IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Special Issue-1 E-ISSN: 2347-2693

A Comprehensive Analysis on Vowels Present in Bodo Language

Uzzal Sharma*
Assistant Professor: Dept. of CSE & IT
School of Technology, ADBU
Guwahati, India
uzzal_sh@rediffmail.com

***Abstract*— Research on Speech signal processing has gained a high degree of importance due to the broad areas of application. The success of any speech recognition system and speaker identification system depends on the level of accuracy and correctness. To accomplish a high level of accuracy, it is very important to use the proper tool for the purpose. The present paper tries to use two popular tools namely MFCC and Formant Frequency Measure used for the purpose of feature extraction and to give a comparison between the two. The Bodo language is considered for the present research.**

Keywords- Formant frequency measure; MFCC, OSV, ASR, DCT.

I. INTRODUCTION

Bodo is one of the tonal languages widely spoken in India, especially in the North-East region of India and more specifically in Assam and in few parts of Nepal. There are only a few tonal languages spoken in India and Bodo is one of them. The tonal languages are the kind of languages where the tone associated with the words plays an important role in distinguishing the uttered words in terms of its meaning. A language consists of sentence and sentence consists of words which comprises of vowels and consonants sounds [1]. Although the presence of both vowels and consonants are indispensable, but vowels plays a very important role as far as the recognition of words and identification of speaker is concerned [2, 3 and 4].

II. BACKGROUND

In the present paper, the available Bodo vowels are studied in terms of Mel-Frequency Cepstral Coefficient (MFCC) and different formants namely first formants (f1) and second formants (f2) to characterize them from two different angles and make a comparative study of the both.

Speech and speaker recognition has gained popularity day by day because of its diversified application in the field of Natural Language Processing (NLP), Speech Recognition System and Speaker Identification System etc. As vowels and consonants together can be considered as the building blocks of any language, their study is very important. But vowels as compared to consonants plays very important role [2, 3 and 4]. So as a first phase, vowels present in a language need to be analyzed. In the current paper, vowels of Bodo language are studied using MFCC and formant frequency. A comparative study is performed to identify the better tool for speech recognition and speaker identification is concerned.

Before we proceed for the actual study, let us see how the Bodo language reaches the today's stage. The Bodo is the second major language of Assam [5]. It is derived from -TIBETAN or TIBETO-CHINESE family [6] and is one of the tonal languages spoken in India especially in the north east region and few parts of Nepal. It consists of six (6) pure vowels and two semi-vowels; it also has sixteen (16) consonant sounds. As a tonal language, Bodo has two distinct tones. Namely High Tone and Low Tone [7].The basic word order of the Bodo language is SOV (Subject-Object-Verb). But, as it has rich case system all other kinds of word order i.e. SVO (Subject-Verb-Object), OSV(Object-Subject-Verb), OVS(Object-Verb-Subject), VOS(Verb-Object-Subject) and VSO(Verb-Subject-Object)

can realize the same meaning which an SOV structure can. Out of them OSV is more frequently found in the native speakers' speech.

III. ANALYSIS

There are a number of tools available for the purpose of feature extraction in Speech Recognition System. Each of the different tools has its strengths and weaknesses. The analysis of feature phase plays an essential role in any Automatic Speech Recognition System (ASR). There are a number of tools available for the feature extraction purpose. In this paper, two of the important methods namely MFCC and Formant Frequency are used for the purpose of feature extraction and comparative study among the two in performed.

A. Feature Extraction Using MFCC

MFCC is considered as one of the standard method for feature extraction as a part of speech recognition process [2]. MFCC is perhaps the most popular feature for speech recognition which can be obtained by the following steps:

- The signal under study is first subjected to Fourier analysis and then
- Convert the power spectrum to a Mel frequency spectrum.
- Take the logarithm of the above spectrum and by computing the inverse Fourier transform, the MFCC can be obtained.

The reason behind the popularity of MFCC over other tool is the accuracy rate and a minimal amount of computational complexity [8].

