LPC Analysis of Vowels and Formant Analysis of Some Typical CV and VC Type of Words in Bodo Language

Monoj Kr. Singha
Deptt. of Physics
Rangia College
Rangia-781354, Assam, India
reiki_sinha@yahoo.com

Jogen Boro
Deptt.of Bodo
Gauhati UniversityGauhati University
Guwahati, Assam, India

Biren Sarma
Deptt. of Instrumentation & USIC
Gauhati University
Guwahati, Assam, India

***Abstract*-Bodo language belongs to the Tibeto Burman family-a sub group of Sino Tibetan family of languages .This study deals with an acoustic representation of vowels and some typical Bodo words through homomorphic analysis of speech signal in the form of LPC (Linear Predictive Coding) model and formant frequency. The speech signals used here are vowels and a set of words in the form of CV (consonant-vowel) and VC (vowel-consonant) type. In our observation we measured the range of variation of the LPCC coefficients and formant frequencies for male and female speakers of Bodo language.**

Key words- LPCC, Formant frequency

I. INTRODUCTION

In the Indian subcontinent, the languages that belongs to the Sino-Tibetan family of languages like Bodo language is said to be syllable timed, that is, syllable occur at regular intervals of time. Syllables are often considered as the phonological “building blocks” of words. They can influence the rhythm of a language, its prosody and its stress pattern etc. In all languages, vowels from the nucleus or peak of the syllables, whereas the consonants before the nucleus are called the onsets and those after it are called coda. A coda-less syllable of the form such as V, CV, CCV etc is called open syllable (or free syllable). The syllable that has a coda (VC, CVC etc) is called closed syllable (or checked syllable).The syllable nucleus is typically a sonorant, usually making a vowel sound, in the form of a monophthong, diphthong and triphthong.

Being a sub group of Sino-Tibetan family of languages, Bodo language is monosyllabic in nature(Basumotary,2005). But keeping the vowel at the nucleus, Bodo language prefer open syllables and close syllables in their word structure. The onset of the syllable and cluster with the nucleus forms the structure like V, VV, CV etc which are found common in Bodo words. Similarly, the rhyme (nucleus and coda) form the structure like VC, CV etc, which are also have a common use in Bodo language.

II. SPEECH MODEL

An ideal model of a speech signal is that which most accurately reflects different aspect of the speech production system. To study the basic components of the speech signal, the individual effect of speech production organs are to be separated. Homomorphic Analysis (Rabiner et al, 1993) is a useful method in this case.

A.M.Noll, (1967) first applied the homomorphic analysis to speech processing which led to the success of cepstral pitch detection. In the following section a homomorphic analysis of speech signal in the form of LPC (Linear Predictive Coding) model, has been illustrated and its application in the present analysis of Bodo speech signals have been highlighted.

The human speech production system can be modeled using a rather simple structure: the lung which generates the air or energy to excite the vocal tract are represented by a white noise source. The acoustic path composed comprising the larynx, vocal cords, pharynx, mouth and nose, can be modeled by a time-varying digital filter.

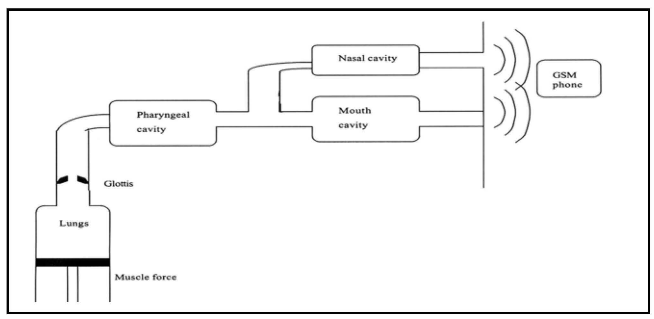

Fig.1: A block diagram of human speech production system.

III. LINEAR PREDICTIVE CODING (LPC)

A. Introduction

Linear prediction is a method for signal source modelling dominant in speech signal processing and having wide application in other areas. Linear Predictive Coding (LPC) is one of the most powerful speech analysis techniques. The glottis (the space between the vocal cords) produces the sound, which is characterized by its intensity (loudness) and frequency (pitch). The vocal tract (the throat, the mouth and the nasal cavity) forms the tube, which is characterized by its resonance frequencies, which are called formants.

