Cepstral & Mel-Cepstral Frequency Measure of Sylheti phonemes

Suchismita Sinha, Jyotismita Talukdar, Purnendu Bikash Acharjee, P.H.Talukdar
Dept. of Instrumentation & USIC, Gauhati University, Assam.
phtassam@gmail.com

ABSTRACT

This paper deals with the different spectral features of Sylheti language, which is the major link language of southern part of North-East India and Northern region of Bangladesh. The parameters condidered in the present study are the cepstral coefficients, Mel-Cepstral coefficients and LPC . It is found that the cepstral measure is an efficient way for sex identification and verification for sylheti native speakers. Further, the vowel sound and their spectral feature dominate the features of the sylheti language.

Keywords : Cepstral coefficients, Mel-Cepstral coefficients, LPC, Pitch, Formant frequency

1. INTRODUCTION

Sylheti , native name Siloti, Bengali name Sileti, is the language of Sylhet, the northern region of Bangladesh and also spoken in parts of the north-east Indian states like: Assam(the Barak valley),and Tripura. Sylheti language is considered as a dialect of Bengali and Assamese[11]. This language has many common features with Assamese , including the existence of a larger set of fricatives than other East Indic languages. Sylheti language is written in the Sylheti Nagri script which has 5 independent vowels, 5 dependent vowels attached to a consonant letter and 27 consonants.Sylheti is quite different from standard Bengali, in its sound system, the way in which its words are formed   and   in   its   vocabulary. Unfortunately, due to lack of proper attention given to this language and increasing popularity of the Bengali and Assamese language among the common mass , which might be due to socio-economic and political reasons, this century old language is gradually dying out. But it has to be admitted that once it the only link language between Assam, Bangladesh and Bengal. Through this paper an attempt has been given to explore the different features of Sylheti language.

In the present study the analysis of cepstral co-efficients, has been done to explore the structural & architectural beauty of sylheti language. The Cepstral co- efficients allow to extract the similarity between two Cepstral feature vectors. They are considered as important features to separate intraspeaker variability based on age, emotional status of an individual speaker of a language [1].

The extraction of information from speech signal has been a common way towards the study of the spectral characteristics of the utterances of the phonemes of a language. One of the most widely used methods of spectral estimation in signal and speech processing is linear predictive coding(LPC). LPC is a powerful tool used mostly in audio signal processing and speech   processing   technique[2].   The spectral envelope of a digital signal of speech in compressed form are represented by using the information of a linear predictive model. It is a useful speech analysis technique for encoding quality speech sound at a low bit rate that provides a way for estimation of speech parameters, namely,   cepstral   coefficients,   formant frequencies   and   pitch   like   cepstral features, Mel-Cepstral features,

Formant analysis and LPC analysis.[2,3]

2. ESTIMATION OF LPC BASED CEPSTRAL CO-EFFICIENTS

The different steps involved in the present work includes the following:

1) Speakers have been selected randomly from the sylheti speaking area .e Barak Valley, karimganj, Hailakandi, and Indo Bangladesh border areas.

2) Speech has been recorded using Cool Edit Pro 2.0 with respect to different age groups i.e 14yrs-21yrs, 22yrs-35yrs and 36yrs-50yrs.

3) The recorded speech signals are then sampled at a sampling frequency of 8 KHz

$$C[1] = a[1]$$
$$C[n] = a[n] + \sum (m/n)a[m].c[n-m], 2 \le n \le p$$
$$C[n] = \sum (n-m/n) \, a[m] \, c[n-m] \, , \, n>p$$

-------------------- (1.0)

The cepstral analysis is generally used in the field of signal processing and particularly used in speech processing. As already mentioned speech signals are digitized at the sampling rate of 8KHz per second. Each of the spectra is divided into 32 frames, where every frame contains 250 samples. The cepstral coefficients of eight sylheti vowels namely,'a', 'aa', 'i ', 'ii ', 'u', 'uu', 'e', 'o', have been calculated for both male and female utterances. The maximum and minimum cepstral coefficient values corresponding to the 16th frame which is a middle frame, for male and female utterances of different age groups have been given in Table 1.0

4) The sampled speech signals have been divided into 32 frames and for each frame the maximum and minimum cepstral coefficients have been calculated corresponding to female and male of different age groups.

In the present study, the cepstral analysis of eight sylheti vowels have been made by the technique as proposed by Rabiner and Juang [3]. From the pth order Linear Predictor Coefficients a[i], the LPC cepstral coefficients c[i] are computed by the following equation (1.0) .

and Table 2.0. The plots for the utterances of the eight sylheti vowels have been shown in Fig 1.0 and Fig 2.0. Also the comparative plots of male and female utterances have been shown in Fig 3.0.