The MFCCs is calculated for signal in time domain S using the equation (1) [6]:

$$C_n = \sum_{n=1}^{K} (\log S_K) \left[n \left(K - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad \text{..... (1)}$$

$$\text{where } n = 1,2, \dots K$$

Here K represents the number of Mel cepstrum coefficient, which has been taken as 12 in the present study. C_0 is excluded from the DCT as it represents the mean value of the input signal which carries little speaker specific information. For each speech frame of about 20ms with overlap, a set of Mel-frequency cepstrum coefficients is computed. This set of coefficients is called an acoustic vector used to represent and recognize the speech characteristic of the speaker.

The vowels of Bodo language uttered by both male and female speakers are subjected to MFCC analysis and following results are obtained:

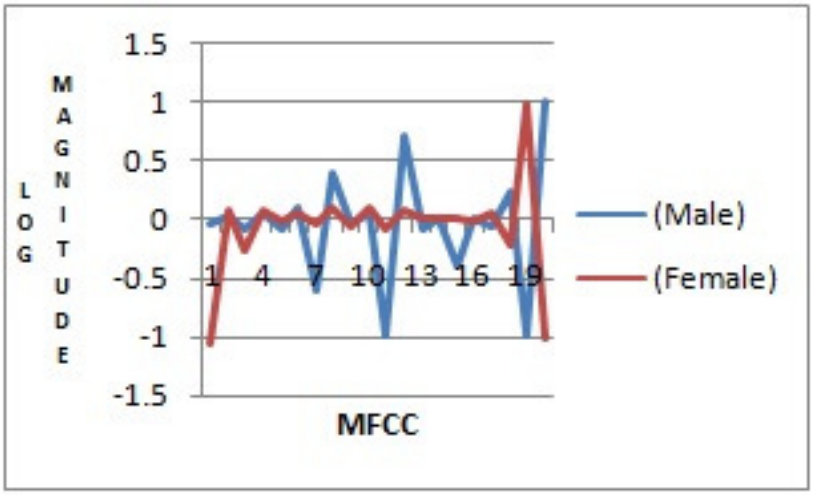

Figure 1: MFCC comparison of Male and Female in vowel /a/

- It has been observed that the values are changing from negative to positive scale more drastically in case male speaker as compared to female speaker.
- The changes in values are steadier in case of female informants ranging from -0.25 to +0.1. Where as in case of male informants the values are ranging from -1.0 to +0.75.
- It has also been observed that speaker recognition is much more accurate in case of male speaker as compared to female speaker.

B. Feature Extraction Using Formant Frequency

*F*ormant frequency measure can be considered as one of the important tool to extract speech features. In the current paper, the formant frequency measure is used to extract the characteristic feature of Bodo vowels. The main distinction between vowels and consonants is that vowels resonate in the throat where as consonants does not. Formants are exactly the resonant frequencies of the vocal tract when pronouncing a vowel [9]. The variation in the shape of the vocal tract forms the different vowel 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 resonance of significance below 3500 Hz. The formant frequency representation is a highly efficient and compact representation of the time varying characteristics of speech [10].

The formant frequency is given by,

$$P_f = \text{across}\left[-\alpha_k(1-\beta_k)/4\beta_k\right] \quad \text{.... (2)}$$

So, the value of α_k and β_k which are prediction coefficients that are defined as,

$$\alpha_k < 2, \quad \text{and} \quad \text{... (3)}$$

$$-1 < \beta_k < [-|\alpha_k|/(4-|\alpha_k|)] \quad \text{....(4)}$$

The variation of formant frequency for the vowel /a/ is shown in the Table 1 when uttered by female informant and in the Table 2 when uttered by male informant.

