The Government of the Union of Myanmar
Ministry of Education

Department of Higher Education (Lower Myanmar)
and
Department of Higher Education (Upper Myanmar)

Universities Research Journal

Vol. 1, No. 3 December, 2008.

Universities Research Journal 2008
Vol. 1, No. 3

Editorial Board

Editors in Chief

Prof. Dr. Saw Hla Myint, Head of the Department of Chemistry, University of Yangon

Prof. Dr. Thein Thein Win, Head of the Department of Chemistry, Yangon Institute of Education

Prof. Dr. Thida Win, Department of Chemistry, University of Mandalay

Prof. Dr. Win Win Thar, Head of the Department of Physics, University of Yangon

Dr. Khin Tint, Associate Professor, Head of the Department of Physics, Yangon Institute of Education

Prof. Dr. Yin Mya, Head of the Department of Physics, University of Mandalay

Prof. Daw Nwe Nwe Win, Head of the Department of Computer Studies,University of Yangon

Editors

Prof. Tin Kyaing, Head of the Department of Chemistry, Yangon University of Distance Education

Prof. Khin Khin Saw, Head of the Department of Chemistry, University of Dagon

Prof. Dr. Aye Aye Tun, Head of the Department of Chemistry, University of Sittway

Prof. Dr. Tin Tin, Head of the Department of Chemistry, University of Pyay

Prof. Dr. Daw Hla Than, Head of the Department of Chemistry, University of Dawei

Prof. Dr Phway Phway, Head of the Department of Chemistry, University of East Yangon
Prof. Dr. Khin Myo Nwe, Head of the Department of Chemistry, University of Hinthada
Prof. Dr. Myint Myint Sein, Head of the Department of Chemistry, University of Mandalay
Prof. Dr. Aye Aye Wai, Head of the Department of Chemistry, University of Magway
Prof. Khin Hnin Lwin, Head of the Department of Chemistry, University of Monywa
Prof . Dr. Sein Sein Aung, Department of Chemistry, University of Panglong
Prof. Dr. Tha Zin, Head of Chemistry Department, University of Loikaw
Prof. San San Wai, Head of the Department of Chemistry, University of Lashio
Prof. Dr. Aye Myatt Mynn, Head of the Department of Physics, University of Mawlamyine
Prof. Dr. Than Hla, Head of the Department of Physics, Yangon University of Distance Education
Prof. Dr. Khin Mar Kyu, Head of the Department of Physics, University of Dagon
Prof. Dr. Thet Tun Aung, Head of the Department of Physics, University of Pathein
Prof. Dr. Daw Thein Win, Head of the Department of Physics, University of Sittway
Prof. Dr. Khin Soe Win, Head of the Department of Physics, University of Pyay
Prof. Dr. Myint Myint Moe, Head of the Department of Physics, University of Dawei
Associate Prof. Dr Ni Ni Zin, Head of the Department of Physics, University of East Yangon
Prof. Dr Tin Tin Win, Head of Department of Physics, University of West Yangon
Prof. Dr. Mya Mya Win, Head of the Department of Physics, University of Hinthada
Prof. Myint Yee, Head of the Department of Physics, University of Maubin
Prof. Dr. Myint Myint Tun, Head of the Department of Physics, University of Hpaan

Prof. Dr. Than Win, Head of the Department of Physics, University of Myeik
Prof. Swe Swe Yi, Head of the Department of Physics, University of Monywa
rof. Dr Thet Thet, Head of the Department of Physics, Mandalay University of Distance Education
Prof . Dr. May Yee Thein, Head of Physics Department, University of Panglong
Prof. Dr. Soe Soe Nwe, Head of the Department of Physics, University of Lashio
Prof. Dr. Khin May Oo, Head of Department of Physics, Myinchan Degree College

Contents

Spectral Analysis on Voices of Myanmar Characters and Words

Ye Chan[1] and Win Win Kyi [2]

Abstract

Voices of Myanmar characters and words have been analysed in either the temporal domain or the frequency domain. The voice samples have been captured by the built-in sound card of personal computer (PC) with sampling rate of 44.1 kHz accompanying with microphone while running LabVIEW. The Fast Fourier Transform (FFT) technique has been used to calculate the spectrum from a signal. It provides a measure of the frequencies found in a given segment of a signal by decomposing it into its sine components. Furthermore, it allows for pitch extraction and the identification of fundamental frequencies, which is an essential component of the experiment. The programme developed uses the necessary tools to record, filter, and analyse different voice samples of some Myanmar characters and words.

Key words: Fast fourier transform, short term fourier transform, spectrogram.

Introduction

Voice analysis includes parameter extraction which consists of preprocessing an electrical signal to transform it into a usable digital form, applying algorithms to extract only speaker-related information from the signal, and determining the quality of the extracted parameters. A microphone first converts the acoustic signature of the uttered word into an electrical signal. An analog-to-digital converter (ADC) converts the electrical signal into a digital representation of the successive amplitudes of the audio signal created by the utterance. The signal is converted from the time domain to the frequency domain which gives the amplitude of the signal in each of a plurality of frequencies over time. Such acoustic signature can also be visualized through display on a spectrogram, as a three-dimensional graph which plots frequency along the vertical axis, time along the horizontal axis, and the intensity of the sound at any given frequency and time by degree of coloration (Julius, 2007; Kahrs, 2002; Mulgrew, 2001).

1. Lecturer, Department of Physics, University of Yangon

2. Demonstrator, Department of Physics, West Yangon University

In this work, voice samples of Myanmar characters and words from different speakers are recorded by microphone and graphical representations of the voiceprint take the form of both FFT spectrums and speech spectrograms because some voice signals do have frequency components that change with time. The Short Time Fourier Transform (STFT) is used to evaluate the frequency contents changing with time.

Theory and Method

A voice in the time domain is captured by 16-bit built-in sound card of personal computer and the voice data are saved as archive samples. The major frequency components of interest are analyzed by FFT and features of the voice are identified in the joint time-frequency domain. Features of a voice in the time domain are either evident in the raw signal, or they become more evident by deriving a secondary signal that has particular properties. The major time-domain parameters of interest are duration and amplitude. Features of the voice are identified by spectral analysis in the frequency domain (Madisetti, 2000).

