Analytical Study of CV Type Bodo Words using Formant Frequency Measure

Uzzal Sharma

Dept. of CST & IT,
DBCET, Assam Don Bosco University,
Guwahati, India,
Email : uzzal.sharma[@]dbuniversity.ac.in

Abstract — Words can be categorized into different types according to the position of occurrences of vowels and consonants in the word. Accordingly we have CV (Consonant-Vowel), VC (Vowel-Consonant), CVC (Consonant-Vowel-Consonant), CVCC (Consonant-Vowel-Consonant-Consonant), CVVC (Consonant-Vowel-Vowel-Consonant etc type of words in most of the languages. As a first step towards the recognition of any speech signal, it is very much important to study the different types of words using some of the available techniques. Some of the approach which produces reliable and good results are Formant Frequency measure, Mel-frequency cepstral coefficients (MFCC) etc. In this paper, a step has been taken to measure the formant frequency of CV type Bodo words to identify the distinct features of it. Formant Frequency, based on Formant Tracking Model can be defined as the spectral peak of the sound spectrum |P(f)|.

Keywords— MFCC, Formant Tracking Model, FFT, Formant Frequency, Resonance Frequency

I. INTRODUCTION

In the determination of phonetic content of speech, formant frequencies are found to be very important which is revealed from past researches [1], [2], [3]. Speech sounds, called "phonemes" are classified either as vowels or consonants. In the same way words are comprises of combination of vowels and consonants in a particular sequence. Accordingly we have CV (Consonant-Vowel), VC (Vowel-Consonant), CVC (Consonant-Vowel-Consonant), CVCC (Consonant-Vowel-Consonant-Consonant), CVVC (Consonant-Vowel-Vowel-Consonant) etc types of word. The transfer function of energy from the excitation source to the output through a tube can be referred as natural frequencies of resonances of the tube. The formant frequency depends upon the dimensions and shape of the vocal tract, which is considered as a tube or combination of some tubes of varying cross-sectional area where each shape is characterized by a set of formant frequencies. The variation in the shape of the vocal tract produces different sounds. Thus as the shape of the vocal tract changes, the spectral characteristics of the speech signal vary with time. Typically, a human vocal tract exhibits about three significant

resonances below 3500 Hz. The formant frequency representation is a highly efficient and compact representation of speech sound [7].

The placement of vowels and consonants in a words of any language are very important towards the identification of the different types of words. Analysis of formant for different type of words helps in distinction of speakers when the same words are uttered by different speaker as formant values contains large amount of speaker information. In this paper, an attempt is made to analyze the words of CV types of Bodo language, a major language of NE India, by applying Formant frequency measure and some remarkable results has been observed.

II. SPEECH RECOGNITION AND FORMANT ANALYSIS

Formants can be defined as the spectral peak of the sound spectrum |P(f)| [4]. These are the peaks which are known as the Resonance Frequency, |T(f)| that are observed in the spectrum envelop [5]. Although in most of the cases, it is seen that the Resonance Frequency, |T(f)| and Formant Frequency, |P(f)| is same, but in some particular cases it may be different.

To develop a perfect speaker identification system, a number of approaches have been developed for the analysis and synthesis of speech signal. Among all, Formant Tracking Method [9, 10], Articulatory model [12], and Auditory model are considered as the basic models for speech recognition and research. Out of these, Formant Estimation Model based on the Linear Predictive Coding (LPC) has been found to be very successful [11, 13]. The formant model used in the present study for the determination of Formant Frequency of Bodo CV type words are based on the model proposed by Welling et. al. [14]. Applying the proposed technique, the entire frequency band is segmented into a fixed number of segments, where each of these segments represents frequency. A second order resonator for each segments K, with a specific boundary is defined. A predictor polynomial defined as a Fourier Transform of the Corresponding second order predictor is given by [15],

$$A_k(e^{jw}) = 1 - \propto_k e^{j\omega} - \beta_k e^{-j2\omega} \quad (1)$$

Where, $\propto_k$ and β_k are the real valued prediction co-efficient. The formant frequency is given by,

$$P_f = across[-\alpha_k(1-\beta_k)/4\beta_k] \quad (2)$$

The value of α_k and β_k are defined as,

$$\alpha_k < 2 \quad (3)$$

and,

$$-1 < \beta_k < [-|\alpha_k|/(4-|\alpha_k|)] \quad (4)$$

Now, using equation (2), the Formant Frequencies of CV type Bodo words are estimated for both Male and Female informants. A good care was taken to ensure the correct pronunciation, which was verified by some Bodo Phonetic experts. For obtaining the Formant frequency, the spectrum is subjected to First Fourier Transformation (FFT).

