Detection of Voiced, Unvoiced and Silence Regions of Assamese Speech by Using Acoustic Features

Bidyut Kumar Das#**1** **,** Ajit Das***2** , Utpal Bhattacharjee€**3**

#*Asstt. Professor, Dept. of Computer Sc.,M.C.College, Assam, India*

**Asstt. Professor, Dept. of Computer Sc. and Engineering, Bodoland University, Assam, India*

€*Professor, Dept. of Computer Sc and Engineering, R.G.University, Arunachal Pradesh, India*

Abstract- **Voiced, Unvoiced and Silence regions are detected by extracting certain information from speech signals. Earlier, Zero crossing and Energy values have been used for this purpose. In this paper, Pitch and MFCC are also calculated from speech signal and used along with zero crossing and energy values to identify Voiced, Unvoiced and Silence region. It is tried to find out which parameter can give best result.**

Keyword: Zero-crossing, Energy, Pitch, MFCC

I. INTRODUCTION

A speech signal contains three different regions – voiced, unvoiced and silence. It is to determine whether a particular segment of speech is voiced, unvoiced or silence. Sounds made with the vocal folds together and vibrating are called voiced. Sounds made without this vocal cord vibration are unvoiced [1]. There is no excitation during silence region. A number of measures can be used in determining these three different sections. In this paper four measurements have been used which are Zero-Crossing rate, Energy, Pitch and MFCC.

Zero-crossing parameter is the indicator of frequency at which the energy is concentrated in the signal spectrum. Voiced speech usually shows a low zero crossing count - typically in the range 0 to 30. Unvoiced speech has a concentration of energy at high frequencies and shows a high zero crossing count – typically range from 10 to 100. Zero crossing count for silence would be lower than unvoiced section. [2].

Energy of speech signal is a parameter to classify voiced, unvoiced and silence part. Voiced part of the speech signal has high energy because of its periodicity and unvoiced part unvoiced part has low energy. [3]. At silence part, energy is zero.[4]

Pitch is the perceived fundamental frequency of speech. Silence sections are easily detected and are categorized by a constant DC component, generally zero. Unvoiced section do not contains much periodicity and have very little information of pitch. Voiced section contains periodicity characterized by period of source. This period is called pitch period.[5]. In this paper Autocorrelation function has been used for pitch detection. This function searches for maximum value. When the value exceed 0.3 the section is classified as voiced, otherwise unvoiced.[7]

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency.[6]. It can be calculated as follows:

1. Frame the signal into short frames.
2. For each frame calculate the periodogram estimate of the power spectrum.
3. Apply the mel filterbank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filterbank energies.
5. Take the DCT of the log filterbank energies.
6. Keep DCT coefficients 2-13, discard the rest.

II.   EXPERIMENT

A number of Assamese Language words are recorded and processed. It has been observed that in some types of word a silence region is there inside the processed speech, but in some other words unvoiced section is there. All these three types of sections are tried to detect by using Zero crossing, Energy, Pitch and MFCC.

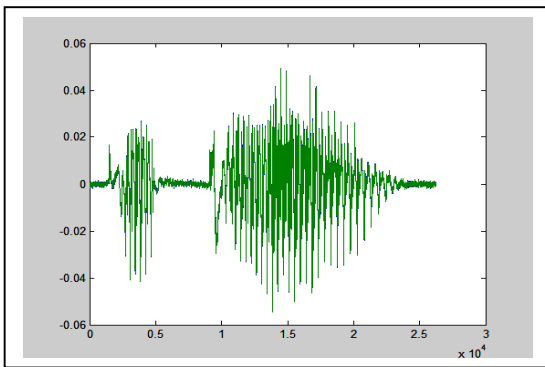Whole speech signal is decomposed in frames each of which contains 500 samples. For each frame Zero crossing, Energy. Pitch and MFCC   are calculated separately.