The basic problem of the LPC system is to determine the formants from the speech signal. The solution of this

978-1-4577-0748-3/12/$26.00 © 2012 IEEE

problem is a difference equation, which expresses each sample of the signal as a linear combination of previous samples. Such an equation is called a linear predictor i.e. Linear Predictive Coding. The coefficients of the difference equation (the prediction coefficients) characterize the formants. Therefore, the LPC system needs to estimate these coefficients. The estimation is made by minimizing the mean square error between the predicted signal and the actual signal.

The basic idea behind the LPC model is that a given speech sample s(n)at time ***n***, can be approximated as a linear combination of the past ***p*** speech samples (Rabiner & Juang,1993) such that

$$s(n) \approx a_1 s(n-1) + \cdots + a_n s(n-p) \quad \ldots\ldots(1)$$

Where the coefficients are $a_1, a_2, \ldots a_n$ assumed to be constants over the speech analysis frame. The equation (1) can be converted to an equality by including an excitation term Gu(n),

$$s(n) = Gu(n) + \sum_{i=1}^{p} a_i s(n-i) \quad \ldots\ldots.(2)$$

Where $u(n)$ is normalized excitation and G is the gain of excitation.

B. LPC analysis

The relation between s(n) and u(n) is defined as (based on the speech production model Fig-1.

$$s(n) = Gu(n) + \sum_{i=1}^{p} a_k\, s(n) \quad \ldots ..(3)$$

consider the linear combination of past speech samples as the estimate s(n), defined as,

$$\hat{s} = \sum_{k=1}^{p} a_k\, s(n-k) \quad \ldots(4)$$

The predictor error e(n), is defined as

$$\mathrm{e(n) = s(n) - \hat{s}\,(n) = s(n) - \sum_{k=1}^{p} a_k\, s(n-k)} \quad (5)$$

And the error transfer function is,

$$A(Z) = 1 - \sum_{i=1}^{p} a_k\, Z^{-k} \quad \ldots ..(6)$$

The basic problem of linear prediction analysis is to determine the set of predictor coefficient a_k, directly from the speech signal so that the speech properties of the digital filter match those of the speech waveform within the analysis window.

C. The LPC- Cepstral Coefficient

In the present study, LPC-based cepstral coefficients and phonetically important parameters are used as feature vectors. Cepstral weighted feature vector is obtained for each frame by block processing of continuous speech signals. The analog speech waveform is then sampled and quantized analog-to-digital converter. To spectrally flatten the signal, the speech signal has been subjected to the pre-emphasis procedure through a first order digital filter whose transfer function has been given by

$$H(Z) = 1 - az^{-1}$$

$$\text{Where } 0 \le a \le 1.0 \quad \ldots\ldots.\ (7)$$

Consecutive speech signal are taken as a single frame. To reduce the undesired effect of Gibbs phenomenon, the frames are multiplied by a windows function (Hamming window), which is given by (Proakis, & Manolakis, 2004)

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad \ldots\ldots\ldots(8)$$

Where $0 \le n \le N-1$

N is the number of sample in a block. Now, each frame of the windowed signal is next auto correlated to give

$$r_f(m) = \sum_{n=0}^{N-1-m} \hat{x}_f(n)\hat{x}_f(n+m) \quad \ldots\ldots\ldots(9)$$

Where $m = 0,1,2, \ldots p$. The highest auto correlated value, p, is the order of the LPC analysis.

D. LPC parameter conversion to Cepstral Coefficients

The LPC cepstral coefficients, which are a set of values that have been found to be more robust, reliable feature set for speech recognition than the LPC coefficients (Rabiner & Juang,1993). These coefficients are obtained recursively as follows.

$c = \ln[\sigma^2]$ where σ^2 is the gain term in the LPC model.

$$c_m = a_m + \sum_{k=1}^{p} \left(\frac{k}{m}\right) c_{m-k} a_k$$

$$\text{Where } 1 \le m \le p \quad \ldots.(10)$$

$$c_m = \sum_{k=1}^{p} \left(\frac{k}{m}\right) c_{m-k} a_k$$

$$\text{Where } m > p \quad ..(11)$$

Equation (11) shows the computation of cepstral coefficients C_{p+1}, $C_{p+2 \ldots}$ C_p. Generally, q > p is taken for cepstral representation.