In the frequency domain, for continuous signals x (t), the Fourier transform X (ω) is

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \qquad (1)$$

Feature extraction in the joint time-frequency domain for a voice begins by dividing the speech signal (a pattern) into frames. An acoustic representation is then extracted for each frame. STFT is a sequence of Fourier Transforms of a windowed signal. STFT provides the time-localized frequency information for situations when frequency components of a signal vary over time, whereas the standard Fourier Transform provides the frequency information averaged over the entire signal time interval.

The STFT is based on the Fourier Transform but the input signal, x(t), is multiplied by a shifted window function, $w(t-\tau)$ as given in equation 2.

$$X(\tau, j\omega) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-j\omega t} dt \qquad (2)$$

Short-time Fourier Transform (STFT) is a sequence of Fourier Transforms of a windowed signal. STFT provides the time-localized frequency information for situations when frequency components of a signal vary over time, whereas the standard Fourier Transform provides the frequency information averaged over the entire signal time interval (Nasser, 2005; Steven, 1999).

Software Implementation

The programmes to capture the voice and to extract the feature of the voice samples are written by LabVIEW. The wave-capture, FFT, peak detection and STFT programmes are developed in this work. Wave capture programme is used to capture the voice signals of different speakers and then put into individual files. The function of the 'sound file read' is to search out the file (*.wav) of the subject, converts the data to mono from stereo and outputs to 'Filter' function to be filtered. The type of filter used in this work is Butterworth band-pass filter. The 'FFT' converts the filtered voice data into frequency domain. Then the 'Normalize' function scale the FFT results within 0 and 1 and plots the power spectrum and original waveform. The normalization of FFT data was done using equation 3.

$$Y = \frac{X-\mu}{\sigma}, \mu = \frac{\sum_{i=0}^{n-1} x_i}{n}, \sigma = \sqrt{\sum_{i=0}^{n-1}(x-\mu)} \quad (3)$$

where Y represents the output sequence normalized vector, X represents the input sequence vector of length n, and x_i is the i^{th} element of X (Cory, 2000; Steven, 1999).

The 'Peak Finder' function analyses FFT data and picks up peaks of the FFT up-to a preset threshold value. The number of data points used in the fit is specified by width. For each peak or valley, the quadratic fit is tested against the threshold level. Peaks with heights lower than the threshold level are ignored. Fig. 1, Fig. 2, Fig. 3, Fig. 4 and Fig. 5 show the normalized FFT spectrum for voice sample 'ka', 'ki', 'ku', 'kay' and 'ke' and the location of their peaks. It is found that peak detector programme is able to identify the peaks on a FFT accurately.

STFT programme computes the signal energy distribution in the joint time-frequency domain using Short-Time Fourier Transform algorithm. The time step, frequency bin, window type and window length are configured for STFT calculation. The 'Hanning' window is used for all frequencies and the FFT block size of 44100 and window length of 512 samples are chosen. Time step specifies the number of samples to shift the sliding window. The time and frequency resolution are adjusted by modifying the window length. A small window length will provide good resolution in the time domain but poor resolution in the frequency domain and vice versa. The step of 2048 samples was used in this analysis.

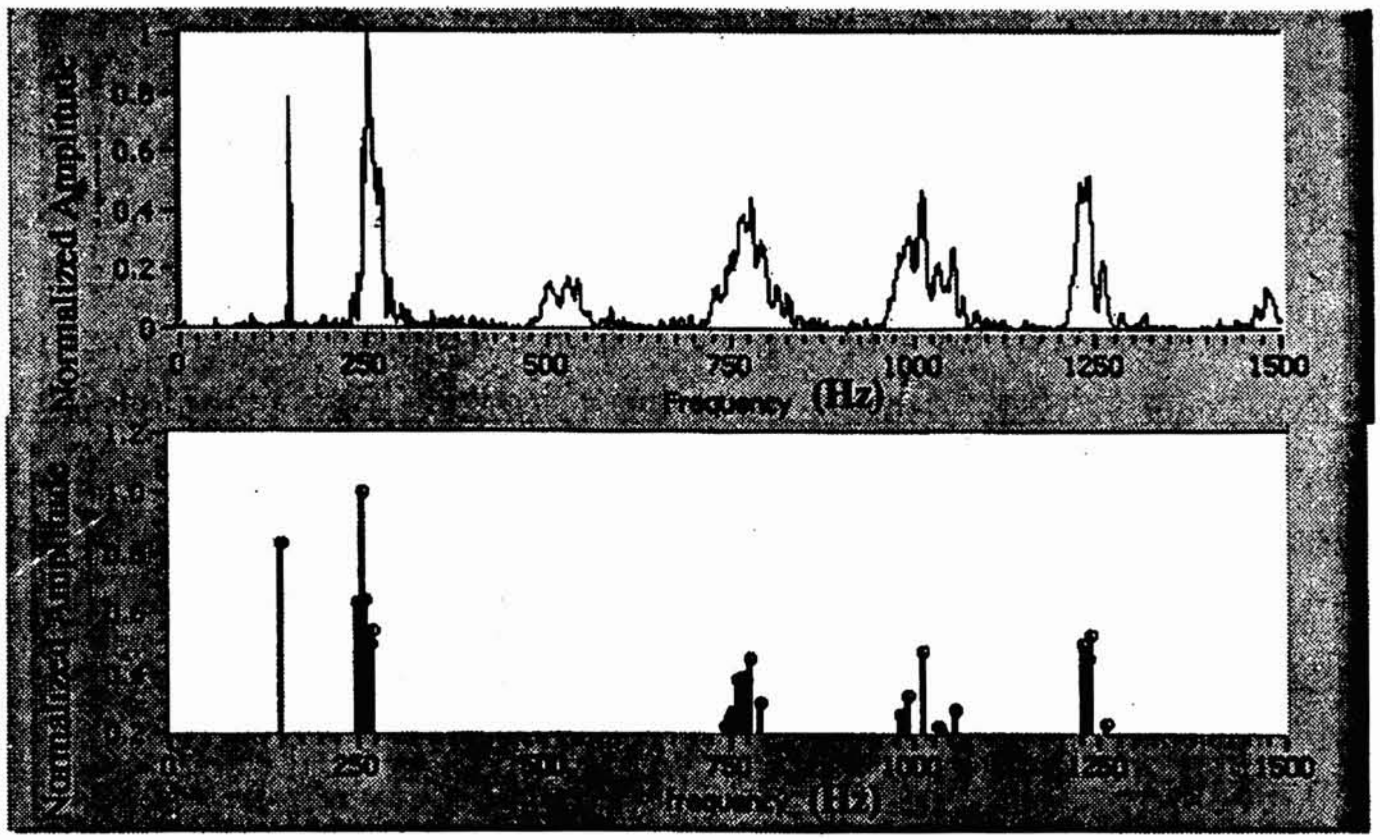

Figure 1. Normalized FFT spectrum of voice 'ka' and its major frequency components

