Analysis of Formant Frequency F1, F2 and F3 in Assamese Vowel Phonemes using LPC Model

Dr. Bhargab Medhi
Department of Applied Science
Gauhati University
Guwahati, Assam, India

Abstract—**Formant frequency plays an important role in speech as well as speaker recognition. Formants are the spectral peaks of a sound wave which means the specific resonance frequencies of vocal tract which have maximum energy concentration during the vowel utterances. In a speech spectrum, there may be any number of formants, but for speech the most informative are the first three formants referred to as F1, F2, and F3. In this paper, the paths of these three formants are analyzed in Assamese vowel phonemes. LPC model is used to identify the formant frequencies.**

Keywords— Formant, LPC, Vowel Phoneme, Spectrum, Filter.

I. INTRODUCTION

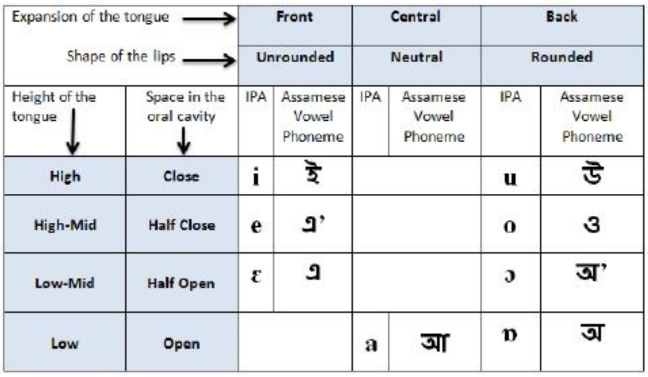
Assamese (IPA: ɔxɔmija) is a native language of Assam which is a major language in the north-eastern India. Its root is Indo-European family of languages. Assamese scripts is derived from the Devanagari scripts consisting thirty nine consonants and eleven vowel symbols which are arranged in a well structured scientific manner[6]. Though there are eleven vowel symbols in Assamese script, but the number of vowel phonemes is only eight.

A **phoneme** is nothing but a single unit of sound that has a meaning in that language. The vowel is the largest phoneme group as the source for vowel is quasi-periodic puffs of airflow through the vocal folds vibrating at a certain fundamental frequency. Each vowel phoneme corresponds to a different vocal tract configuration. Different studies say that the first three formant frequencies measured in the steady-state part of a vowel play an important role in its characterization. The formants of the same vowel uttered by different speakers, in different contexts, at different speaking rates and with different stress patterns, show a lot of variability [5]. From the last few decades, a number of well approaches have been developed for analysis and synthesis of speech signal with a view for speaker/speech recognition. Among those approaches formant estimation is considered as one of the basic models for speech recognition and research.

The **formant model** is used for the determination of formant frequency of Assamese vowels based on the model proposed by L. Welling. The basic idea behind the LPC model is that a given speech sample at time can be approximated as a linear combination of the past speech samples [1,5]. The formant frequencies are computed by taking peak the LPC spectrum. In the first phase during this work, a small database for eight Assamese vowel phonemes is created recording each phoneme 10 times, uttered by ten numbers of Assamese native speakers of equal number of male and female. The recording is done in an acoustic studio in a noise free environment where the utterances are kept normal, stress free and intonation flat.

The written symbols of Assamese vowel scripts and their corresponding vowel phonemes are presented in the following TABLE I.

TABLE I: Assamese vowel phonemes and their positions



Expansion of the tongue →		Front		Central		Back	
Shape of the lips →		Unrounded		Neutral		Rounded	
Height of the tongue ↓	Space in the oral cavity ↓	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme	IPA	Assamese Vowel Phoneme
High	Close	i	ই			u	উ
High-Mid	Half Close	e	এ’			o	ও
Low-Mid	Half Open	ɛ	এ			ɔ	অ’
Low	Open			a	আ	ɒ	অ

II. LPC MODEL AND FORMANT FREQUENCY

Speech signal is formed by the convolution of excitation source and time varying vocal tract components.LPC is a method of separating out the effects of source and filter from a speech signal. The cepstral analysis is the deconvolution of speech into source and system components by traversing through frequency domain. LPC is a tool used mostly in audio signal processing and speech processing for representing the spectral envelop of a digital signal of speech in compressed form, using the information of a linear predictive model [5].

In LP analysis of speech, an all pole model is assumed for the system producing speech signal **s(n)**.The predicted sample can be represented as by (1).

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

Where, $\mathbf{a_i}$ ($i=1, 2, 3, \ldots, p$) are the co-efficients assumed to be constant over the speech analysis frame. The **u(n)** is the normalized excitation and G is the gain of excitation. If $\hat{s}(n)$ is the estimate value of $s(n)$ calculated from the linear combination of past p-samples, then we can get the (2).

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

---(2)

Now, the predictor error can be defined as by (3) given below.