IV. Formant Analysis Of Some Typical Bodo Words

Formant frequency is the distinguishing frequency components of human speech. It refers to specific resonance frequencies of vocal tract which have maximum energy concentration during the vowels utterance. It can be qualitatively distinguished by the frequency component of the vowel. Generally, three formants frequencies (F1,F2 and F3) are considered for perception and discrimination of vowels by a listener (Kewley, 1982, 1983). A variety of approaches, such as formant tracking articulator model and auditory model have been used for the analysis and synthesis of speech. The formant tracking method, based on Linear Predictive Code (LPC), has received considerable attention. Based on digitalized technique, the entire frequency range is divided into a fixed number of segment and each segment is represents a formant frequency. A second order resonator for each segment k with a specific boundary is defined. A predictor polynomial defined as the Fourier transform of the corresponding second-order predictor is given by (Welling. I, and Ney, II, 1998)

V. Speech Data Collection For Acoustic

A. Representation

In the present study, speech data is collected from the native speakers of Bodo language who are fluent in speaking and writing the language. Male and female speaker of age between 15 to 30 years, possessing a pleasant and a good voice quality are chosen to record the data. The recording is done one-by one manner. The speakers were instructed to read each word or sentence naturally, without emotions and expression. They were asked to speak clearly and to keep their normal speaking rate and volume. (Talukdar et al 2004, 2006)

The recording is done in audio editing software Cool Edit Pro. Each digitized voice uttered, is divided or blocked into 50 frames of duration 20 millisecond. Every frame contains 441 samples and for each frame 20 cepstral coefficients have been calculated. The spectral characteristics of six Bodo vowels, corresponding to male and female speakers were investigated. Approximately 12 samples were averaged to obtain one coefficient. Firstly, 10^{th} frame of all utterances of male and female speakers have been considered for analysis. The variation of the cepstral coefficients for the Bodo vowels corresponding to the selected speakers have been shown in Table 1 and depicted in Fig.2. However, from continuous frame wise analysis, it is observed that Bodo speaker (Fig. 3) have shown distinct variation of the cepstral coefficients for male and female speakers.

TABLE I: Range Of Variation Of The Cepstral Coefficients Corresponding To The Male And Female Bodo Speaker.

Cepstral Coefficients						
	Male			*Female*		
Vowel	*Maximum*	*Minimum*	*Range of variation*	*Maximum*	*Minimum*	*Range of variation*
/o/	2.2237	-1.5940	3.8177	1.9492	-0.0086	1.9578
/a/	1.6260	-0.9615	2.5875	0.9492	-0.0641	1.0133
/i/	1.1528	-0.1253	1.2781	0.9059	-0.0217	0.9276
/e/	1.2355	-0.6532	1.8887	0.9847	-0.0578	1.0425
/u/	1.0922	-0.0601	1.1523	1.1385	0.0690	1.0695
/w/	1.1832	-0.1541	1.3373	1.1843	-0.1674	1.3517