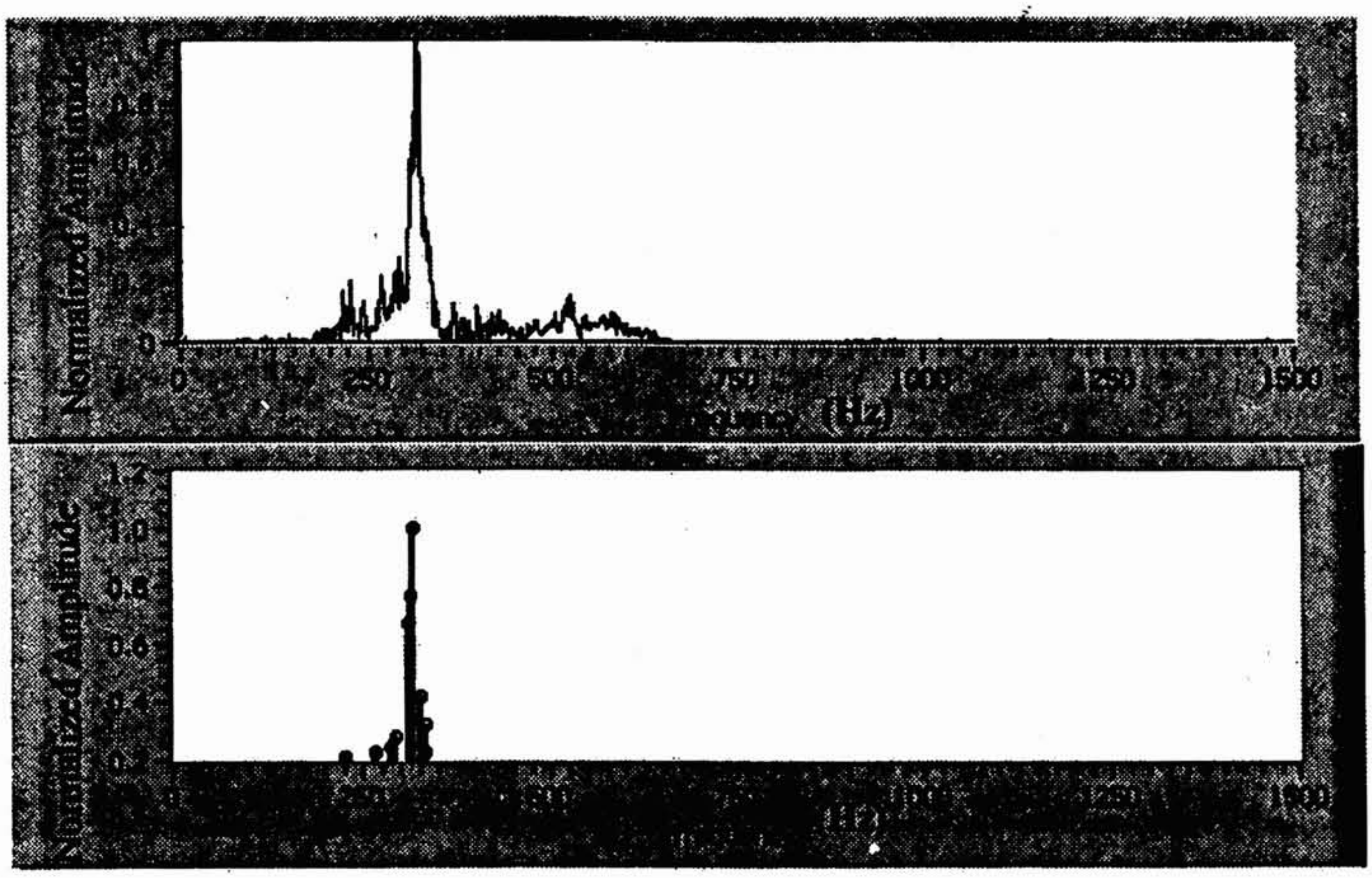

Figure 2. Normalized FFT spectrum of voice 'ki' and its major frequency components