To determine the cepstral coefficients, Matlab7.0 Data acquisition Toolbar which works elegantly with Windows XP is used. The cepstral coefficients so obtained from LPC model seems to be more robust and representing more reliable features for speech recognition than LPC coefficients.

In my study, these co-efficients have been derived and analyzed to make an in depth study of the spectral characteristics of the Sylheti phonemes.

Table 1.: Range of variations of cepstral co-officients of eight sylheti phonemes. corresponding to sylethi female utterance

| Age groups → Vowels ↓	14yrs - 21yrs	22yrs - 35yrs	36yrs - 50yrs
a	-0.77 to 1.58	-0.20 to 1.24	-1.90 to 3.43
aa	-1.65 to 4.3	-6.43 to 2.6	-0.58 to 1.5
i	-0.68 to 1.37	-1.33 to 1.47	-0.93 to 1.32
ii	-0.87 to 1.44	-1.0 to 1.60	-0.4 to 1.04
u	-0.48 to 1.27	-0.26 to 1.26	-0.58 to 1.07
uu	-0.64 to 1.34	-0.07 to 1.14	-0.63 to 1.38
e	-3.43 to 2.54	-3.75 to 11.12	-1.89 to 1.87
o	-1.88 to 1.35	-0.56 to 1.68	-0.74 to 1.50

Table 2: Range of variations of cepstral co-officients of eight sylheti phonemes. Corresponding to male utterances

