INTERNATIONAL JOURNAL OF RESEARCH – GRANTHAALAYAH
A knowledge Repository
Science

SPECTROGRAM STUDY OF BODO VOWELS

Dr. Jogen Boro*[1]
*[1]Guest Faculty, Department of Bodo, Gauhati University, Assam, INDIA

ABSTRACT

Bodo Language is phonetically rich language but it is not analysed systematically and scientifically still now. So each and every parts of language is to be studied with the help of modern equipments for find out the accurate results. In the age of Information Technology where the Information Technology (IT) is threading the entire into a Global Village with knowledge as the prime currency and sole differentiator, development of appropriate access technology take part vital role for gaining the information. Especially for India, with its multi-lingual requirements and not so fortunate achievements in terms of overall literacy, development of speech technology in each of its recognized language demands utmost attention. Moreover, development of such Speech Technologies in Indian Languages, with their core dependence on linguistic and cultural ethos, need to be developed largely in India. From the present analysis and study of Bodo vowels spectrograms, it is seen that the lower frequency regions for almost all vowels are very clear. It is a very uncommon characteristic observed in case of the Bodo vowel utterances in comparison with other local languages of Assam. So this will help the speech researcher's on Bodo Language in various directions in future.

Keywords:
Bodo, equipments, phonetically, speech, spectrogram, vowel.

Cite This Article: Dr. Jogen Boro, "SPECTROGRAM STUDY OF BODO VOWELS" International Journal of Research – Granthaalayah, Vol. 3, No. 9(2015): 122-128.

1. INTRODUCTION

The speech signal is a slowly time varying signal. When such signals are examined over a sufficiently short period of time, the characteristics of speech signal are fairly stationary; however, over long periods of time, the signal characteristics change to reflect the different speech sounds being spoken**.** Speech waveform consists of sequence of different events. The time variation corresponds to highly fluctuating spectral characteristics over time. A single Fourier transform of the entire acoustic signal cannot capture the time varying frequency contents for all the harmonics present. In order to capture the time varying nature of the speech signal, another Fourier Transform, called Short-Time Fourier Transform (STFT) is used. It consists of a separate Fourier Transform on a piece of the waveform under a sliding window, which is represented by $w\,[n, \tau]$, where $\boldsymbol{\tau}$ is the position of the window centre and $\boldsymbol{n}$ is the number of sample per window.

The Short-Time Fourier Transform of the windowed speech waveform, i.e., **STFT** is given by equation **(1)**:

$$x(\omega,\tau) = \sum_{n=\alpha}^{\alpha} x[n,\tau]\exp[-j\omega n] \qquad \textbf{(1)}$$

Where $x[n\ \tau,] = w[n, \tau]x[n]$ represents the windowed speech segments as a function of the window centre. The spectrogram is a two dimensional representation of the time dependent spectrum in which the vertical dimension on the paper represents frequency and the horizontal dimension represents time. The spectrogram magnitude is given by equation **(2):**

$$S\ [\omega, \tau]= |x(\omega, \tau)|^2 \qquad \textbf{(2)}$$

Basically, there are two types of spectrograms - *(a) wideband spectrogram and (b) narrowband spectrogram*. The difference between these two types of spectrograms is the length of the window ***w* [*n*, *τ*]**. The wideband spectrogram displays good temporal resolution and poor frequency resolution. On the other hand, the narrowband spectrogram displays the good frequency resolution and poor time resolution. The wide band spectrogram corresponds to performing a spectral analysis on 15 milisecond (typically) sections of waveform using a broad analysis filter (125 Hz bandwidth) with the analysis advancing in intervals of 1 millisecond**.**The spectral intensity at each point in time is indicated by the intensity of the plot at a particular analysis frequency. On the other hand, the narrowband spectrogram corresponds to performing a spectral analysis on 50 milisecond sections of waveform using a narrow analysis filter (40 Hz bandwidth), with the analysis again advancing in intervals of 1 millisecond**.**

