] P.H. Talukdar, U. Bhattacharjee, C.K. Goswami, and J. Barman, Cepstral Measure of Boro Vowels through LPC-Analysis, *Journal of the CSI*, 34(2004).

[9] “Speaker Recognition”, Project Report, *Website: www.softwarepractice.org//mediawiki*

[10] B. P. Bogert, M. J. R. Healy and J. W. Tukey, The quefrency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking, *Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15,* (1963), pp 209-243. New York: Wiley.

[11] D. G. Childers, D. P. Skinner and R.C. Kemerait, The Cepstrum: A Guide to Processing, *Proceedings of the IEEE,* 65(10)(1977), 1428-1443.

[12] S. Stevens, J. Volkmann and E. Newman, A Scale for the Measurement of the Psychological Magnitude of Pitch, *Journal of the Acoustical Society of America*, *8* (1937), 185–90.

[13] L. Welling and H. Ney, Formant Estimation for Speech Recognition, *IEEE Transactions on Speech and Audio Processing,* 6 (1998), 36-48.

[14] R. T. Ingo, What is meant by non-linear and Interactive in voice Science?, *Journal of Singing,* 60(3) (2004), 277-278.

[15] G. Chollet, A. Esposito, M. Faundez-Zanuy and M. Marinaro, *Nonlinear Speech Modeling and Applications: Advanced Lectures and Revised Selected Papers*, Springer, 2005.

[16] M.F. Zanuy, L. Janer, A. Esposito, A. Satue-Villar and J. Roure, Nonlinear Analyses and Algorithms for Speech Processing: International Conference on Non-Linear Speech Processing, *NOLISP*, Barcelona, Spain, 2005.