| Age groups → Vowels ↓	14yrs - 21yrs	22yrs - 35yrs	36yrs - 50yrs
a	-0.626 to 1.62	-1.44 to 4.14	-0.44 to 1.17
aa	-1.43 to 2.08	-1.21 to 1.87	-0.83 to 2.19
i	-1.12 to 1.77	-0.34 to 1.45	-0.5 to 1.48
ii	-0.53 to 1.50	-0.43 to 1.36	-0.27 to 1.24
u	-1.19 to 4.34	-0.09 to 1.09	-0.65 to 1.16
uu	-0.09 to 1.13	-0.1 to 1.18	-0.62 to 1.68
e	-2.32 to 2.44	-1.56 to 2.94	-0.64 to 1.67
o	-0.37 to 1.35	-2.03 to 1.94	-0.46 to 1.58
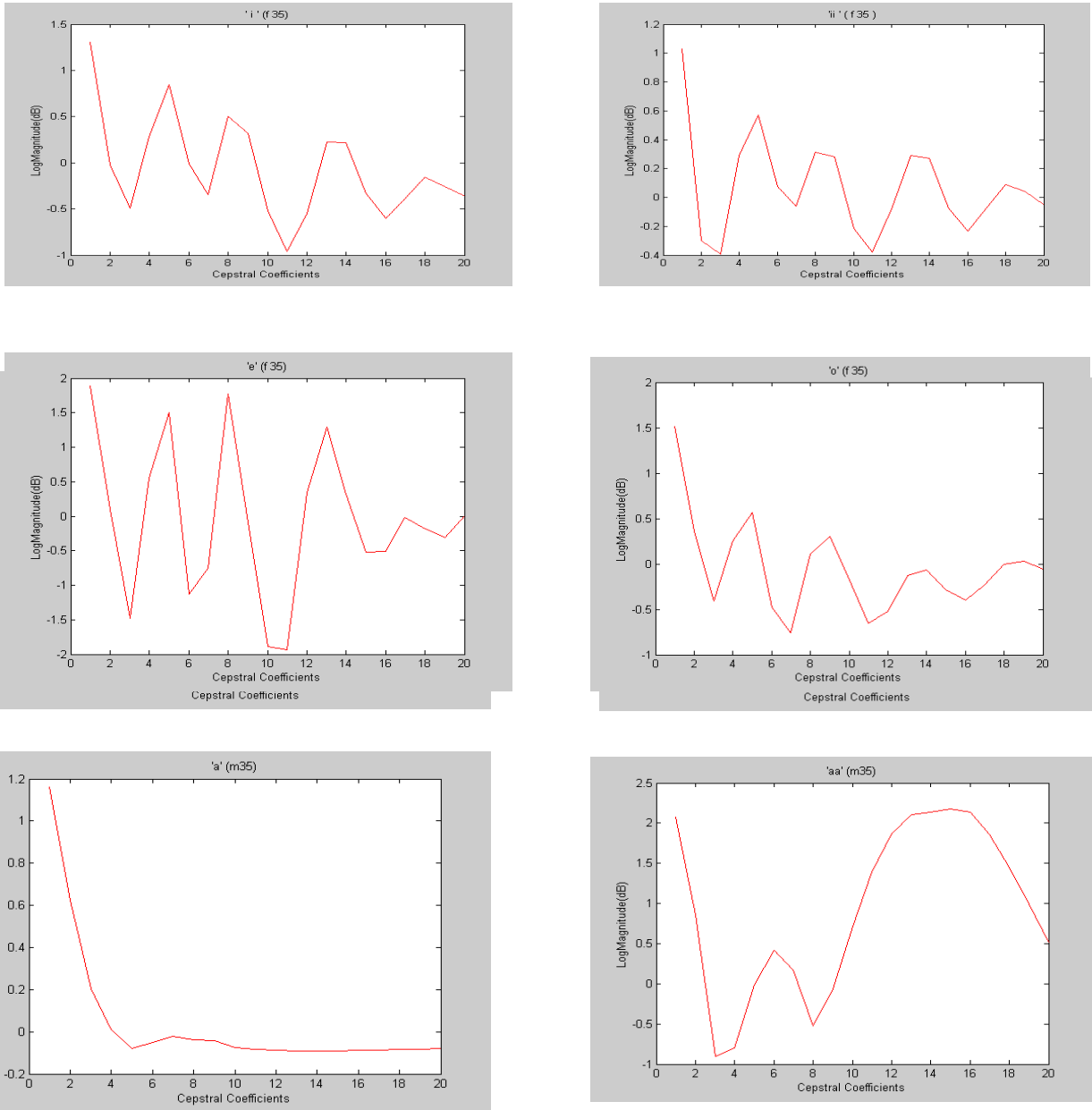