For voiced speech, the output of a linear time-invariant system with impulse response h[n] and with a glottal flow input given by convolution of the glottal flow over one cycle g[n], with the impulse train, is given by equation (3):

$$p[n] = \sum_{k=-\infty}^{\infty} \delta[n-kp] \qquad \textbf{(3)}$$

In the windowed speech waveform the result can be expressed as

$$x[n, \tau] = w[n, \tau]\{(p(n)*g(n))*h(n)\} = w[n,\tau](p[n]*\hat{h}[n]) \qquad \textbf{(4)}$$

Where, the glottal waveform, over a cycle, and vocal tract impulse response are lumped into, $\hat{h}[n] = g[n] * h[n]$

Using Multiplication and Convolution theorem, the Fourier Transform of the speech is given by,

$$X(\omega,\tau) = \frac{1}{p} w(\omega,\tau) * [H(\omega)G(\omega)\sum_{k=-\infty}^{\infty}\delta(\omega-\omega_k)]$$
$$= \frac{1}{p} H(\omega_k)G(\omega_k)w(\omega-\omega_k,\tau)$$
$$= \frac{1}{p}\sum_{k=-\infty}^{\infty}\hat{H}(\omega_k)G(\omega_k)w(\omega-\omega_k,\tau) \qquad \textbf{(5)}$$

Where $\hat{H}(\omega_k) = H(\omega_k)G(\omega_k)$ and $\omega_k = \frac{2\pi k}{p}$ and $\frac{2\pi}{p}$ is the fundamental frequency.

Therefore, the spectrogram of x[n] can be expressed as

$$s(\omega,\tau) = \frac{1}{p^2}\left|\sum_{k=-\infty}^{\infty}\hat{H}(\omega_k)w(\omega-\omega_k,\tau)\right|^2 \qquad \textbf{(6)}$$

The wideband and the narrowband spectrograms display a great deal of information about the properties of a speech utterance.

Usually, the formant frequencies are greater than the corresponding pitch or fundamental frequency for particular speech signal. While encoding, synthesizing and recognizing the speech signal, the formant frequencies and pitch or fundamental frequencies find extensive use. The time-frequency analysis of speech signal is also used extensively in the study of human speech. The spectrograms which are nothing but the squared magnitude of the STFT, plays important role while visualizing the time-varying frequency content of a speech signal**.**

However, the STFT has limited resolution. This limitation is subsequently overcome the use of mixed time-frequency signal representation, which is substantially different from the spectrogram. This was first proposed by Wigner (1932) and Ville (1958) and the technique is known as **Wigner-Ville distribution (WVD)**. The WVD is the FT of the autocorrelation of the signal obtained from the Hilbert transform of the original speech signal. A major problem with WVD technique is the interferences between two signal components located at different regions in the time-frequency plane**.**

2. SPECTROGRAMS OF BODO VOWELS

The speech signal, though taken as non-stationary signals, is assumed as stationary one for speech periods (10ms to 50ms)**.** A popular way of characterizing the speech signal and representing the information associated with the sounds is via spectral representation. Perhaps, the most popular representation of this type is the sound spectrogram in which a three dimensional representation of the speech intensity, in different frequency bands, over time is portrayed**.** The basic software tools used in the formant study is *Matlab (version7.1).* Other C++ programs were specially developed to check and validate the accuracy of our computational results. **Low-pass and band-pass filtering** operations were implemented with MATLAB filter design functions. This typical window length is chosen for the computation of the spectrum of the vowels utterances in the present study. In ***Figure (1), Figure (2)***, the spectrograms of the six Bodo vowels utterances corresponding to male and female speakers has been shown. From these Figures, the pitch of the six Bodo vowels can be determined from the first line along the frequency axis. However, these frequencies are observed, very roughly, while formants are hardly seen directly from the graph.