There are few other techniques available for the recognition of speech e.g. MFCC, Pitch, etc. But in case of CV type words of this language, formant frequency gives a higher accuracy rate which is revealed from the study.

III. METHODOLOGY

The dataset is prepared very carefully and for the purpose of accurate study the same set of words are given to the speaker as well as to the phonetic expert. The pronunciation was verified by the phonetic expert. The speakers were advised to speak in stress-free manner while maintaining a constant pitch as far as possible. While preparing the dataset, the words of required type under study are embedded in the natural running sentence, than the required words are separated from the sentence and stored as a corpora entry. The speakers were given a rest of 10 to 15 minutes after every session of recording.

The Male and Female informants of age between 15 to 30 years possessing a pleasant and good voice quality are chosen to record the data. Only native speakers being graduate or post-graduate are selected. To ensure accuracy and consistency, the recording process is supervised by acoustic phonetic experts of Bodo language. The block diagram to find formant frequency is shown in Figure 1.

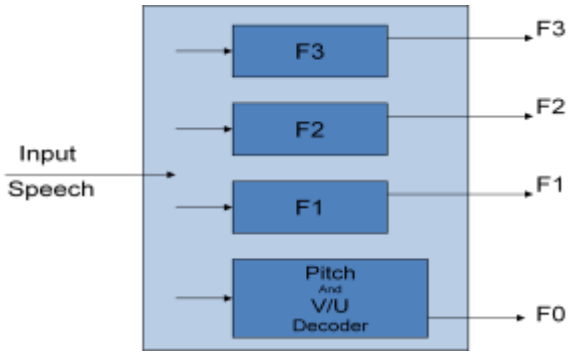


Fig. 1 Formant Frequencies estimation

IV. RESULTS

For the current study, the recorded data set is analyzed for First(F1), Second(F2) and Third(F3) formant frequencies as under :

The recording and separation of the important segment is performed using audio editing software Cool Edit Pro and after that the analysis was done using MATLAB 7.1, and COLEA (subset of a COchLEA Toolbox), a special speech signal analysis tool belongs to MATLAB [6]. Each digitized voice recorded, is divided into 50 frames of duration 20 millisecond (ms) each. Every frame contains approximately 441 samples and for each frame Formant Frequencies (F1, F2, F3) are calculated and investigated. The variation of the formant frequencies for the CV type words "Bu" and "K^{h}a" for both high and low tone corresponding to the selected speakers have been shown in Table-I for male and Table-II for female and depicted in Fig. 2(a) and Fig. 2(b) for the word "Bu" for high and low tone respectively and Fig. 3(a) and Fig. 3(b) for the word "K^{h}a" for high and low tone respectively when uttered by male informant and the female version is shown in the fig. 4(a) and Fig. 4(b) for the word "Bu" for high and low tone respectively and Fig. 5(a) and Fig. 5(b) for the word "K^{h}a" for high and low tone respectively. The "H" and "L" represents High and Low tone respectively

V. DISCUSSION AND CONCLUDING REMARKS

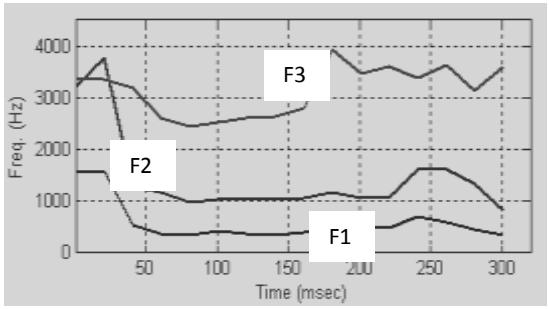
It has been observed that, in case of CV type of words uttered by female informants, the change in frequency is gradual in most of the cases with one or two exceptions and this characteristic is observed mostly in case in F1. On the

TABLE I
RANGE OF VARIATION OF FORMANT FREQUENCIES OF FEW BODO CV TYPE WORDS (MALE)