Figure 1: Cepstral coefficients extracted from the 16th frame of female utterances for the eight sylheti vowels
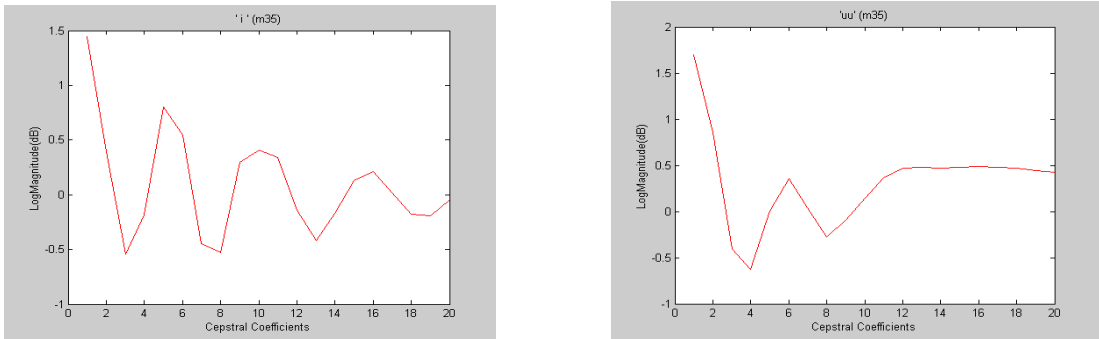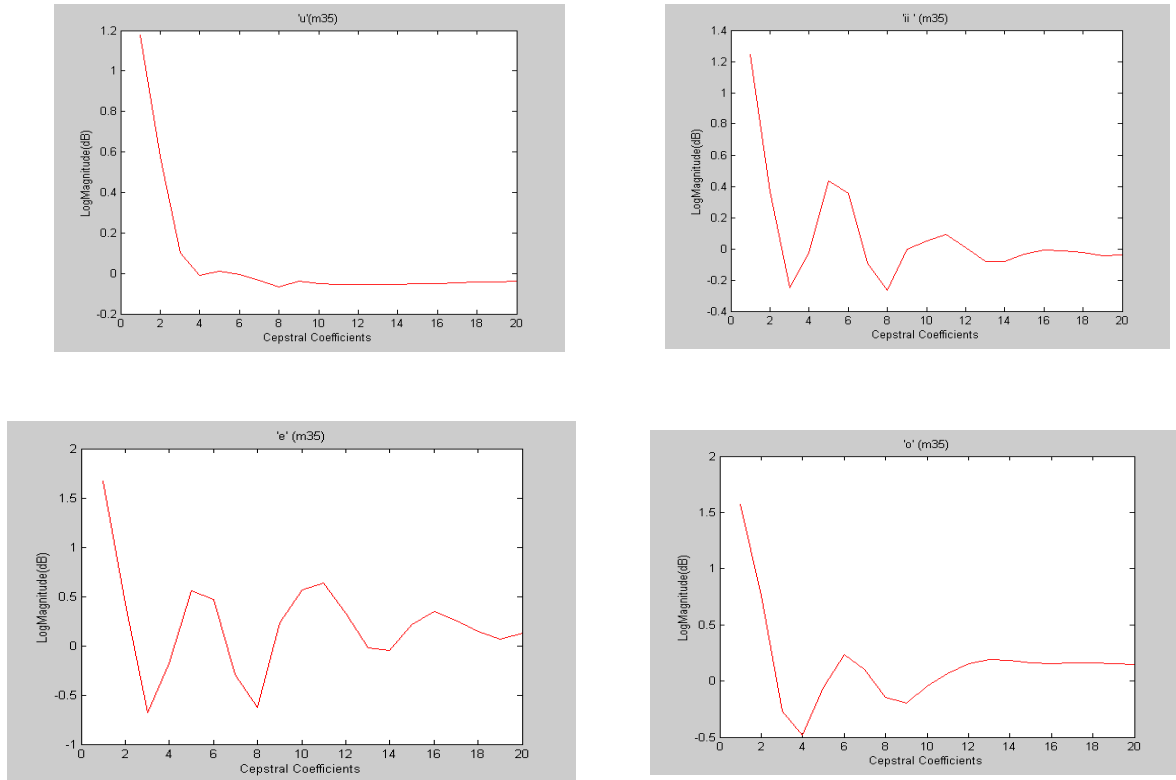
Figure   2:. Cepstral coefficients extracted from the 16th frame of male utterances for   the eight sylheti vowels
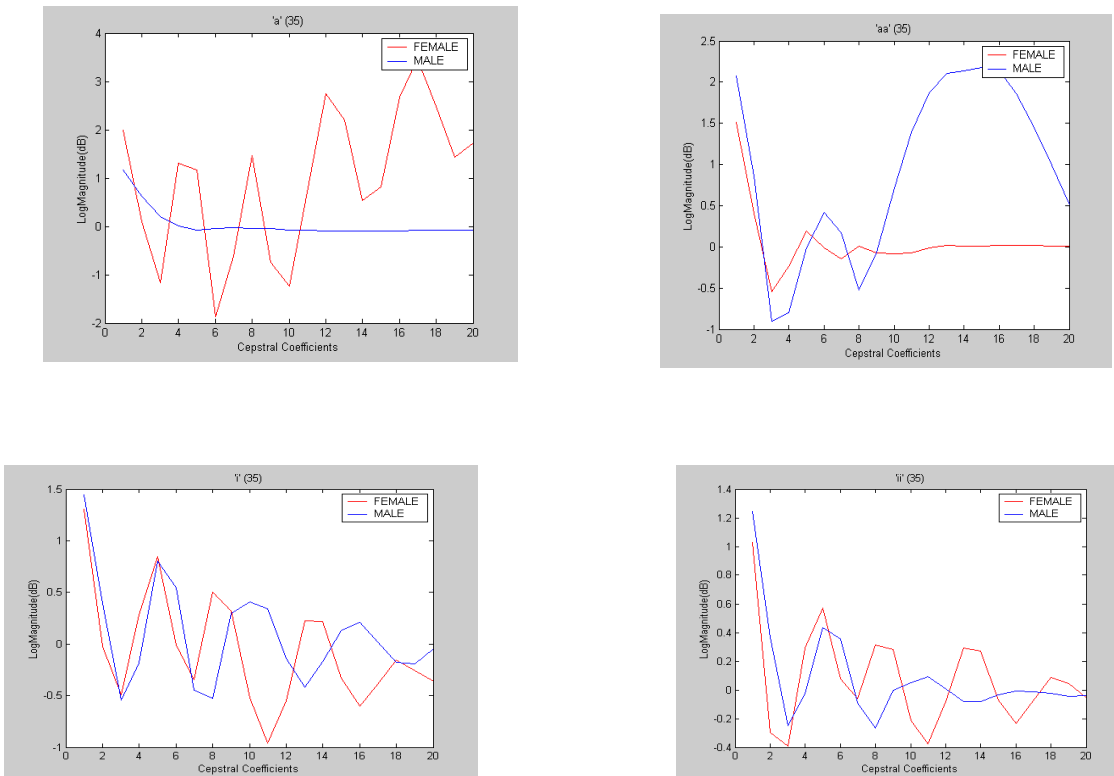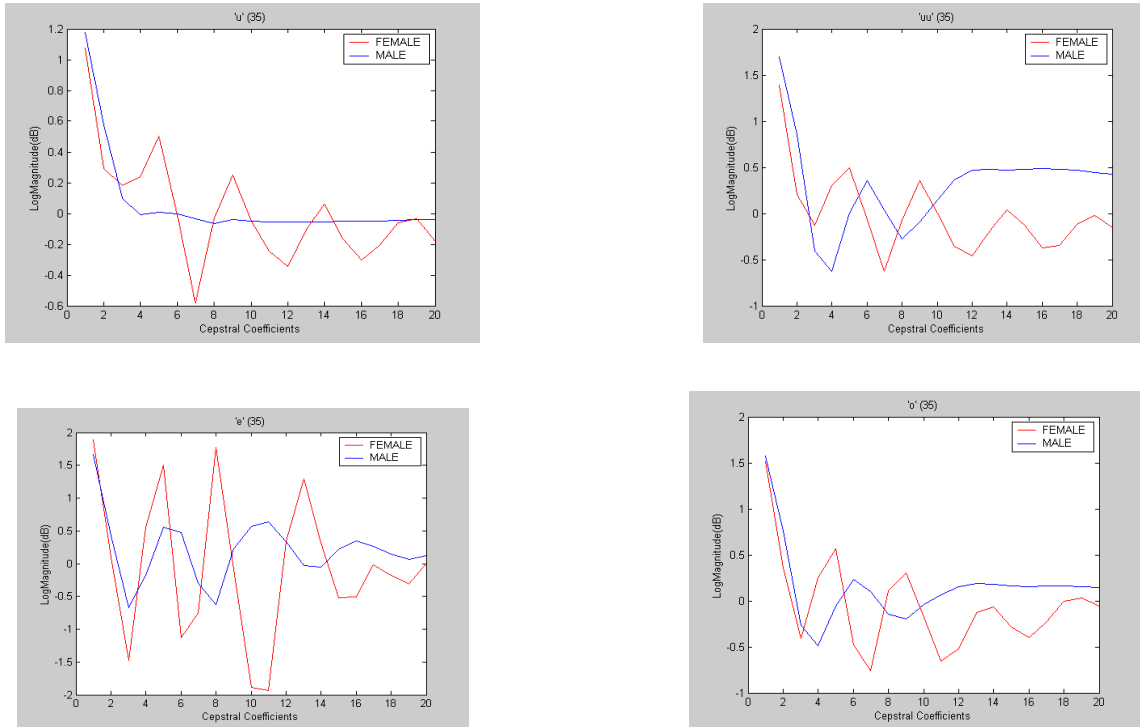
Figure 3: Comparative plots of female and male utterances of the eight sylheti vowels

