Voiced/Unvoiced Decision with a Comparative Study of Two Pitch Detection Techniques

Priyanka Gupta[1], S Sengupta[2]

[1]Dept. of Electrical Engg. , Shri Govindram Seksaria Institute of Technology and Science, Indore, INDIA
[2]Dept. Of Applied Instrumentation and Control, Glasgow Caledonian University, Glasgow, UK

---***---

ABSTRACT-*Pitch tracking is one of the challenging tasks in the field of speech processing and technology. Yet the robustness and accuracy level of the pitch tracking algorithms are unsolved. In this paper a comparative study of two pitch detecting algorithms (PDA) are conducted. These are modified autocorrelation method [12] and On-The- Fly algorithm [1]. In here, these two algorithms are used as a part to decide the voiced-unvoiced parts of speech. One of the difficulties in pitch detection is to differentiate between low level voiced part and unvoiced part of speech. In this paper to reduce this problem On-The-Fly algorithm [1] is proposed and shown by a comparative study with modified autocorrelation method [12]. On-The-Fly [1] is basically a time based algorithm used to estimate the pitch of a speech signal, divided into frames. Therefore, the results of these two methods are proved helpful in making decision of effective voiced/unvoiced speech.*

Key Words: Pitch detecting algorithms, pitch of a speech, autocorrelation, On-The-Fly Algorithm

1. INTRODUCTION

Pitch detection plays a vital role in various speech processing techniques such as speech and speaker recognition, speech synthesis, speech coding etc. The fundamental frequency (f0) is the main cue of pitch. However, it is difficult to build a reliable statistical models involving fundamental frequency f0 because of pitch estimation errors and discontinuity of the f0 space [9].

In the past years, there exist several pitch detecting algorithms (PDA) but the problems remain unresolved. Various techniques which are commonly used for this purpose are Energy Based, Cepstral based, pitch in the difference function and autocorrelation based. In this paper the two PDA algorithms are tested, the On-The-Fly based [10] and autocorrelation based [4][5][9].These two algorithms are further applied to determine the voiced and unvoiced part of speech signal.

Speech can be subdivided into voiced and unvoiced regions. The classification of speech signal into voiced, unvoiced provides a preliminary acoustic segmentation for speech processing applications such as speech synthesis, speech enhancement and speech recognition[3].Speech is composed of phonemes which are produced from the air passing through the vocal folds and vocal tract. Voiced signals are produced when vocal chords are vibrating and air cannot pass through it. These signals are periodic and can be extracted easily. Unvoiced signals do not imply on vocal chords and air can pass through these chords. Unvoiced signals are non-periodic.

Various techniques have been developed to determine the voiced/unvoiced part of speech. One of the techniques used in this paper is by the use of zero crossing rates [3].In this paper the waveform of a speech signal is divided into frames and then the algorithm for voiced/ unvoiced separation is applied. The decision for voiced and unvoiced exist between 0 and 1 i.e.1 for voiced and 0 for unvoiced.

2. PITCH DETECTION ALGORITHMS

2.1 Autocorrelation Method

The autocorrelation is a short-term calculation of a speech waveform. The basic idea behind an autocorrelation method came from the probability theory, which indicates the linear relationship between any two random variables spaced n apart i.e. the changes in one random variable are accompanied by changes in the other. In terms of signal, the autocorrelation function is the relation waveform with itself. This concept of autocorrelation is used for pitch detection.

The mathematical definition of the autocorrelation function is given by Equation (1) for an infinite duration function x[n]

$$R_x(v) = \sum_{-\infty}^{\infty} x[n]\, x[n+v] \qquad (1)$$

Equation (2) for a finite duration function x' [n] of size N

$$R_{x'}(v) = \sum_{n=0}^{N-1-v} x'[n]\, x'[n+v] \qquad (2)$$

To make the signal closely approximate a periodic impulse train we must use some kind of spectrum flattening. To do this we have been chosen to use “Center clipping spectrum flattener” [10].