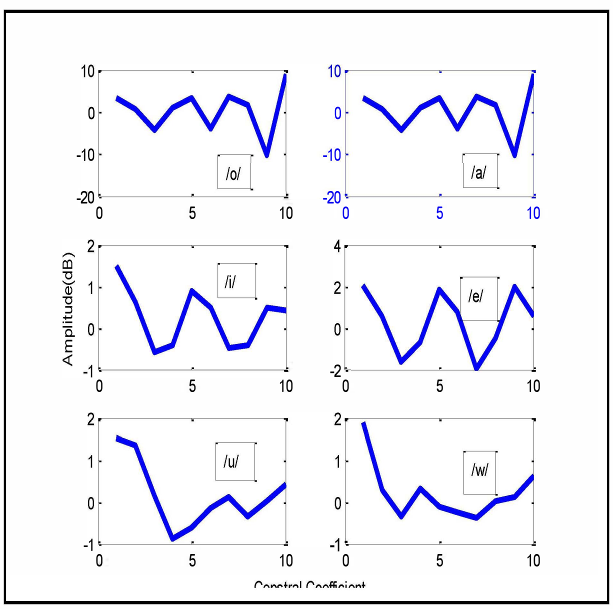

Fig. 2: LPCC of Bodo vowels

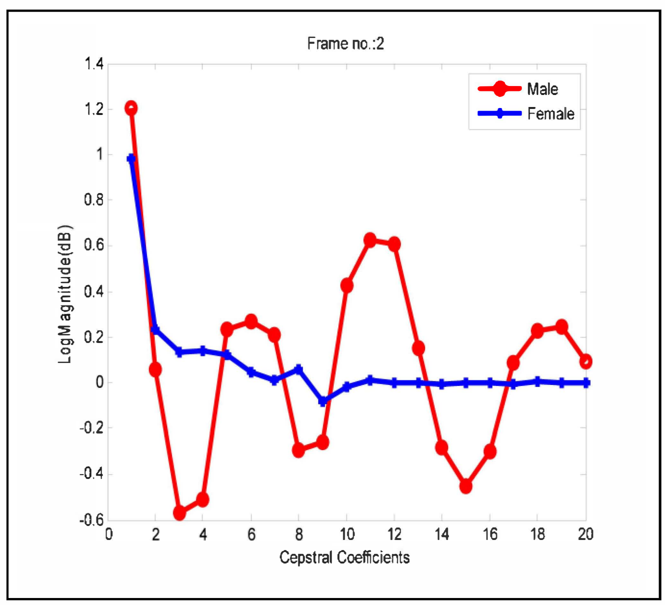

Fig. 3: Distinct variation of LPCC for Male and Female speaker .

VI. Formant Frequencies Estimation

TABLE 2: Represents formant frequencies of Bodo phonemes (vowels) and some typical Bodo words

		Formant frequency					
Vowel		/o/	/a/	/i/	/e/	/u/	/w/
Female	F1	319.1	380.3	411.3	387.5	249.6	292.7
	F2	833.3	1194.5	2409.8	2240.8	997.7	1527.2
	F3	3030.4	3650.4	2911.8	3165.0	3044.3	3165.3
Male	F1	309.3	343.8	394.6	384.9	244.7	206.4
	F2	764.0	1172.0	2341.6	2178.1	837.5	1147.1
	F3	2748.8	2494.5	3002.4	3577.1	3690.6	2486.9
VC		/or/ (fire)	/ aŋ / (I)	/ich/ (pain)	/un/ (back side	/ul/ (confuse)	/em/ (bed)
Female	F1	326.4	326.1	293.3	300.5	347.2	311.4
	F2	1623.4	1717.5	2371.3	1424.7	2353.1	2452.7
	F3	3023,8	3006.2	3455.9	3276.9	2853.5	2765.3
Male	F1	539.1	714.0	299.3	280.2	442.8	398.5
	F2	2293.9	2365.5	2932.2	2240.0	2544.9	1265.7
	F3	3242.6	3199.6	3189.1	2636.7	3350.7	2435.8
CV		/hw/ (to give)	/bu/ (to swell)	/ru/ (to boil)	/ bu / (to beat)	/be/ (this)	/gi/ (to fear)
Female	F1	320.7	382.1	311.1	337.6	354.6	334.8
	F2	1687.9	1661.1	1623.5	1853.7	1699.5	1617.9
	F3	3120.24	3077.1	3445.5	2996.8	3001.65	2947.7
Male	F1	494.4	690.1	633.0	375.5	283.4	393.0
	F2	2109.8	2545.8	2386.2	2536.2	2250.1	2223.5
	F3	3216.3	3355.9	3298.9	2842.9	3220.0	3287.7