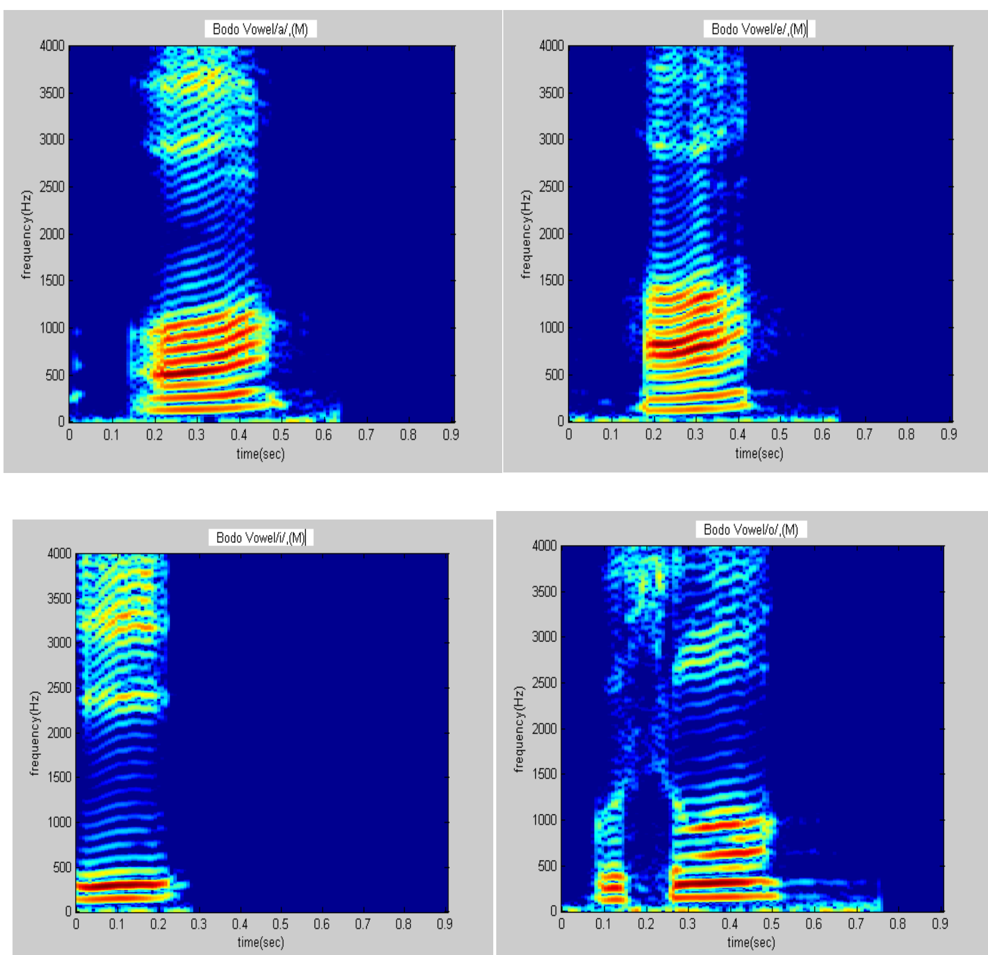
3. ANALYSIS AND RESULTS

From the present analysis and study of Bodo vowels spectrograms, it is seen that the lower frequency regions for almost all vowels are very clear. It is a very uncommon characteristic observed in case of the Bodo vowel utterances in comparison with other local languages of Assam. In the cases of Assamese Language which is the major link language in the north-east region, Rabha Languages, both the higher and lower frequency components are faded out or noisy.

It is found here that the formant frequencies lying within the range of 100Hz to 2 KHz are clear. This may include the 1st, 2nd, 3rd and 4th formant frequencies. Therefore, it is seen from the spectrograms of the six Bodo vowels, that the pitch and first formant frequency and second formant frequency are very distinct corresponding to both male and female informants. Thus, **low pass filter (LPF)** seems more relevant in the present study of pitch and formant frequency analysis of the vowel's speech utterances.