2.1.1 Center clipping spectrum flattener

In this case the clipping limits are set to ±30% of the absolute maximum amplitude of the waveform are taken to get the high and low pitches only. Mathematically, the limits are given by Equation (3): CL=% of Amax (e.g. 30%)

$$y(n)=\begin{cases} x(n) - C_L, & x(n) > C_L \\ 0, & x(n) <= C_L \end{cases} \qquad (3)$$

2.1.2 Implementation

In this paper we used a modified version of autocorrelation method for pitch detection which is based on center clipping to get a periodic waveform. The block diagram of the pitch detection algorithm using autocorrelation is shown in figure. In this method, firstly the speech signal is divided into frames of window size of 20ms then step-by-step each frame is analyzed. Before analysis, the speech is passed through low pass filter at 900Hz. Calculate the clipping level (CL) and then perform center clipping. The center clipping level is set which is 68% of the maximum input samples. Autocorrelation function is calculated after performing clipping level. Determine the maximum value in autocorrelation function. Then calculate the average pitch of whole sample.

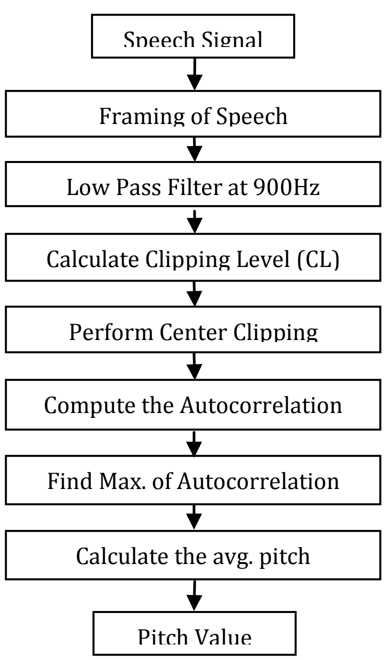

Figure 1. Block Diagram of Pitch Detection Algorithm using Modified Autocorrelation Technique

2.2 On-The-Fly Method

On-The-Fly method is a time-domain pitch detection technique. In time domain, the pitch period is computed directly from the speech waveform. On-The-Fly method is based on difference function (dt). Mathematically, it is represented as

$$d_t(\tau) = \sum_{j=1}^{W-\tau}(x_j - x_{j+r})^2 \qquad (4)$$

where, W is the length of window, t is time domain calculation,

In this difference function, instead of multiplying the signal with its delay a difference signal is created between the original and the delay part of speech and at each delay value the absolute magnitude is taken [9].Here in this method the fundamental frequency is located at the global minima of the difference function. Due to the difficulty in locating the fundamental frequency, the output of the difference function is incorporated. The eigen values of the difference function are given by:

$$\lambda_1(\tau) \propto \sum_{j=1}^{W-\tau}(x_j - x_{j+r})^2 \qquad (5)$$

$$\lambda_2(\tau) \propto \sum_{j=1}^{W-\tau}(x_j - x_{j+r})^2 \qquad (6)$$

These eigen values and vectors helps in deriving the new incorporated difference function given by:

$$d'_t(\tau) = \frac{\lambda_2(\tau)}{\lambda_1(\tau)+\lambda_2(\tau)} = \frac{\sum_{j=1}^{W-\tau}(x_j - x_{j+r})^2}{2.\sum_{j=1}^{W-\tau} x_j^2 + x_{j+r}^2} \qquad (7)$$

2.2.1 Implementation

On-The-Fly method is time domain method and based on difference function. Firstly, speech signal is segmented into frames and determine the original and the new incorporated difference function using Equation (4) and (7). Following the difference equation find the global minima and 4 local minima. The no. of local minimas are maximum then the signal considered as noise and does not proceed further. Now, the parabolic interpolation is carried out to minimize the quantization error. After which the threshold values are calculated which helps in locating the fundamental frequency of speech signal. Finally, the fundamental frequency or pitch is computed.