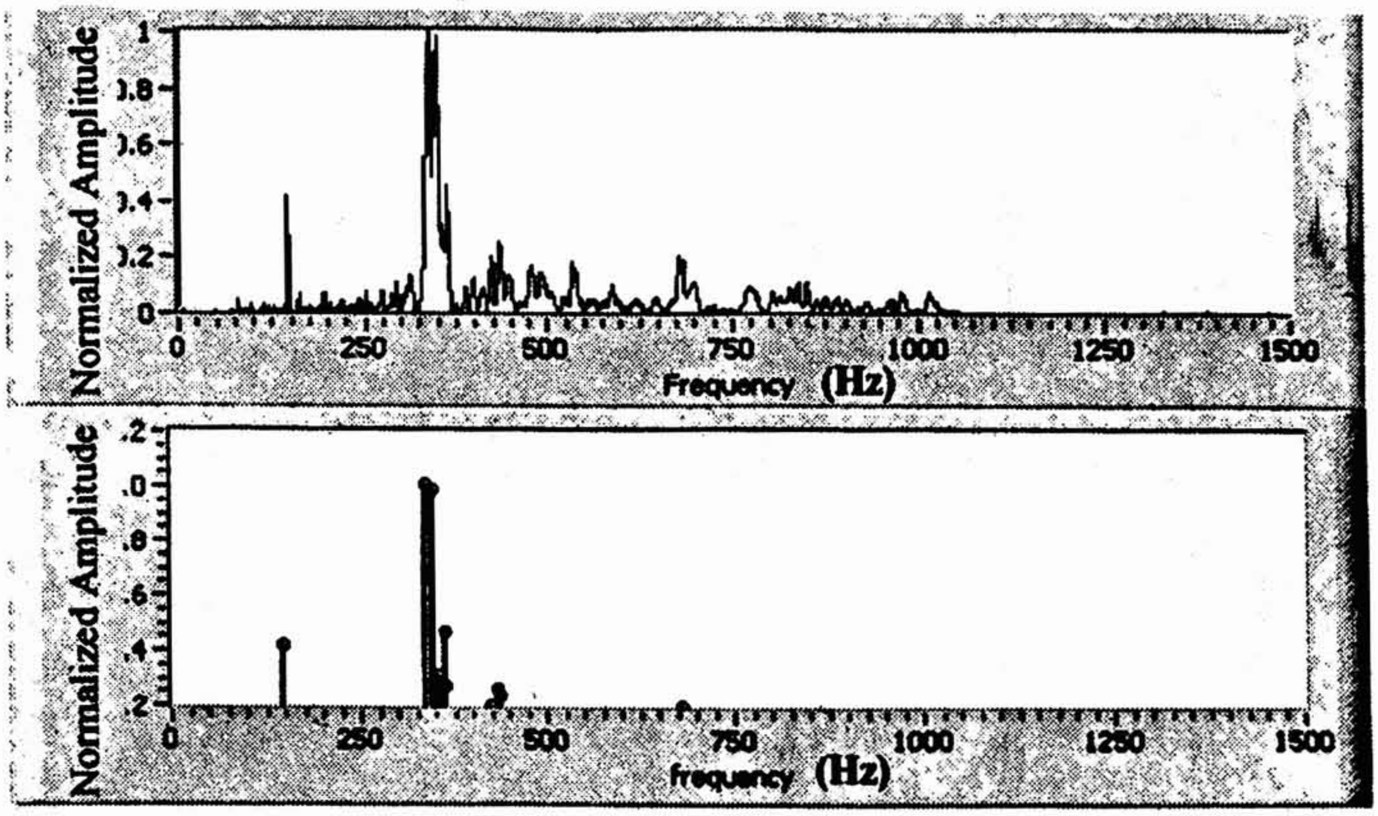

Figure 3. Normalized FFT spectrum of voice 'ku' and its major frequency components

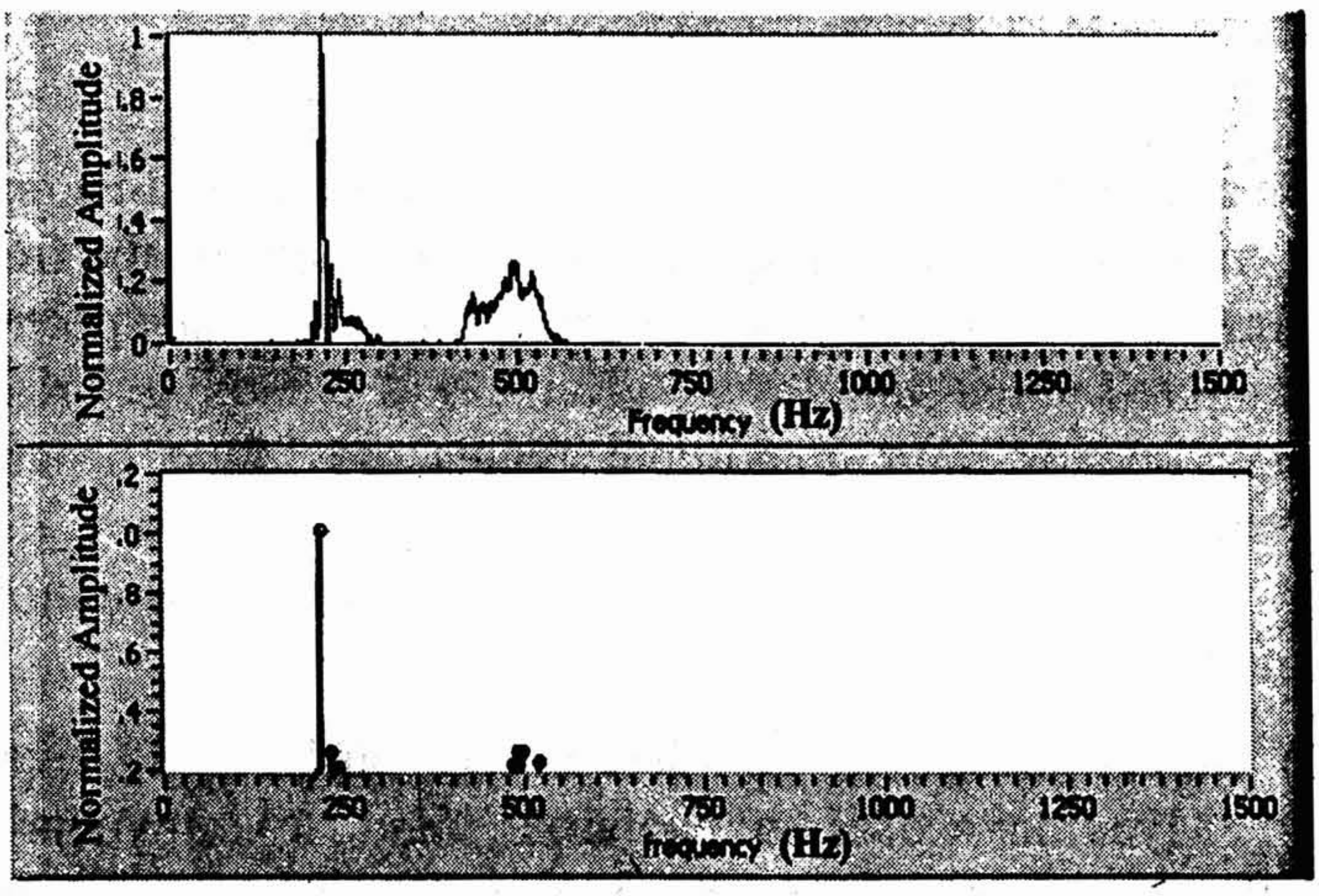

Figure 4. Normalized FFT spectrum of voice 'kay' and its major frequency components