In the frequency range from 500Hz to 2 KHz, there is total disappearance of spectra of the vowels. This is found basically in case of vowels **/e/, /o/, /u/, /i /** for both male and female informants. It is seen from the ***Figure 3(a, b, c and d)*** that the frequency-time spectrograms corresponding to both the male and female speakers, the spectrograms within the time scale ranging from 0.2 sec. to 0.3 sec., are totally viewed as scattered dots. The same patterns found repeated within 0.5sec.to 0.7sec also. The pitch and formant frequency characteristics of the Bodo vowel's spectras are found in agreement, in principle, with the WVD method as mentioned earlier. However, the only exception in the present study with respect to the formant and pitch detection, as proposed by Zaho et al using the WVD technique, is that the gradual blackout of the entire speech spectra, mostly obtained in case of female speakers. This may be attributed to the irregular vocal dynamics and asymmetric vocal fold characteristics of the speakers concerned.

The asymmetric behavior of the vocal fold dynamics may arise due to drift of various vocal fold parameters i.e. effective length of vocal folds, mass and tension of the vocal fold which are controlled by muscle action etc. As the control of these quantities is much slower than the vibration of the folds, so this may produce the uneven and scattered part of the spectrograms.



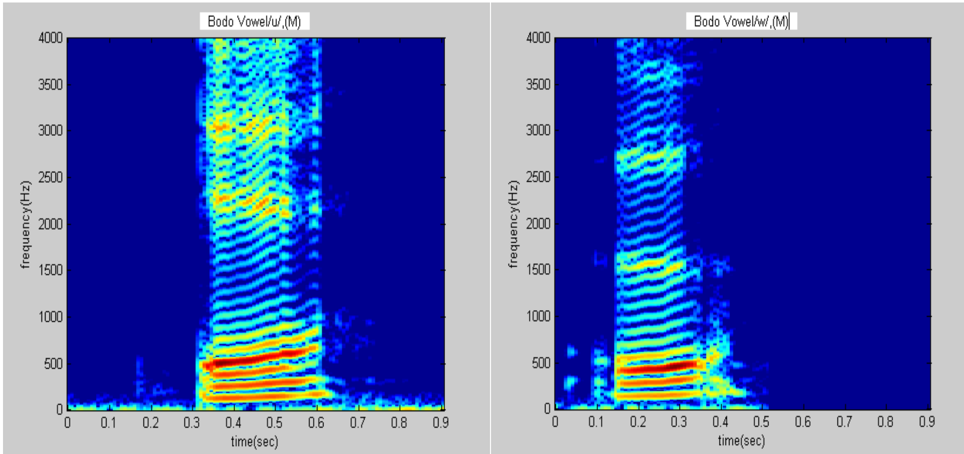
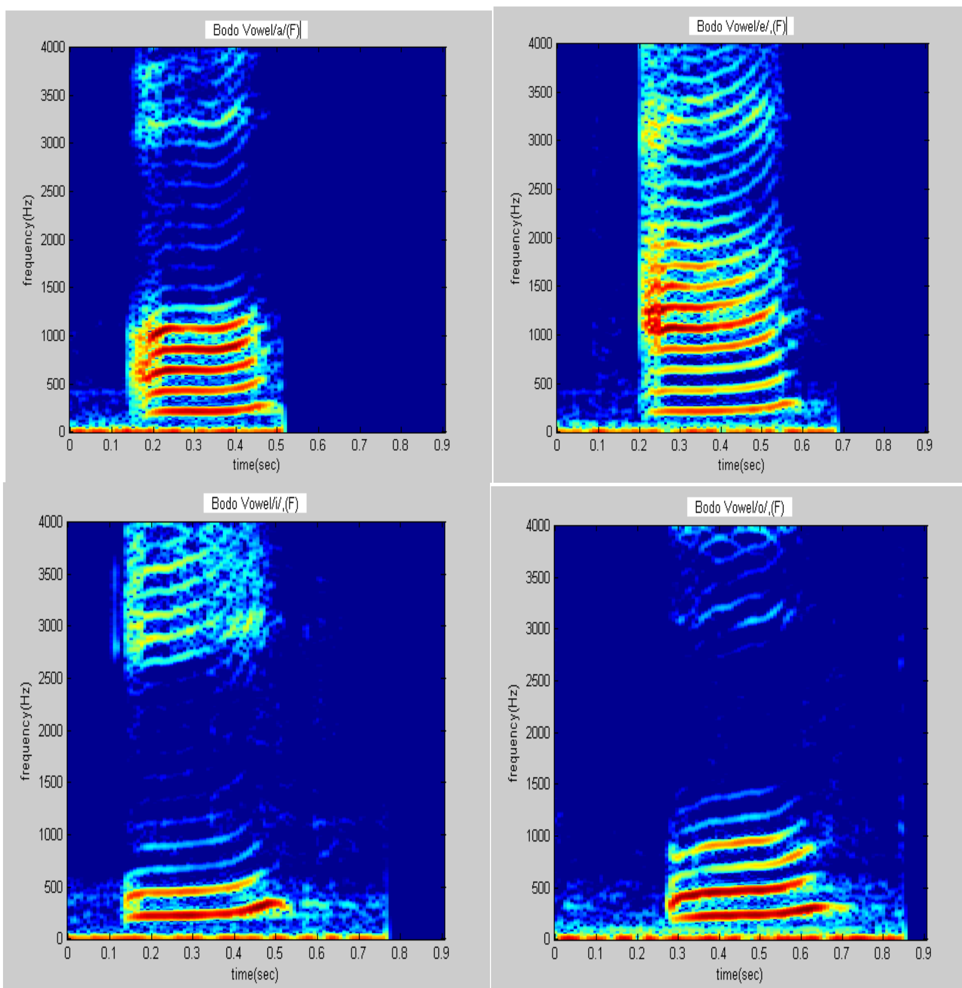