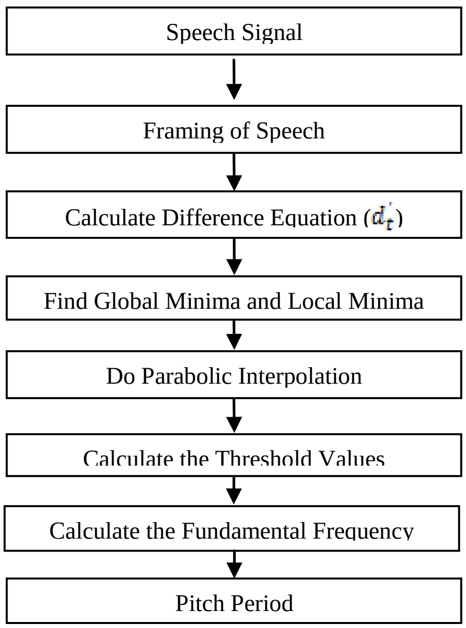

Figure 2. Block Diagram of Pitch Detection Algorithm using On-The-Fly Algorithm

3. VOICED/UNVOICED DETECTING ALGORITHM

Speech is subdivided into voiced and unvoiced regions. If the input excitation is nearly periodic impulse sequence, then the corresponding speech looks visually nearly periodic and is the Voiced Speech. The periodicity of the vocal chords vibration is termed as pitch of the sound vowels sound are one of the example of Voiced sounds [13].If the excitation is random noise-like. Then the resulting speech will also be random noise-like without any periodic nature and is Unvoiced Speech. The vocal folds are open for this sound. Several methods have been developed to classifying the voiced/unvoiced regions of speech. In this paper zero crossing rate method [3] is used to decide the voiced/unvoiced regions. The zero crossing count is an indicator of the frequency at which energy is concentrated in the signal spectrum [3].

3.1 Zero Crossing Rate Algorithms

The zero crossing rates is the rate of changes in the speech waveform crosses the zero axis or the number of times the waveform transition from positive to negative.

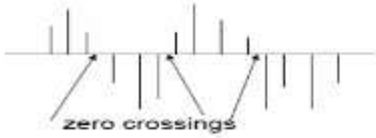

Figure3. Definition of zero-crossing

In this analysis the voiced/unvoiced decision is performed using zero crossing rates. The speech signal is divided into several segments and zero crossing rates are computed for each and every segment. According to the theory the zero crossing rate value are low for voiced part and high for unvoiced part.

Mathematically, the zero crossing rate(ZCR) value is given by equation(8):

$$ZCR = \sum_{-\infty}^{\infty} | sgn[y(n+1)] - sgn[y(n)] | \, w(m-n) \quad (8)$$

Where, y(n) is the signal, (m-n) frame size and sgn is the signum function is given by

sgn[y(n)]=1; y(n)>0

sgn[y(n)]=0; y(n)=0

sgn[y(n)]=-1; y(n)<0 (9)

w (n) = 1/2 (10)

In this design a threshold value is chosen and compared with the zero crossing count and frames less than zero crossing count is set to '1' otherwise '0'.

3.2 Magnitude Sum Function

Magnitude sum function is used to calculate the absolute sum of the filter output which is used for pre-emphasis filtering. Butterworth filter is used for pre-emphasis filtering which increases the magnitude of higher frequencies. This sum function takes part in taking detecting the voiced/unvoiced part of speech. According to the theory a threshold value for magnitude sum function is calculated and if the function value is greater than threshold value than the part of speech is voiced or otherwise unvoiced.

3.3 Implementation

The main objective of this algorithm is to classify whether the part of speech is voiced or unvoiced. In this algorithm the speech is segmented into frames of size 30ms and sampled 8000Hz. Followed with the segmentation of speech sample, pre-emphasis filtering is carried out. In here a high pass filter is used which increases the relative magnitude of the higher frequencies with respect to the lower frequencies. After pre-emphasis filtering the normalized sampling frequency for Butterworth filter is computed to determine the filter coefficients then calculate the sum of all the output magnitudes from the filter. Now the zero crossing count with the help of equation (8) is calculated. In the next part of algorithm a pitch period is calculated with the help of a pitch detection technique. In this paper, two pitch detection techniques i.e.

Modified Autocorrelation and On-The-Fly techniques are used for detecting the pitch period of a speech signal. The purpose of using these two algorithms is to show the better performance amongst the two for detecting the voiced\unvoiced. The three parameters i.e. magnitude sum function, zero crossing rate count and pitch of the signal is calculated. The threshold values of all the three parameters are chosen. As per the theory if zero crossing count is less than the threshold value then the part of speech is voiced. Similarly, the magnitude sum function and pitch period is greater than their threshold value then that part is set as voiced or otherwise unvoiced. In this design, the voiced part is set to 1 or unvoiced part is set to 0.