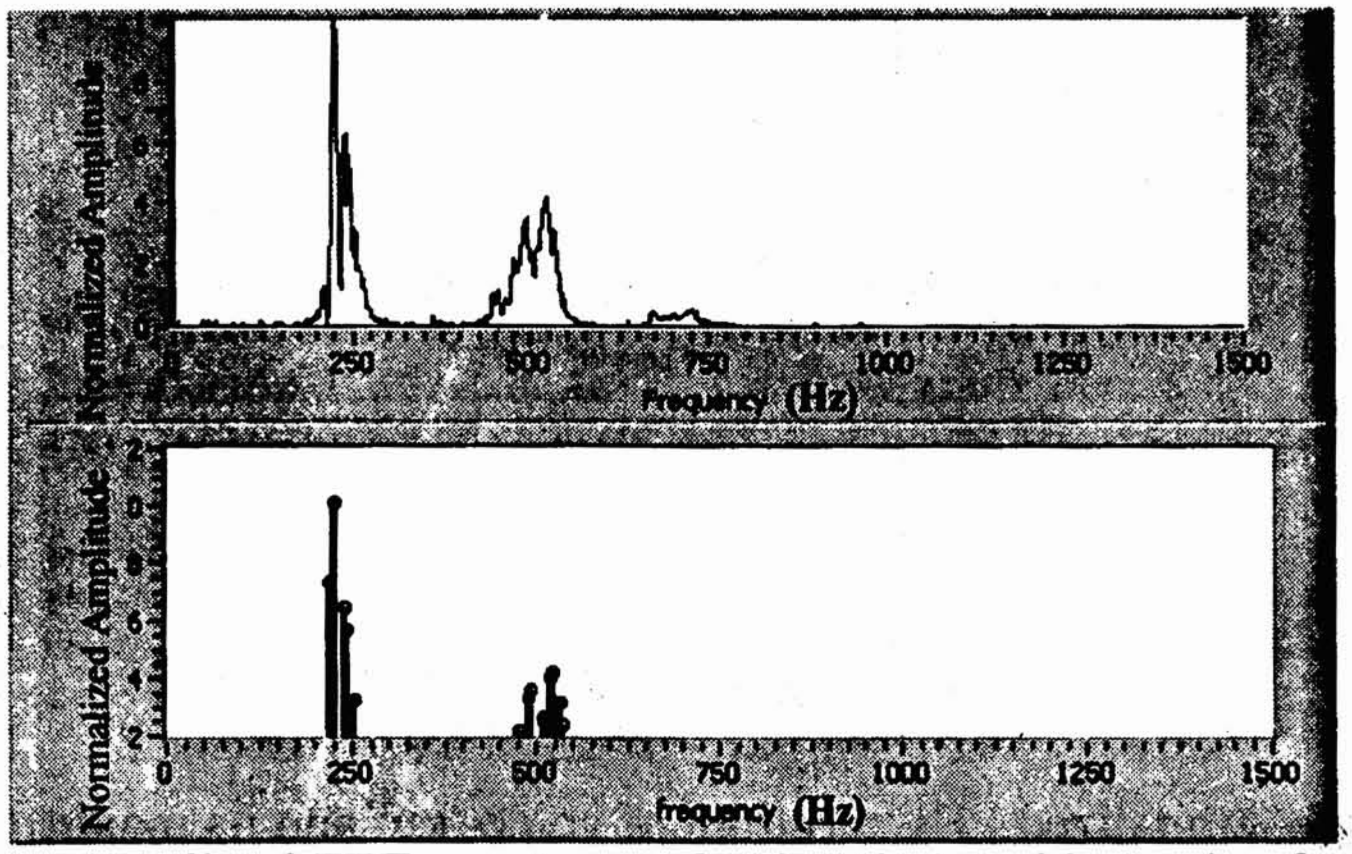

Figure 5. Normalized FFT spectrum of voice 'ke' and its major frequency components

Results and Discussion

The major peaks with maximum intensity in the FFT spectrum are compared with those in STFT spectrum. In this analysis, the frequency and time resolutions are reduced to save computing time. The spectrograms and FFT spectrums of voice samples are given in Fig. 6 and Fig. 7 respectively. In spectrograms, one or more maximum points are found because of poor frequency resolution but the major components do not lose in both spectrums. The arrow pointed on the spectrogram is the major frequency component found in FFT spectrum.

To collect the voice samples, seven females and five males are chosen. The Myanmar words "mingalarbar," and "kwintpyubar" spoken by each person are recorded by the microphone at different times and their FFT spectrum are compared as illustrated in Fig. 8 and Fig. 9. The peak finder and compared program matches the peaks correspond to the deviation in the frequency axis for different subjects. The major frequency components are found at round about 200 Hz and 900 Hz for the word "mingalarbar" spoken by the most subjects. Harmonics for some subjects are found at 400 Hz.