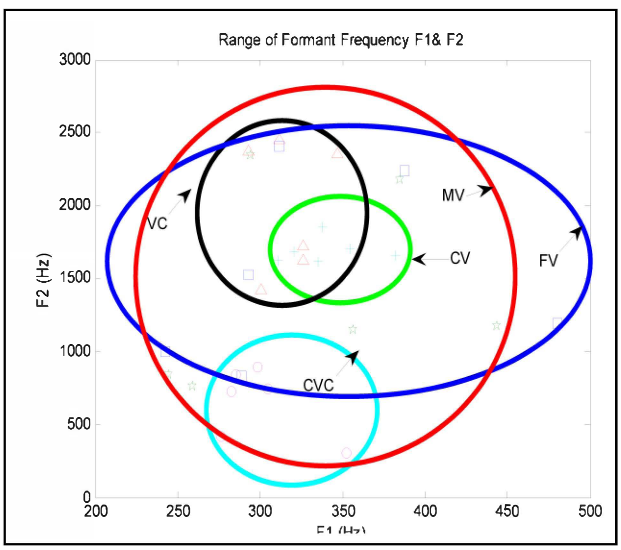

Fig. 4 Shows the F1-F2 plot for range of formant frequencies for different words.

Result And Discussion

Depending on the analysis made in this study on cepstral features and formant frequencies of Bodo phonemes and words the following observations were made-

a) Significant variation of cepstral coefficients are observed among the Bodo vowels as shown in Table 1. The cepstral variation is found to be maximum with respect to vowel /o/ and minimum corresponding to vowel /u/, in case of male speakers. Similarly, for female Bodo speakers, the maximum variation of cepstral measure is found corresponding to vowels /o/ and minimum in case of /i/.

b) Significantly, cepstral coefficients of Bodo vowels for different frames have shown distinctive characteristic (Fig.3) for male and female speaker. The variation of the cepstral coefficients for male is very irregular in contrast to the stable variation of female cepstral coefficients. This observation may be helpful in sex determination for Bodo speakers.

c) The range of variation of cepstral coefficients for Bodo male is found within the range of $3.8177 > C_{Bodo} > 1.1523$. The range of variation for female is found $1.9578 > C_{Bodo} > 0.9276$ i.e. the variation of cepstral features for Bodo vowels for female is less (Male-2.6654; Female-1.0302)with respect to male i.e the former is stable as compared to the latter.

d) It is found that the vowel location for /u/ (low F1, low F2), /i/ (low F1, high F2) and /a/(high F1, low F2) with respect to other vowels are the extreme points of location.

e) The Fig.4 have shown that the formant frequencies of the selected word sets for Bodo language lies within the range of the formant frequencies of the isolated vowels. The investigation have shown that (Table-2) the range of formant frequency is maximum in case of isolated vowels, but when the vowels are placed in the nucleus of a structure like CV, VC the formant frequency decreases.

REFERENCES

[1] A.M. Noll, "Spectrum Pitch Determination", J. Acousti Society. A.M. Vol.41. pp.293-309, Feb.1967.

[2] Basumatary, P.;"An Introduction to the Boro Language", Mittal Publication, 2005.

[3] Benedict P.K., Sino-Tibetan A Conspectus, p.93.

[4] Devi, Basanti. "Vowel Length in Assamese", Acoustic Studies in Languages edited by B.B. Rajapurohit, Central Institute India Mysore1984.

[5] Fant, G. (1960). Acoustic Theory of Speech Production. Mouton. The Hague.

[6] Goswami, S.N.; "Studies in Sino-Tibetan Languages", 1988.

[7] Proakis, J.G. and Manolakis, D.G., Digital Signal Processing Principles, Algorithm and Applications, pearson edition, Third Indian reprint 2004.

[8] Rabiner, L.R. and Gold. B. ; "Theory and Application of Digital Signal Processing", Englewood Cliffs, N.J. Prentice Hill, 1975.

[9] Rabiner, L.R. and Schafer, R. W.; "Digital Processing of Speech Signals", Englewood Clifts, N. J. Prentice Hill, 1978.

[10] Talukdar P.H., Bhattacharjee, U., Goswami, C.K. and Barman, J., "Measure Boro Vowels through LPC Analysis", J. of Computer Society of India, Vol. 34, No. 1, Jan.2004.

[11] Talukdar, P. H, Kalita, S. K, & Dutta, R. "A Spectral analysis of Bodo and Asamese Vowels ". Abstract s 3rd international conference on Computers and Devices for Communication. CODEC– 06, Kolkata, India, pp. 41, 2006.

[12] Willing I., and Ney, II, Formant Estimation for Speech Recognition, IEEE Transactions on Speech and Audio Processing, Vol.6. No.PP-36-48,1998