Table 1 : Range of variation of formant frequencies of Bodo vowel **/a/** (Female)

Vowel		F1 (KHz)	F2 (KHz)	F3 (KHz)
/a/	Max	1.50	3.88	3.98
	Min	0.46	1.12	2.13
	Average	0.96	1.74	3.49
	Range	1.03	2.75	1.84

Table II : Range of variation of formant frequencies of Bodo vowel **/a/** (Male)

Vowel		F1 (KHz)	F2 (KHz)	F3 (KHz)
/a/	Max	1.70	3.82	3.99
	Min	0.17	1.41	2.86
	Average	1.03	1.76	3.50
	Range	1.52	2.41	1.13

The current study reveals that F3 does not plays very important role as the distinction of vowel. Only F1 and F2 are considered for the present study. It is found that both F1 and F2 could give significant value as shown in Table 1 and Table 2 for female and male informants and depicted in Figure 2 which can be used for the recognition of all the different vowels. The result obtained has an accuracy rate of 77%.

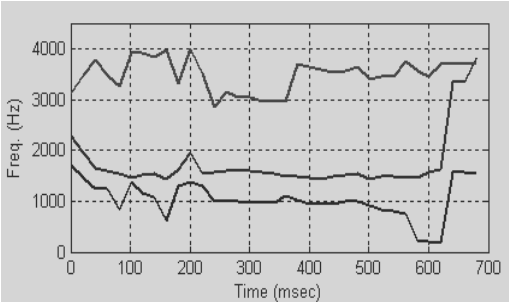

Figure 2: Formant plot for female informants when uttering the vowel /a/

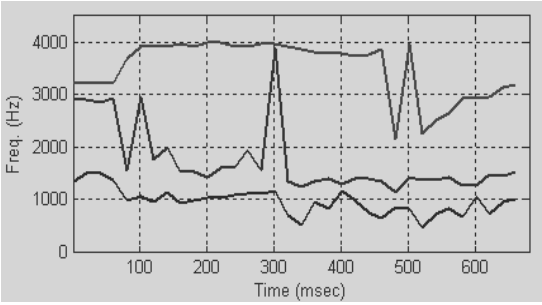

Figure 3: Formant plot for male informants when uttering the vowel /a/

IV. CONCLUSION

Although both MFCC and Formant frequency measure are some of the important and most widely used tools, both of them has some advantages and disadvantages. The current study conducted on Bodo language reveals that MFCC gives better result as compared to Formant frequency measure.

In future, an advanced form of Formant Frequency measure has been planned so that the accuracy rate will be enhanced.

REFERENCES

[1] G. N. Kodandaramaiah, M. N. Giriprasad, and M. M. Rao, "Independent speaker recognition for native english vowels," International Journal of Electronic Engineering Research, vol. 2, 2010.

[2] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations of Monosyllabic Word recognition In Continuously Spoken Sentences," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, pp. 357- 366, 1980.

[3] M. Ursin, "Triphone Clustering In Finnish Continuous Speech recognition," Master Thesis, Department of Computer Science, Helsinki University of Technology, Finland, 2002.

[4] Cole, R.A. Yonghong Yan ; Mak, B. ; Fanty, M. 'Acoustics, Speech, and Signal Processing', 1996 IEEE International Conference Volume:2 pp. 853 - 856.

[5] Biljana Prica and Sinisa Ilic; FACTA UNIVERSITATIS (NI˙S) SER.: ELEC. ENERG. vol. 23, no. 3, December 2010, 379-393.

[6] Welling L. and Ney H. (1998), 'Formant Estimation of Speech Recognition', IEEE trans. Speech and Audio processing Sept 1985, pp-134.

[7] Sharma Uzzal, "Formant Frequency Measure and Analysis of Bodo Vowels – A Core Language of NE India', International Conference on INFORMATION & MATHEMATICAL SCIENCES (IMS-13), 24-26 October, 2013, ISBN : 978-93-5107-162-4, pp. 32-34. (Published by Elsevier)

[8] E. Darren. Ellis "Design of a Speaker Recognition Code using MATLAB "Department of Computer and Electrical Engineering- University of Tennessee, Knoxville Tennessee 37996. 9th May 2001.

[9] Rabinar L.R., Juang B.H. (1986), 'Fundamental of Speech Recognition', Dorling Kinderseley (India).

[10] V. Anantha Natarajan; S. Jothilakshmi International Journal of Computer Applications (0975 – 8887), Volume 56, No.15, October 2012, pp.345-351.