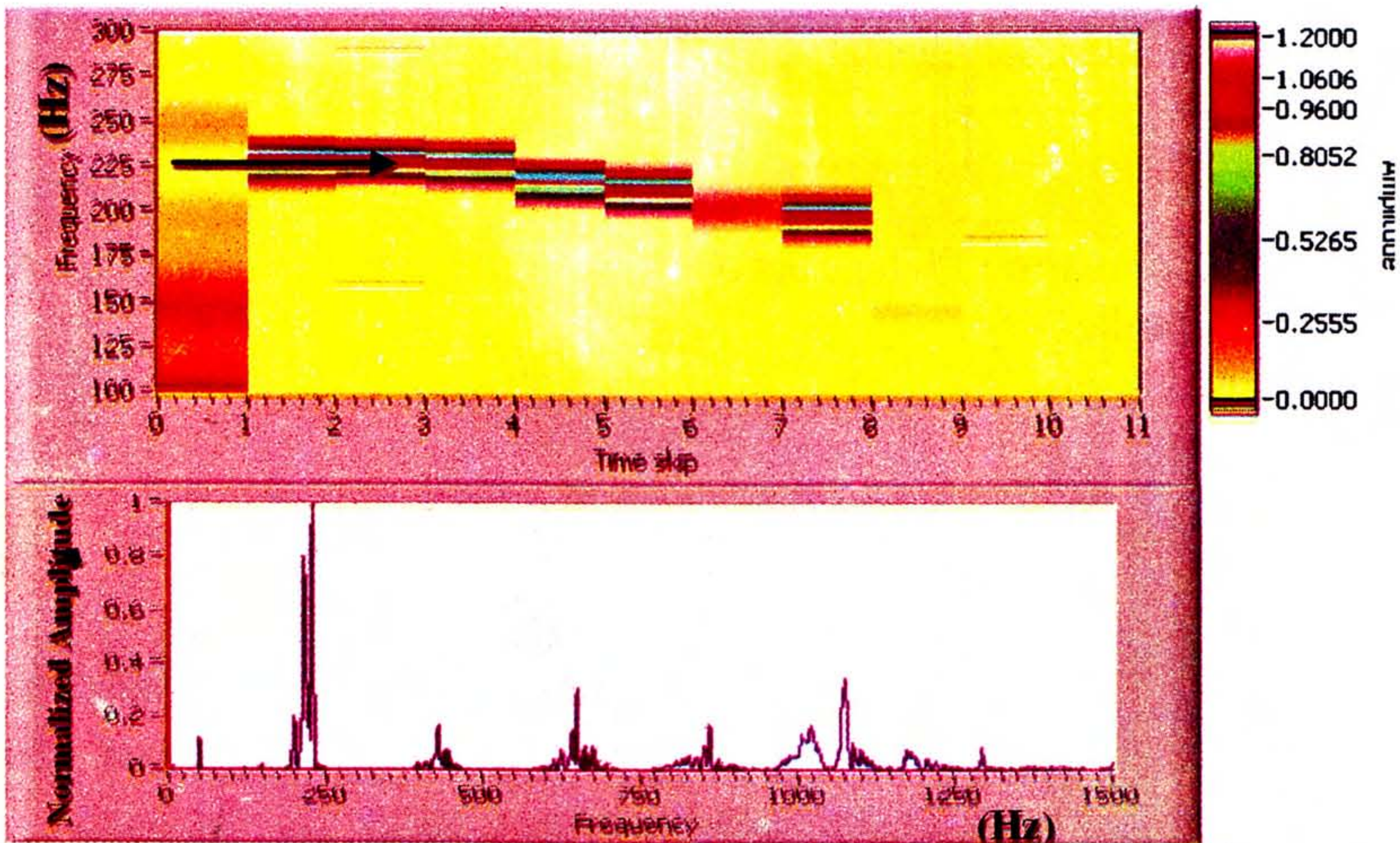

Figure 6. STFT spectrum and FFT spectrum a voice signal spoken by speaker A

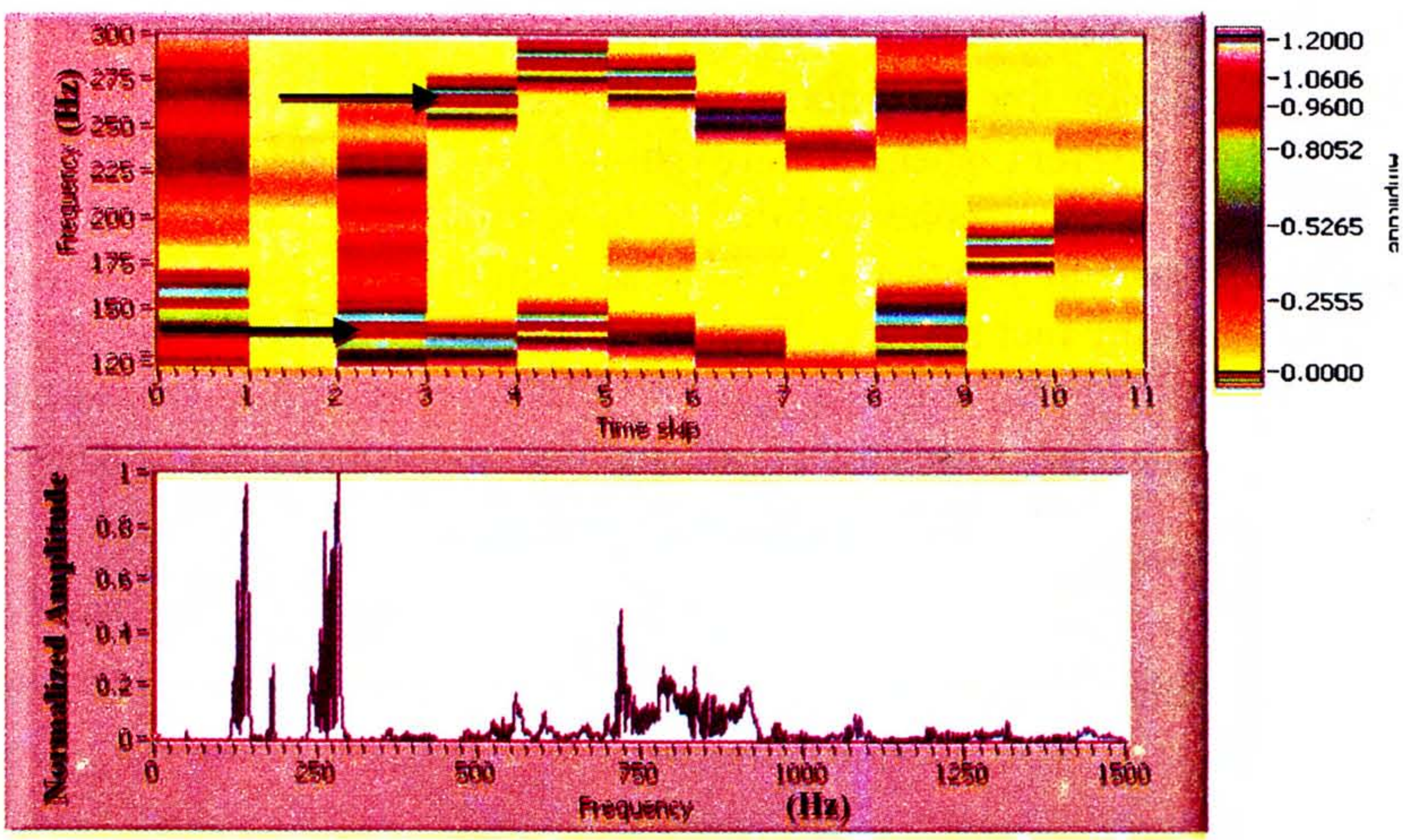

Figure 7. STFT spectrum and FFT spectrum a voice signal spoken by speaker B

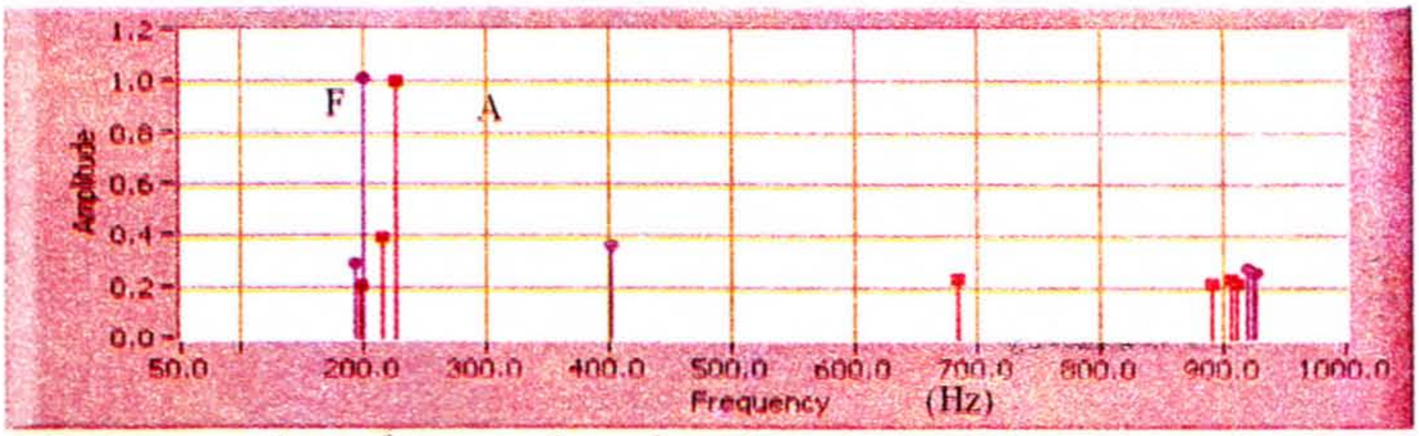

Figure 8. Comparing FFT peaks of "mingalarbra" between speaker A and speaker F

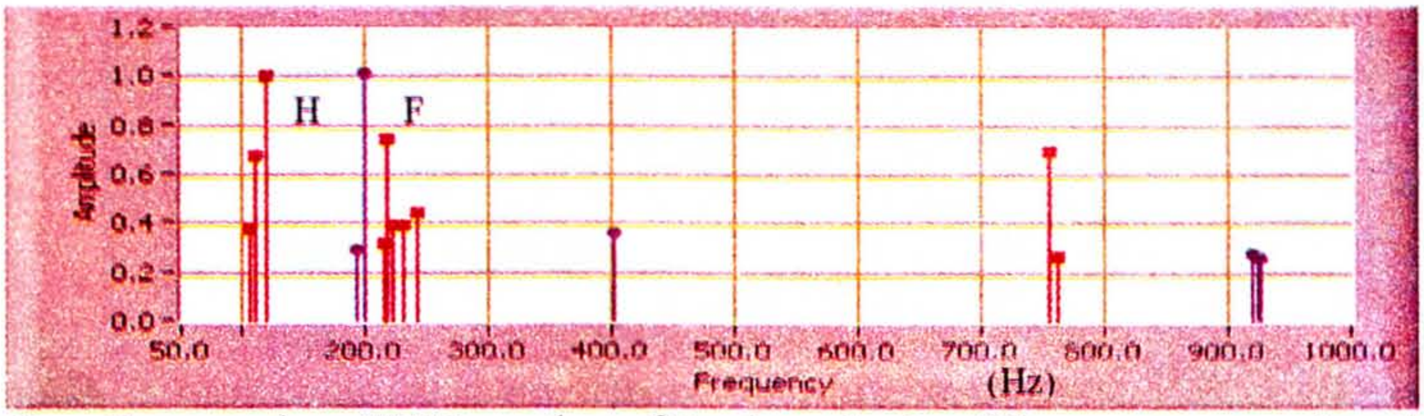

Figure 9. Comparing FFT peaks of "mingalarbar" between speaker F and speaker H

The voice finger prints of words "mingalarbar" and "kwintpyubar" spoken by different speakers are shown in Fig. 10 and Fig. 11. The difference can be a combination of different patterns of the voice and different pitches at which the word was being spoken. Thus, the variation can be due to the fact that the delay in between "min", "ga", "lar" and "bar" , and "kwint", "pyu" and "bar" are different between any two speakers or "min-ga-lar-bar" and "kwintpyubar" all together was spoken at higher or lower pitches.