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= s(n) - \sum_{k=1}^{p} a_k s(n-k) \end{aligned}$$

---(3)

For a speech frame of size **m** samples, the mean square of prediction error over the whole frame is given by (4).

$$E = \sum_m e^2(m) = \sum_m \left[s(m) - \sum_{k=1}^{p} a_k s(m-k) \right]^2$$

---(4)

Optimal predictor coefficients will minimize this mean square error. The minimum MSE criterion of E is given by (5)

$$\frac{\partial E}{\partial a_k} = 0, \qquad k = 1, 2, \ldots, p$$

---(5)

Differentiating the Equation (4), we get

$$Ra = r$$

---(6)

Where,

$a = [a_1 \; a_2 \; \ldots \; a_p]^T$, $r = [r(1) \; r(2) \; \ldots \; r(p)]^T$

and ***R*** is a **Toeplitz** symmetric auto-correlation matrix which given by (7).

$$R = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}$$

---(7)

LP residual is the prediction error e(n) i.e. the difference between the predicted speech sample $\hat{s}(n)$ and the current sample $s(n)$ which is given by (8).

$$e(n) = s(n) - \hat{s}(n)$$

$$e(n) = s(n) + \sum_{k=1}^{p} a_k . s(n-k)$$

---(8)

In the frequency domain, the equation (8) can be represented as
in given below (9).

$$E(s) = S(z) + \sum_{k=1}^{p} a_k . S(z) z^{-k}$$

---(9)

i.e. $A(z) = \dfrac{E(z)}{S(z)} = 1 + \sum_{k=1}^{p} a_k . z^{-k}$

---(10)

So it is clear that LP residual can be obtained filtering the speech signal with A(z). Similarly, it can be defined as (11).

$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k . z^{-k}} = \frac{1}{A(z)}$$

---(11)

Since, A(z) is the reciprocal of H(z), LP residual is obtained by the inverse filtering of speech. LP spectrum provides the vocal tract characteristics from where the vocal tract resonances i.e. **formants** can be estimated taking the peaks from the LP spectrum [2, 4].

III. EXPERIMENT AND RESULT

In the processing part, the first action is to capture the signal of vowel utterances that we require. The Audacity software is used to record the vowel utterances in 16,000 Hz in mono format. The silence part is removed manually which overwrites the original signal. The following parameters are considered in the formant analysis-

- Frame length=256 samples
- Frame overlap= 128 samples
- Sampling frequency= 16,000 Hz.
- Window type=Hamming (256 samples)

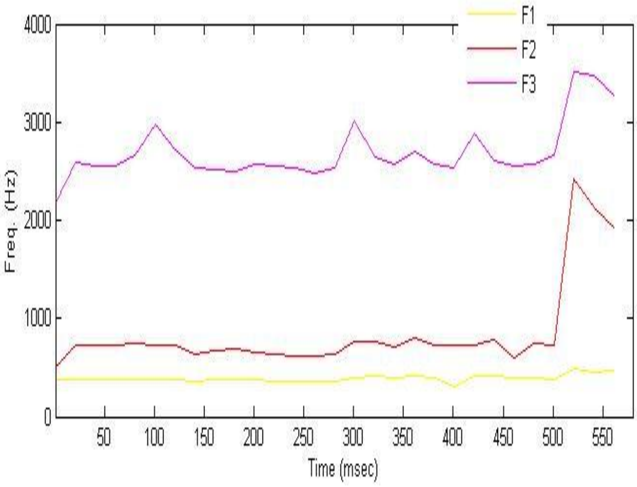


Fig.1. Formants of vowel ও （IPA: /o/) by male speaker.

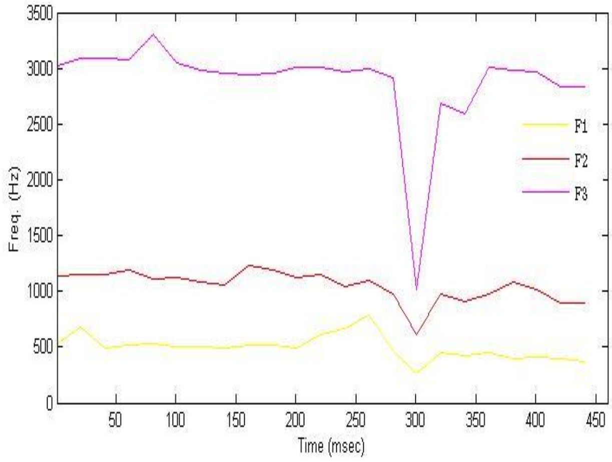


Fig.2. Formants of vowel ও (IPA: /o /) by female speaker.

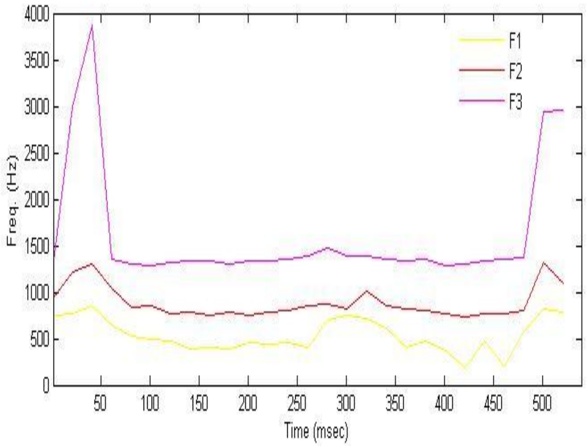


Fig.3. Formants of vowel আ (IPA: /a/) by male speaker.