Fig: 4: Zero crossing Rate (Jonak)

Fig 1: Original Speech Signal (Ketia), (English meaning "When")
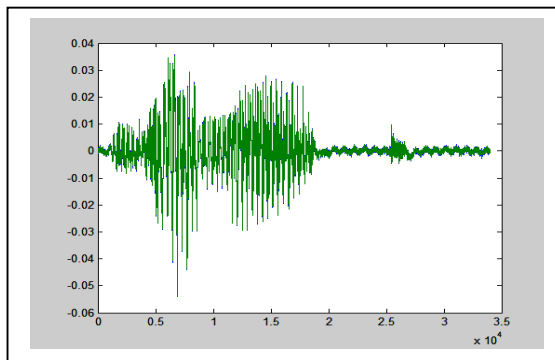
Fig 5: Energy (Ketia)

Fig 2: Original Speech Signal (Jonak)
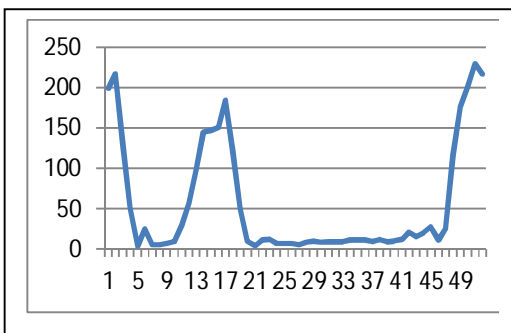
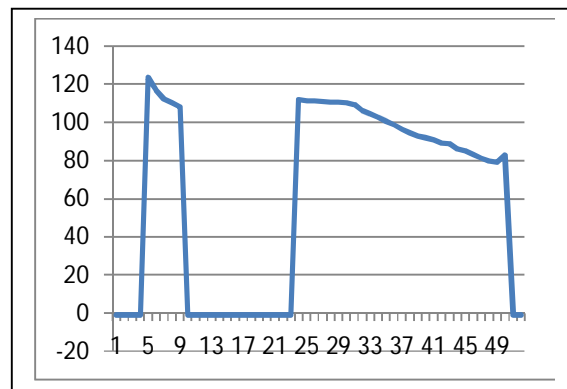Fig 6: Energy (Jonak)

Fig: 3: Zero crossing Rate(Ketia)
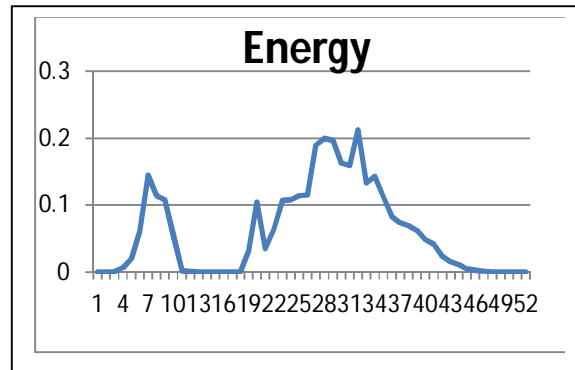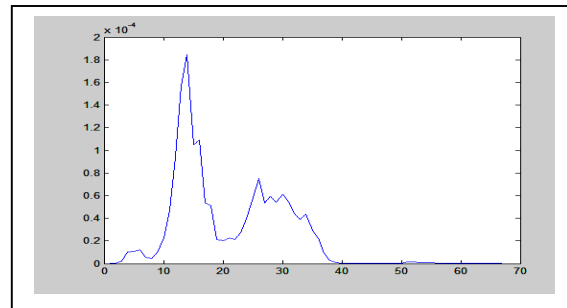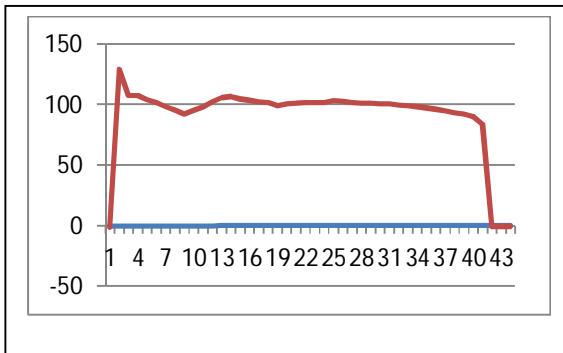