3. DETERMINING MEL FREQUENCY CEPSTRAL CO-EFFICIENTS

The effitiveness 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. The Cepstral co-efficients allow to extract the similarity between two Cepstral feature vectors. They are considered as important features to separate intraspeaker variability based on age, emotional status of an individual speaker of a language. Campbell(1997) proposed the scope of further improvement

In sound processing, the **mel-frequency cepstrum** (**MFC**) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. The difference between of linear cepstra in feature extraction of speech sounds by the use of m**el- Cepstral Coefficients (MFCC)**

The name **mel** Comes from the word **melody** used for pitch comparisions. The **mel scale** was first proposed by Stevens, Volkman and Newman (1937)[12] . This coefficient has a great success in speech recognition application[4,5,10]. Mel Frequency Cepstral Coefficients analysis has been widely used in signal processing in general and speech processing in particular. It is derived from the Fourier Transform of the audio clip.

the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used

120

in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression.

The mel-cepstrum is a useful and widely used parameter for speech

recognition[6,].There are several methods that have been used to obtain Mel-Frequency Cepstral Coefficients (MFCC).

MFCCs are commonly derived through the algorithm as follows:[7]

Step1 : Divide the signal into frames,

Step2 : For each frame,obtain the amplitude spectrum.

Step3 : Take the logarithms.

Step4 : Convert to Mel spectrum.

Step5 : Take the discrete cosine transfrom (DCT).

Step6 : The MFCCs are the amplitudes of the resulting spectrum.

In the present study, the MFCC's have been calculated from the LPC co-efficients using recursion formula. The LPC co-efficients are first transformed to the cepstral co-efficients and then the cepstral co-efficients are transformed to the Mel Frequency Cepstral Coefficients by using the recursion formula[8] Mel Frequency Cepstral coefficients are co-efficients which collectively make up an Mel Frequency Cepstrum(MFC). They are derived from a type of cepstral representation of the speech sound (a non linear 'spectrum of a spectrum').Mfcc's

are based on known variation of the human ear's critical bandwidth with frequency. The speech signal is expressed in the mel frequency scale for determining the phonetically important characteristics of speech.