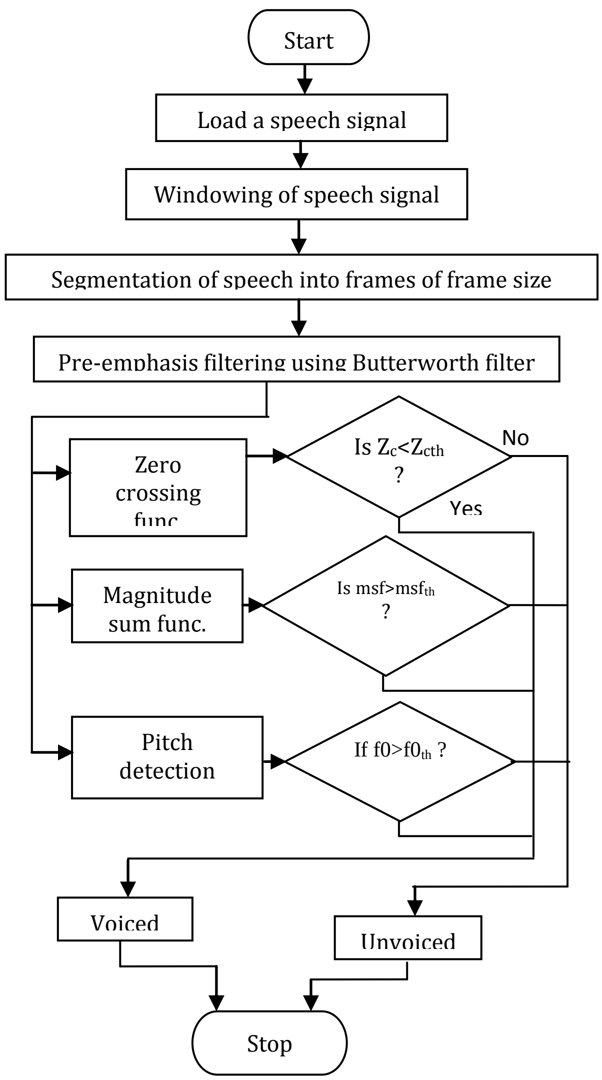

Figure 4. Flow chart of detecting voiced/unvoiced part of speech

4. RESULTS AND DISCUSSION

4.1 Modified Autocorrelation Method

In this paper a sample of a male voice is taken with a sampling frequency of 8000Hz. According to the theory in this method sample is center clipped and limits are set to ±30% of the absolute maximum amplitude of the waveform. The approximate results are shown in Figure5.

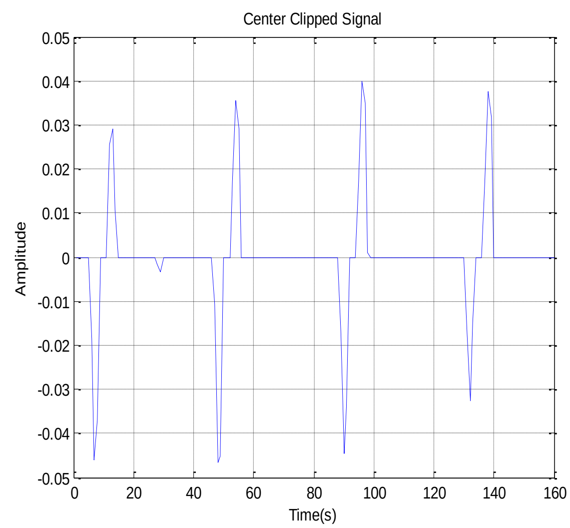

Figure 5. Center clipped part of speech sample

The autocorrelation function is calculated using equation (2).The autocorrelated signal is shown in Figure6.

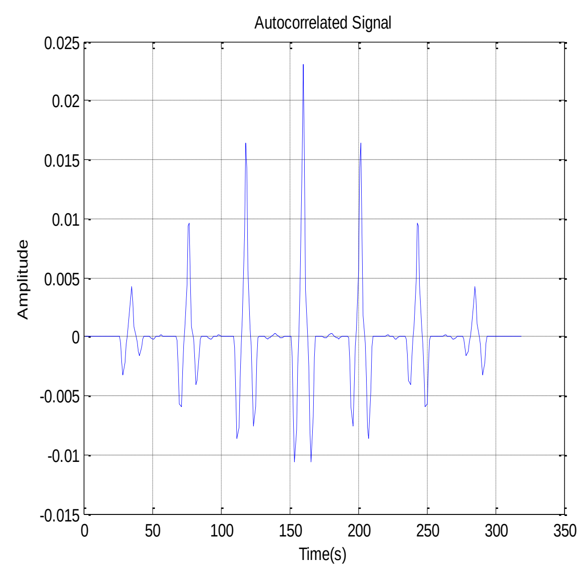

Figure6. Autocorrelated signal

4.2 On-The-Fly Method

According to the implementation discussed in part 2.2.1 the plot for the fundamental frequency or pitch of the signal is shown in Figure7.