Figure 1: Spectrograms of Bodo Vowels (Male)



ISSN- 2350-0530(O) ISSN- 2394-3629(P)
Impact Factor: 2.035 (I2OR)

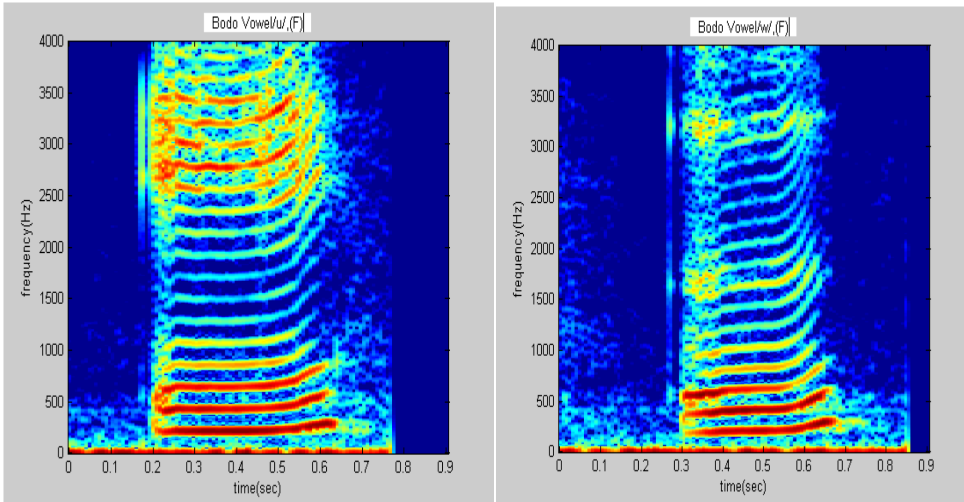


Figure 2: Spectrograms of Bodo Vowels (Female)