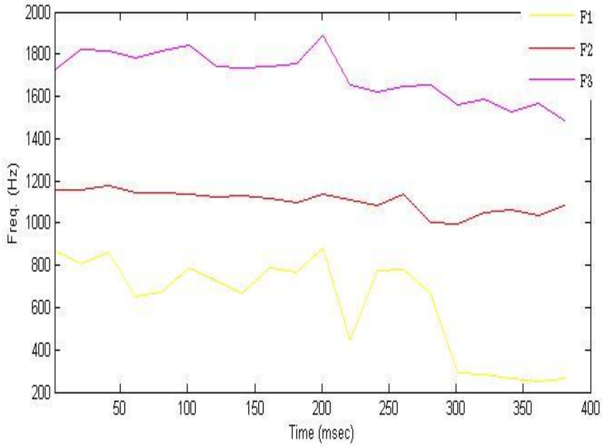


Fig.4. Formants of vowel আ (IPA: /a/) by female speaker.

Englewood Cliffs, NJ,1979, Prentice-Hall.

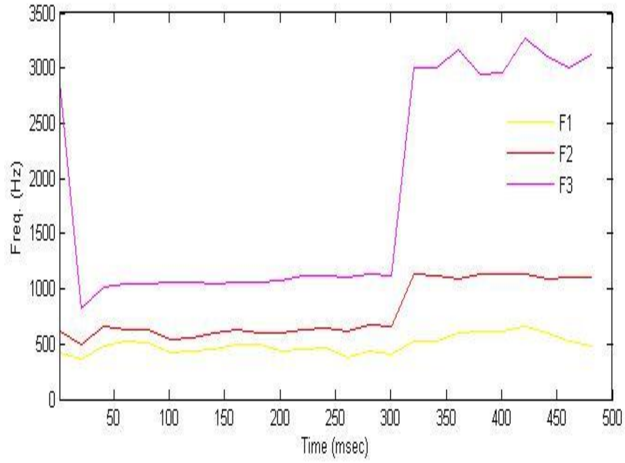


Fig.5. Formants of vowel অ (IPA: /ɒ/) by male speaker.

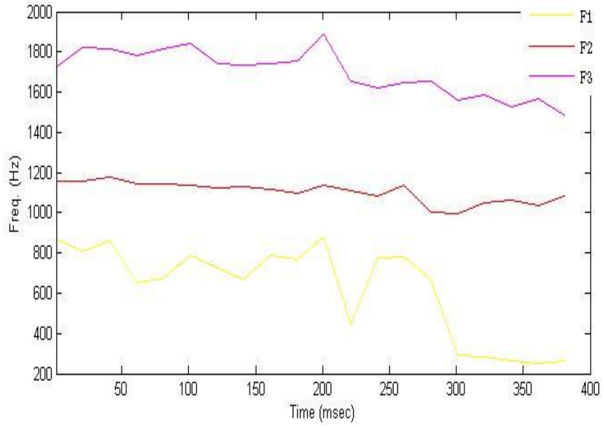


Fig.6. Formants of vowel অ (IPA: /ɒ/) by female speaker.

IV. CONCLUSION

From the analysis of formant frequencies of different Assamese vowel phonemes, we notice that the variation of F1 and F2 with respect to different vowel is quite distinct. Each color line represents a unique formant. In each case the formant values of female speaker is comparatively high than the male speaker. It is also seen that the third formant frequency F3 does not play a crucial role in the identification of a specific vowel spectrum.

REFERENCES

[1] Gold and N. Morgan, “Speech and Audio Processing: Processing and Perception of Speech and Music” ,New York, 2000.

[2] Medhi Bhargab, and P. H. Talukdar. "Isolated Assamese Speech Recognition using Artificial Neural Network." Advanced Computing and Communication (ISACC), 2015 International Symposium on. IEEE, 2015.

[3] Medhi Bhargab, and Pran Hari Talukdar. "Zero Crossing Rate Analysis of Assamese Vowel Phonemes." International Journal of Engineering Research and Technology. Vol. 3. No. 3 (March-2014). IJERT, 2014.

[4] Medhi Bhargab, TALUKDAR P. “Different acoustic feature parameters ZCR, STE, LPC and MFCC analysis of Assamese vowel phonemes”,ICFM 2015.

[5] L.R. Rabiner and R. Schafer, “Digital Processing of Speech Signals”,

[6] Banikanta Kakati, “Assamese, its Formation and Development”, 5th edition, Guwahati, India,LBS Publications, 2007.

[7] F. Jelinek, “Statistical Methods for Speech recognition”, Cambridge, The MIT Press, 1998.