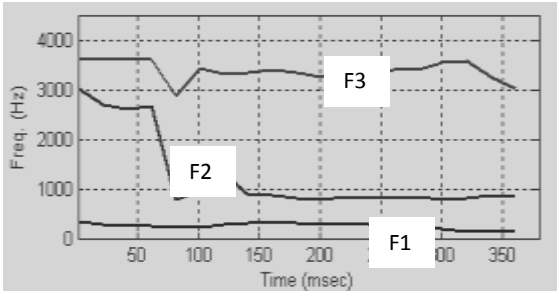
Word	Value	F1(KHz)	F2(KHz)	F3(KHz)
Bu-H	Max	1.56	3.76	3.93
	Min	0.32	0.79	2.43
	Average	0.57	1.43	3.17
	Range	1.24	2.97	1.50
Bu-L	Max	0.32	1.32	3.57
	Min	0.12	0.29	2.59
	Average	0.23	0.79	3.20
	Range	0.20	1.03	0.98
K^{h}a-H	Max	1.61	3.78	3.99
	Min	0.82	1.25	2.30
	Average	1.17	2.30	3.54
	Range	0.79	2.53	1.69
k^{h}a-L	Max	1.55	3.67	3.90
	Min	0.12	1.51	2.83
	Average	1.03	2.10	3.46
	Range	1.44	2.16	1.07

TABLE II
RANGE OF VARIATION OF FORMANT FREQUENCIES OF FEW BODO CV TYPE WORD (FEMALE)

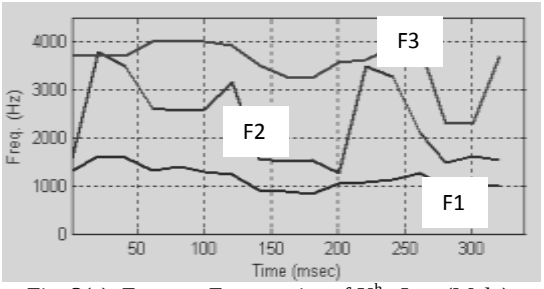
Word	Value	F1(KHz)	F2(KHz)	F3(KHz)
Bu-H	Max	0.46	0.758341	3
	Min	0.25	0.60	3.00
	Average	0.35	0.69	3.00
	Range	0.21	0.16	0.00
Bu-L	Max	0.32	2.93	3.99
	Min	0.17	0.42	3.91
	Average	0.27	0.89	3.96
	Range	0.15	2.51	0.08
K^hA-H	Max	1.73	3.86	3.95
	Min	0.98	1.49	3.76
	Average	1.48	2.61	3.88
	Range	0.75	2.38	0.19
K^hA-L	Max	1.42	1.66	3.92
	Min	0.35	1.47	2.26
	Average	0.95	1.58	3.19
	Range	1.08	0.19	1.66



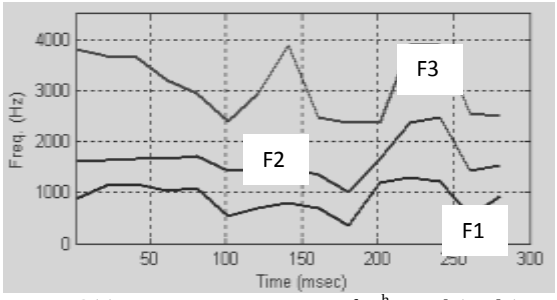
Fig, 2(a) Formant Frequencies of Bu-High(Male)



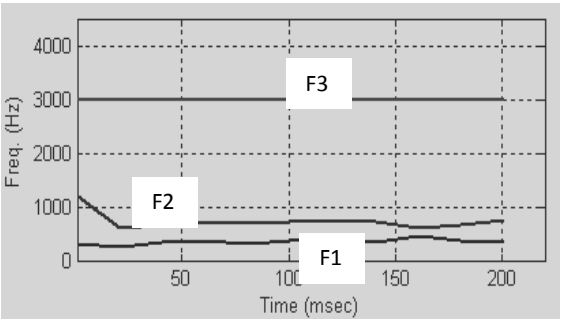
Fig, 2(a) Formant Frequencies of Bu-Low(Male)



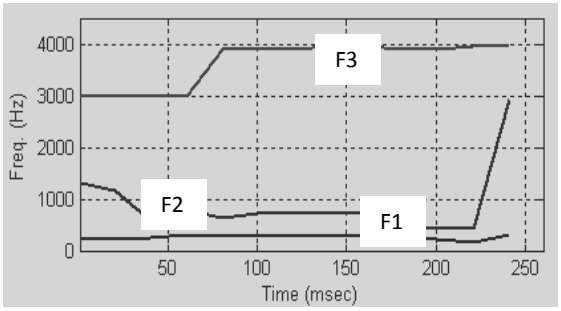
Fig, 3(a) Formant Frequencies of K^ha-Low(Male)