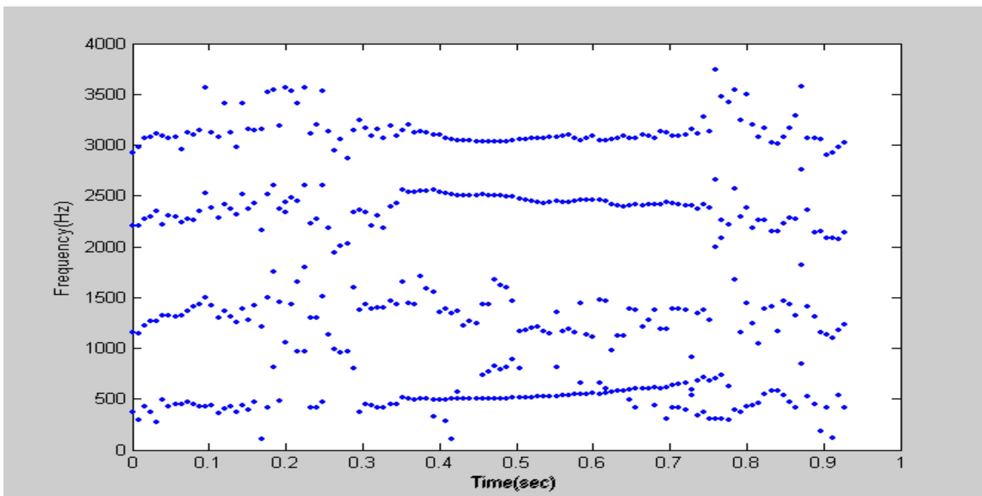


Figure 3(a): Disappearance of spectra for vowel /e/

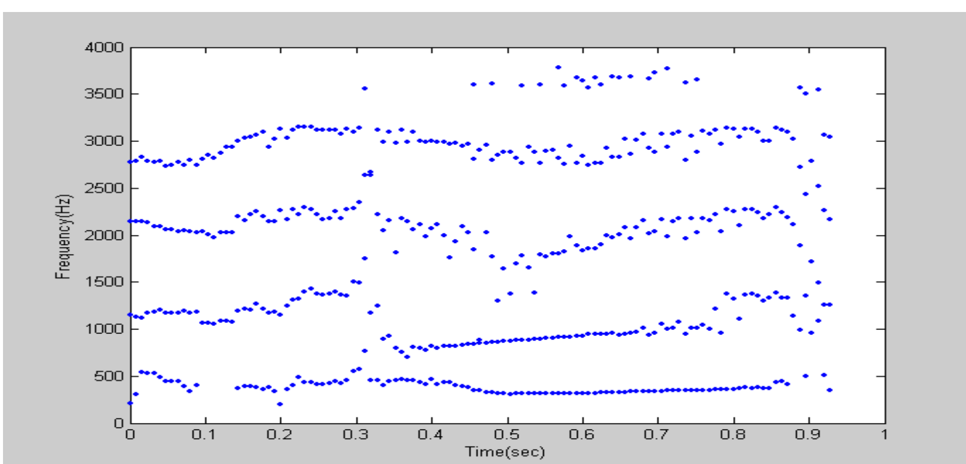


Figure 3(b): Disappearance of spectra for vowel /o/

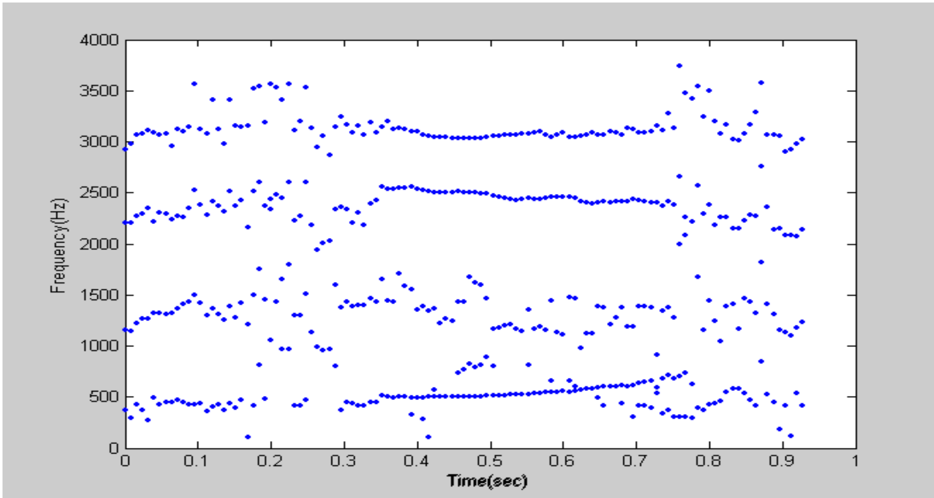


Figure 3(c): Disappearance of spectra for vowel /u/

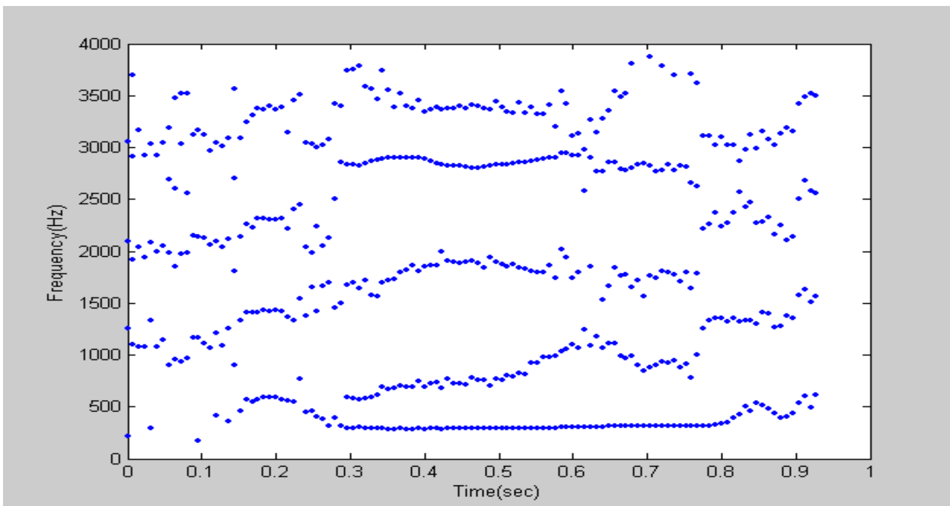


Figure 3(d): Disappearance of spectra for vowel /i/

4. *REFERENCES*

[1] Rabiner, L and Juang, B.: "Fundamentals of Speech Recognitions", Prentice Hall, New Jersey, ISBN:0-13-015157-2, 1993

[2] Key, S.M.: "Modern Spectral Analysis-Theory and Application", Englewood Cliffs, NJ, Percentile Hall, 1998

[3] Ibid

[4] Juang, B.H., Rabiner, L.R. and Wilpon, J.G.: "On the use of band pass filter in speech recognition", IEEE Trans on Acoustics, Speech and Signal processing, 1987, 947-954,

[5] Rabiner, L and Juang, B.: "Fundamentals of Speech Recognitions", Prentice Hall, New Jersey, ISBN:0-13-015157-2, 1993

[6] Bogert, B.P., Healy, M.J.R. and Tukey, J.W.: "The quefrency analysis of time series for echoes: cepstrum, pseudo-auto convariance, cross-cepstrum and saphe cracking", Proceeding of the Symposium on Time Series Analysis (M. Rosenblatt, Ed.) Chapter 15, New York:Wiley, 1963, 2009-243.