As the mel cepstrum coefficients are real numbers, they may be converted to the time domain using the Discrete Cosine Transform(DCT). The MFCC's may be calculated using the following equation [8, 9] corresponding to the female and male utterances.

$$C_n = \sum (log\ S_k\ )[n(k\text{-}1/2)\pi/K]\ \text{-----------}\quad (2.0)$$

Where n= 1,2…………..K

K represents the number of mel cepstrum coefficients,C_0, is excluded from the DCT as it represents the mean value of the input signal which carries less speaker specific information. For each speech frame a set

of mel frequency cepstrum coefficients is computed. This set of coefficients is called an acoustic vector which can be used to represent and recognize the speech characteristics of the speaker.

121

The plot of the MFCC's of the sylheti vowels of the female and male utterances has been shown in fig 1.0 to Fig. 5.0. The maximum and minimum values of the MFCC's of the eight sylheti vowels corresponding to the female and male utterances has been shown in Table 3.0 and Table 4.0 .

Table 3: Range of variation of Mel- cepstral co- officients for sylheti phonemes corresponding to sylheti female utterances

| Age groups → Vowels ↓	14yrs - 21yrs	21yrs - 35yrs	35yrs - 50yrs
a	-8.65 to 6.57	-7.75 to 6.23	-5.77 to 5.52
aa	-8.67 to 6.00	-8.41 to 6.65	-7.50 to 6.48
i	-6.00 to 3.77	-6.33 to 4.34	-3.97 to 3.58
ii	-5.93 to 3.97	-7.18 to 4.68	-4.58 to 2.91
u	-6.68 to 4.72	-7.62 to 7.72	-6.08 to 4.68
uu	-6.29 to 4.19	-7.68 to 8.57	-5.12 to 3.72
e	-6.25 to 2.88	-7.00 to 3.76	-7.26 to 2.00
o	-8.00 to 5.63	-7.27 to 4.98	-6.73 to 6.37

Table 4 : Range of variations of mel frequency cepstral co officients for sylheti phonemes corresponding to sylheti male utterances

| Age groups → Vowels ↓	14yrs - 21yrs	21yrs - 35yrs	35yrs - 50yrs
a	-6.71 to 6.95	-8.93 to 7.12	-6.42 to 8.80
aa	-10.69 to 6.40	-7.98 to 6.28	-8.03 to 8.42
i	-7.54 to 4.37	-7.34 to 7.35	-7.14 to 7.12
ii	-12.58 to 8.77	-7.89 to 6.82	-6.43 to 7.52
u	-6.14 to 8.35	-6.46 to 9.94	-6.57 to 8.92
uu	-6.52 to 8.26	-6.83 to 9.17	-6.53 to 8.89
e	-6.19 to 6.14	-6.35 to 6.56	-6.87 to 6.09
o	-7.89 to 7.21	-7.15 to 7.17	-6.56 to 7.62