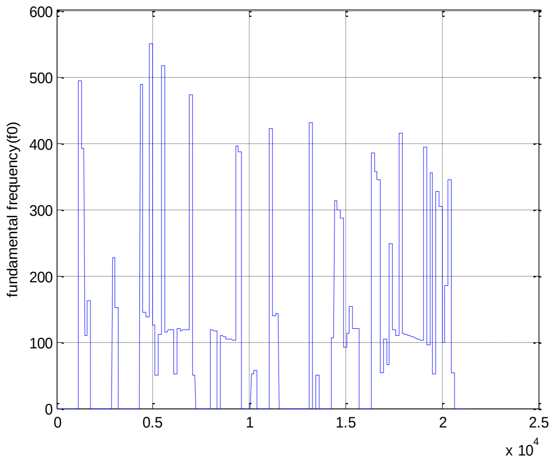

Figure7. Fundamental frequency using On-The-Fly Algorithm

4.3 Voiced/Unvoiced Decision

4.3.1 Zero Crossing Rate

Zero crossing rate function is computed using equation (8). As per the previous discussion the zero crossing rates is low for voiced part and high for unvoiced which is clearly shown in Figure8.

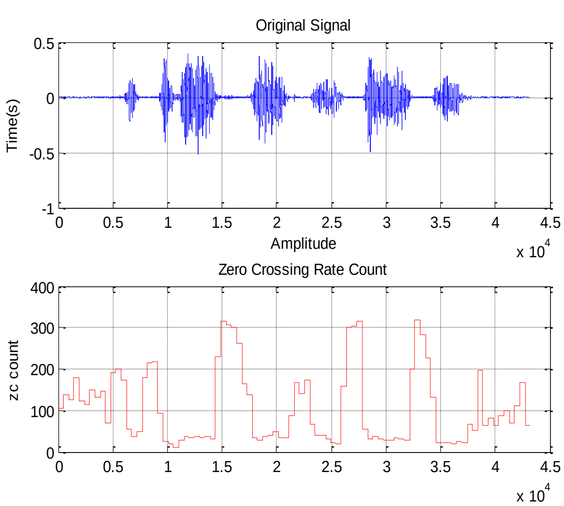

Figure8. Zero crossing rate count

4.3.2 Magnitude Sum Function

Butterworth filter is used for calculating the magnitude sum function. In this design, the magnitude sum function will be high for voiced part and low for unvoiced part. This can be shown in Figure9.