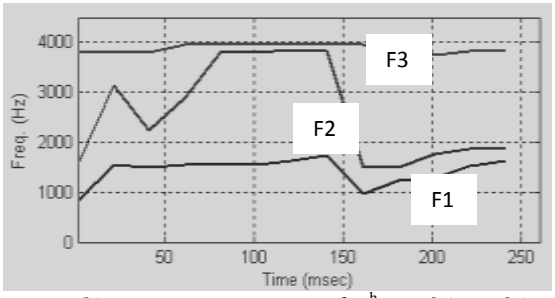
Fig, 3(a) Formant Frequencies of K^ha-High(Male)



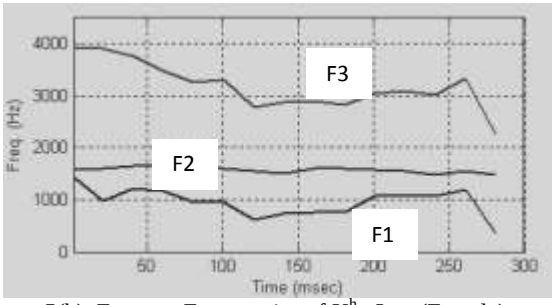
Fig, 4(a) Formant Frequencies of Bu-High(Female)



Fig, 4(b) Formant Frequencies of Bu-Low(Female)



Fig, 5(b) Formant Frequencies of K^ha-High(Female)



5(b) Formant Frequencies of K^ha-Low(Female)

other hand, in case of male informants, the change in the frequency is more gradual with one or two extinctions, and uniform frequency change is observed in all F1, F2, and F3.

It is also observed that, when the same word has a High tone, the frequency tends to raise irrespective of Male and Female informants, where as in case of low tone, the frequency tends to go down, with one or two exceptions.

REFERENCES

[1] M.J. Hunt, "Delayed Decisions in Speech Recognition - The Case of Formants", Pattern Recognition Letters, Vol. 6, pp. 121-137, July 1987.
[2] P. Schmid and E. Barnard, "Robust, N-BestFormant Tracking", Proc. EUROSPEECH'95, pp. 737-740, Madrid, 1995.
[3] L. Welling and H. Ney, "A Model for EfficientFormant Estimation", Proc. IEEE ICASSP, pp. 797-800,Atlanta, 1996.
[4] Fant, G. (1960). Acoustic Theory of Speech Production. Mouton & Co, The Hague, Netherlands.
[5] Benade, A. H. (1976) Fundamentals of musical acoustics, Oxford University Press, London.
[6] Loizou, P. "COLEA: a MATLAB software tool for speech analysis".
[7] Rabinar L.R., Juang B.H. (1986), 'Fundamental of Speech Recognition', Dorling Kinderseley (India).W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
[8] Fatima, N., Zheng, T.F. (2012), 'Vowel-category based Short Utterance Speaker Recognition', Proc. 2012 International.
[9] D. Talkin, (1987), 'Speech Formant Frequency estimation using dynamic programming with modulated transition cost', AT&T Bell labs, McGraw Hill, NJ,
[10] O. Schmidbaner (1990), 'An algorithm for automatic formant extraction in continuous speech', Proc. EUSIPCO-90, Fifth European Signal Processing Conference, Sept 1990, pp-115.
[11] Atal, B. S. and Hanauer, S. L. (1971), 'Speech Analysis and Synthesis by Linear Prediction of the Speech Wave', J. Acoust. Soc. Am., 50, pp. 637-655.
[12] H.B. Richard, Mason J.S. Hunt M. J. and Bridle J.S. (1995), 'Deriving Articulatory Representation of Speech', Proc. of European Conference of Speech Communication and Technology, Madrid, Spain, Sept. 1995, pp- 761,.
[13] Snell R.C. and Milinazzo F. (1993), 'Formant Location form LPC Analysis Data', IEEE trans. Speech Audion, Processing, April 1993, pp-129.
[14] Welling L. and Ney H. (1998), 'Formant Estimation of Speech Recognition', IEEE trans. Speech and Audio processing Sept 1985, pp-134.
[15] Rabinar L.R., and Schafer R.W. (1978), 'Digital Processing of Speech Signal', Prentice Hall, Englewood Cliff, NJ.

Author Profile

Dr. Uzzal Sharma has obtained his MCA from IGNOU and completed PhD from Gauhati University. Dr. Sharma has over 13 years of experience. His research area includes Speech Signal Processing and Software Engineering. He has published more than 25 research papers in journals (International and National) and conference proceedings (International and National). He also has 9 book chapters to his credit in edited book. Currently he is an Assistant Professor at Assam Don Bosco University, Guwahati.