Female utterance   Male utterance
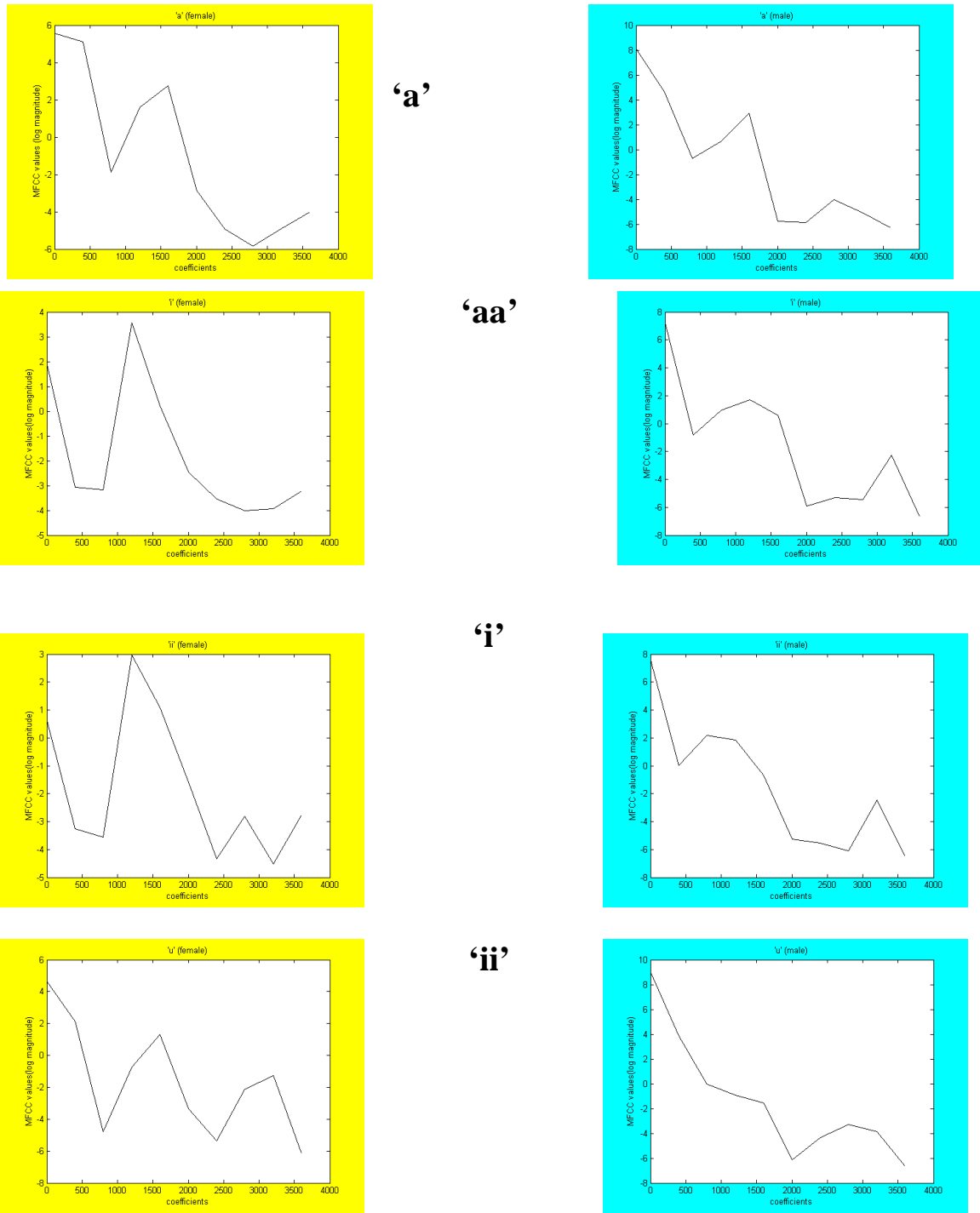
Figure 4: Plots of female and male utterance of 'a', 'aa', 'i' ,'ii'

123
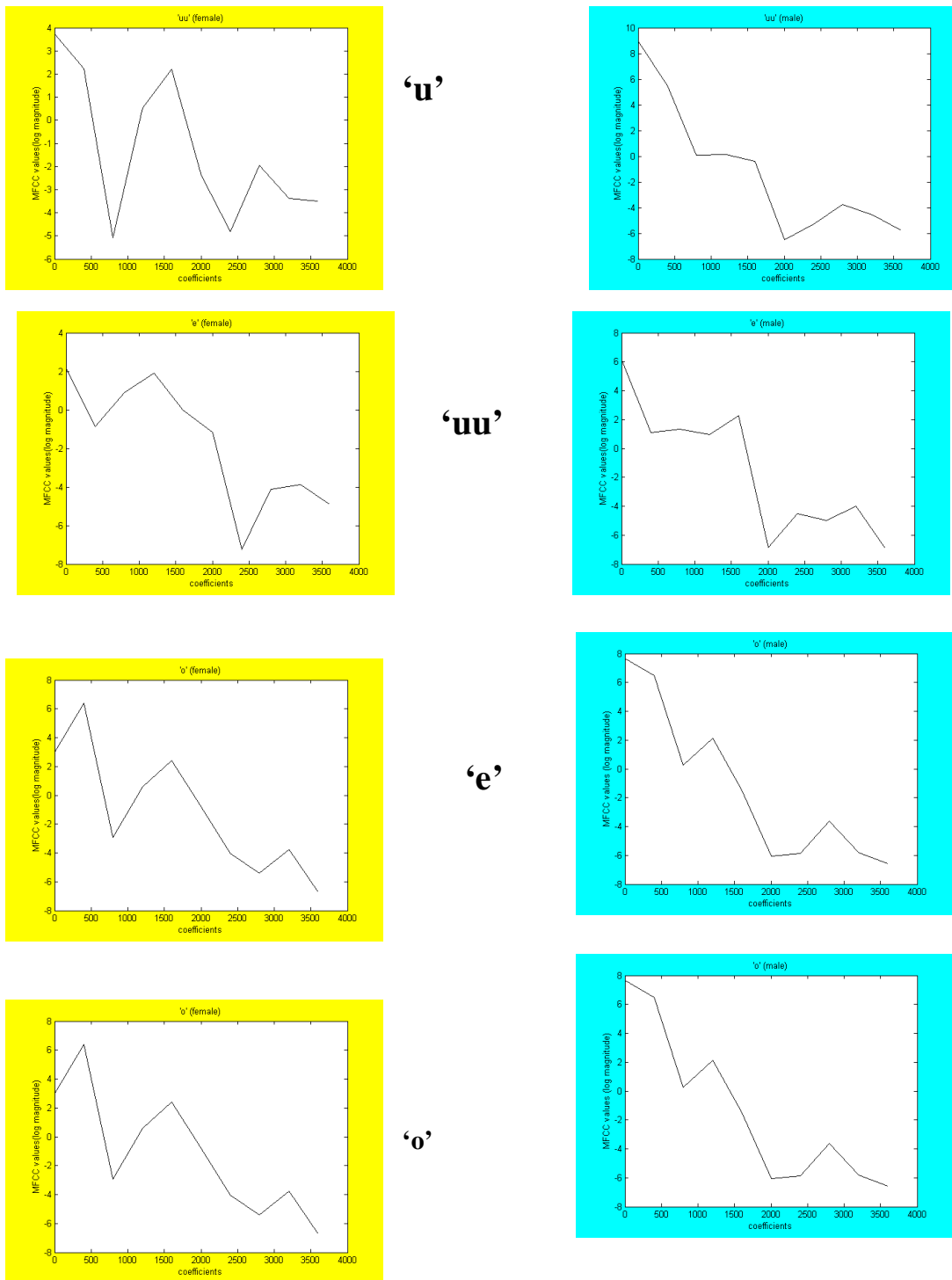
Figure 5: Plots of female and male utterance of 'u', 'uu', 'e' ,'o'