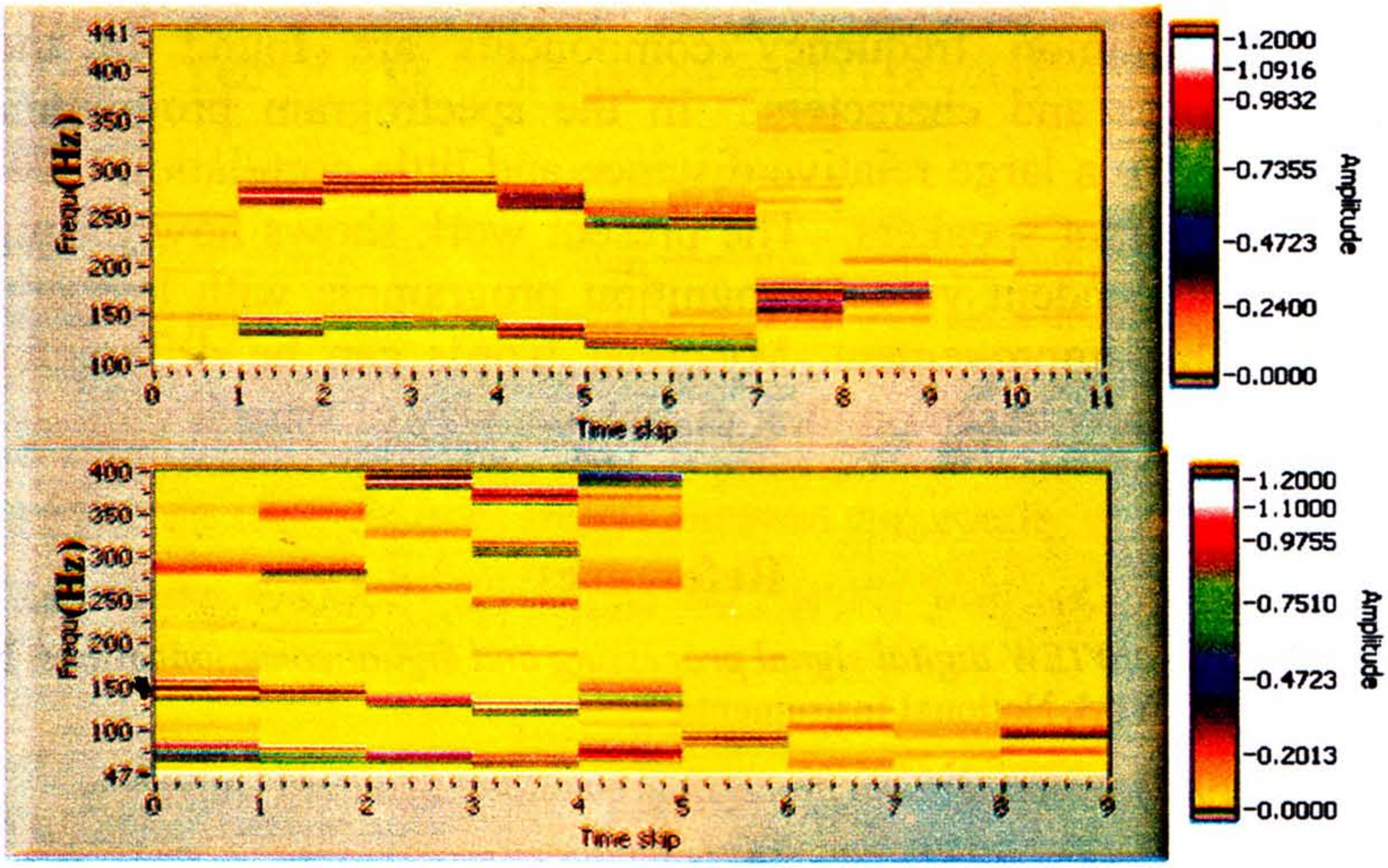

Figure 10. Spectrograms of 'mingalarbar' spoken by two speakers

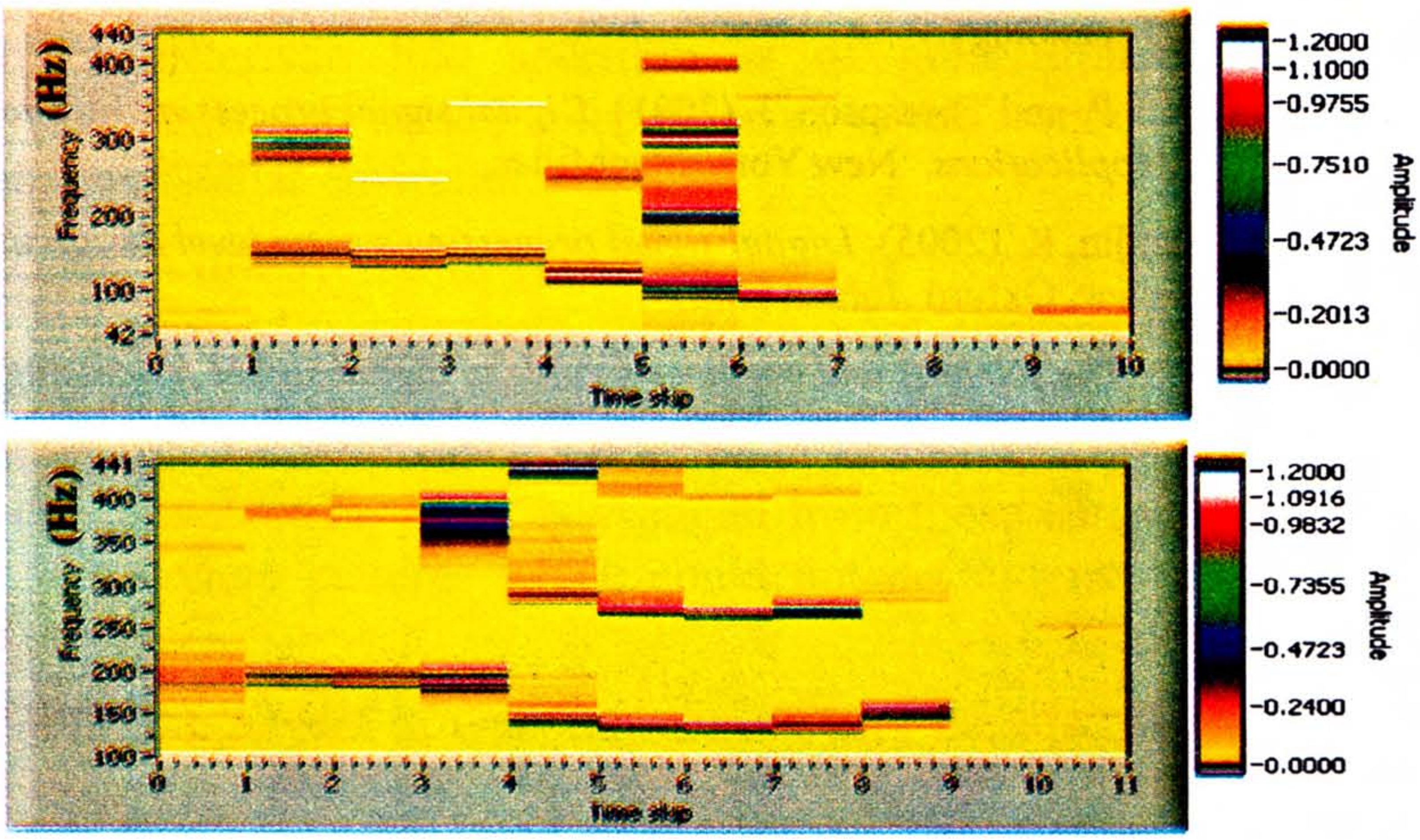

Figure 11. Spectrograms of 'kwintpyubar' spoken by two speakers

Conclusion

The common frequency components are found in the same Myanmar words and characters. In the spectrogram programmes, two fingerprints have a large relative distance and little correlation between the voices for different speakers. The present work shows how to implement the speaker dependent voice recognition programme with less computing cost. With this improvement, Myanmar words can be differentiated the speakers accurately based on their distinctive way of speech.

References

Cory, L. (2000). *LabVIEW digital signal processing and digital communications*. New York National Instruments.

Julius, O. S. (2007). *Spectral audio processing*. stamford Stamford University

Kahrs, M. (2002). *Applications of digital signal processing to audio and acoustics*. New Jersey Rutgers University.

Madisetti, W. (2000). *Digital signal processing handbook*. Atlanta Georgia Institute of Technology.

Mulgrew, B., Grant, P. and Thompson, J. (2001). *Digital signal processing - Concepts and Applications*. New York MacMillan.

Nasser, K. and Namjin, K. (2005). *Digital signal processing system level design using lab-view*. Oxford Jordan Hill.

Steven, W. (1999). *The scientist and engineer's guide to digital signal processing*. California California Technical Publishing.