Fig 7: Pitch (Ketia)
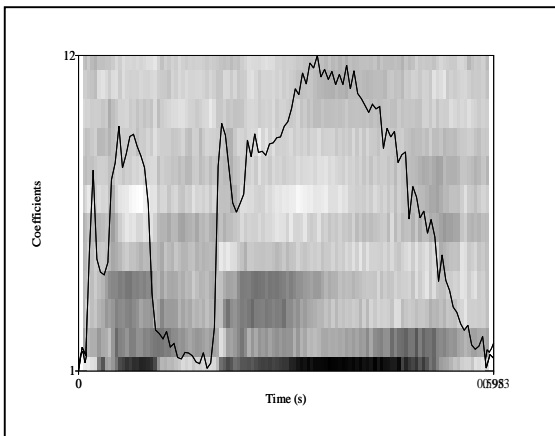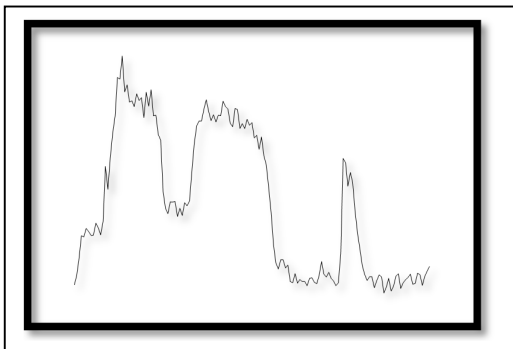
Fig 7: Pitch (Jonak)

Fig 8: MFCC (Ketia)

Fig 5: MFCC (Jonak)

III.   RESULTS

The following results are found from the experiment

i)   For some words only zero crossing become unable to differentiate between the unvoiced and silence region.

ii) For some words Energy is good to identify silence region but show less difference between voiced and unvoiced regions.

iii) Pitch is good parameter to detect Silence region but for some words it is weak to differentiate between voiced and unvoiced regions.

iv) MFCC is very good to detect silence region but for some words it has less difference between voiced and unvoiced region.

CONCLUSION

Use of one or two parameters to detect the three different regions of a speech signal may not guarantee to give best result. The success rate of each parameter may depends on language of the speech and also sometime on speaker.

REFERENCES

[1] D. Jurafsky, J. H. Martin "*Speech and Language Processing*", Publisher Pearson

[2] Bishnu S. Atal, L.R.Rabiner "*A Pattern Recognition Approach to Voiced-Unvoiced-Silence classification with application of Speech Recognition*", IEEE Transactions on Acoustics, Speech, And Signal Processing, Vol.ASSP-24, No. 3, June, 1976.

[3] Bachu, R.G., Koparthi S., Adapa B., Barkana B.D., "*Separation of Voiced and Unvoiced Using Zero crossing rate and Energy of the Speech Signal.*"

[4] M. Greenwood, A. Kinghorn, Dept. of Computer Sc., University of Sheffield, UK, "*SUVing: Automatic Silence/ voiced /unvoiced classification of speech*".

[5]. Moshe Wasserblat, Mikel Gainzay, David Dorranz, Yuval Domb, Dublin institute of Technology, Dublin, "*Pitch tracking and voiced/unvoiced detection in noisy environment using optimal sequence estimation*" ISSC 2008, Galway, June 18-19, Conference paper.

[6] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman "*Speaker Identification Using Mel Frequency Cepstral Coefficients*", **3**rd International Conference on Electrical & Computer Engineering ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh

[7] L.R. Rabiner, M. J. Cheng, A.E. Rosenberg and C.A. McGonegal "*A Comparative Performance Study of Several Pitch Detection Algorithms*", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-24,No. 5, October 1976