124

RESULTS AND CONCLUSION

Frame no. 16 of the sylheti speakers gives distinct difference between male and female with reference to the utterance of **'a' ,'aa'**and **'u'** .From this observation it can be concluded that the **cepstral coefficients** obtained from the utterance of **vowel 'a' ,'aa'** and **'u'** can be implemented to **recognize the sylheti native speaker with respect to sex**. It is found in the Mel- Cepstral analysis that the Cepstral Co-efficients are relatively higher for male than female.The Linear Cepstral Co-efficients are found less in magnitude than the MFCC. It is observed that in the

verification & identification of male and female utterances through the use of Linear Cepstral & MFCC, the Linear Cepstral measure shows more clearity in distinguishing the male & female utterances. More interestingly, out of the eight Sylheti vowels, the vowels **'a', 'aa'** and **'u'** display more clearly in identifying & distinguishing the gender through the Linear Cepstral Co-efficients analysis as shown in **Fig 1.0 to Fig. 5.0** .Thus for Sylheti language, the three vowels **'a', 'aa',** and **'u'** seems playing a major role in **gender verification & identification**

REFERENCES

1. L.R Rabiner and B.H.Junag, *An Introduction to hidden markov models,* IEEE Acoust, Speech Signal Processing Mag, pp4-6,1986.

2. L.R Rabiner and B.H.Junag,*Fundamental of speech recognition, Dorling Kindersley(India).*

3. F.Soong, E. Rosenberg,B. Juang and L.Rabiner,*A Vector Quantization Approach to Speaker Recognition*, AT & T Technical Journal, Vol.66,March/April 1987,pp.14-26.

4. Jr. J. D . Hansen , J. and Proakis, J., *Discrete Time Processing of Speech Signals,* second ed. IEEE Press, New York, 2000.

5. Pran Hari Talukdar, "Speech Production, Analysis and Coding" 2010.

6.Hampshire School
hhtp://www3.hants.gov.uk/education/ema-advice-lcr-bengali.htm.

7.. Joseph,W. P., "Signal modeling techniques in speech recognition", Proceedings of IEEE, Vol.81.no,9,pp.1215-1247, 1993.

8. Kalita S.K., Gogoi M, Talukdar P.H.,"A Cepstral Measure of the Spectral Characteristics of Assamese & Boro Phonemes for Speaker verification", accepted paper for oral presentation at C3IT-2009

9. Jurafsy, M. and Martin, J. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. New Jersey: Prentice Hall 2006

10.. S spear, P. Warren and A Schafer, "Intonation and sentence processing" proc. Of the Internal Congress of Phonetic Science, Barcelona, 2003.

11. "Sylheti Literature". Sylheti Translation And Research.
http://www.sylheti.org.uk/page2.html.
Retrieved 2007-04-24.

12. Stevens, S.S, Volkman J and Newman E.B : 'Ascale for the measurement of the psychological magnitude pitch " J Acoustical soc. America, vol..8,pp-185-190,1937