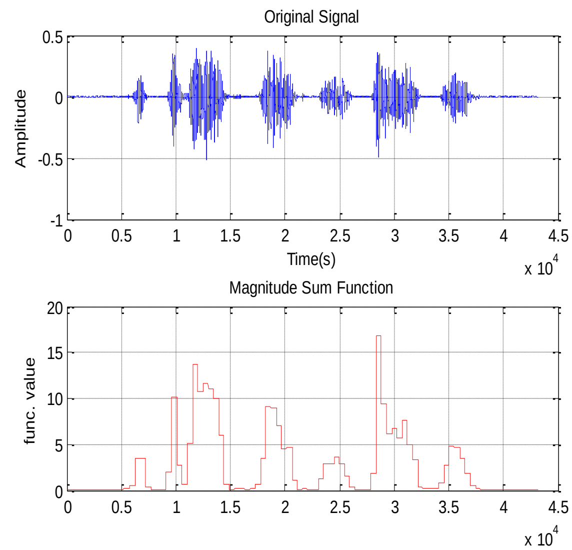

Figure9. Plot of Magnitude Sum Function

4.3.3 Voiced/Unvoiced Decision

A) Using Modified Autocorrelation Pitch Detecting Technique

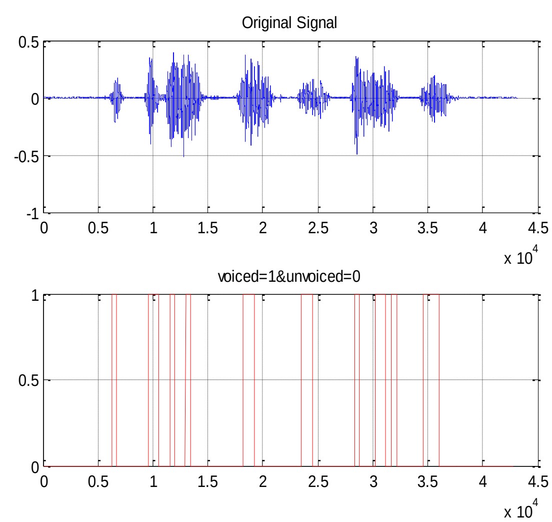

Figure10. Detecting voiced/unvoiced part of speech using modified autocorrelation method

In Figure10 a plot is divided into two subplots. The above plot shows the original signal and the lower part shows the voiced and unvoiced part of speech shown by 1 and 0. The autocorrelation technique could not detect the lower frequency voiced signal. To overcome this problem another pitch detection algorithm is used for detecting the voiced/unvoiced part of speech.

B) Using On-The-Fly Algorithm

On-The-Fly algorithm is based on difference function and reduced the problem of detecting lower frequencies voiced part of speech. The voiced/unvoiced region is shown by 1 or 0 in the Figure11.

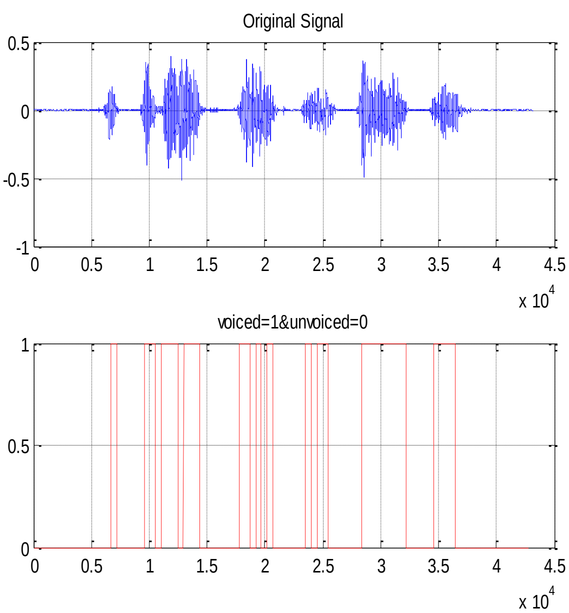

Figure11. Detecting voiced/unvoiced part of speech using On-The-Fly algorithm

5. CONCLUSION

This paper is focused on voiced/unvoiced decision. The detection of unvoiced/unvoiced part of speech suffers from the problem of detecting low voiced and unvoiced part of speech. This problem originates from the pitch detection of speech signal. Therefore, two pitch detecting algorithms i.e. modified autocorrelation and On-The-Fly algorithms are used and a comparison is done in this paper. When these two algorithms are used separately for detecting voiced/unvoiced part of speech the On-The-Fly algorithm reduces the above problem. This is shown in the results through suitable plot. In our future study, we can use better pitch detection technique than On-The-Fly algorithm.

6. REFERENCES

[1] Saurabh Sood, Ashok Krishnamurthy,'' A Robust On-The-Fly Pitch (OTFP) Estimation Algorithm", ACM Multimedia, 2004.

[2] Alain de Cheveigne, Hideki Kawahara," Yin, A Fundamental Frequency Estimator for Speech and Music, Acoustical Society of America, April 2002.

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

[4] Lawrence R. Rabiner, Michael j. cheng, Aaron E. Rosenberg, and Carol A. Mcgonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms", IEEE Transactions on Acoustics, Speech, and Signal processing, vol. No. 5, October 1976.

[5] John J. Dubnowski, Ronald W. Schafer, and Lawrence R. Rabiner, "Real-Time Digital Hardware Pitch Detector", IEEE Transactions on Acoustics, Speech, and Signal processing, vol. No. 5, October 1976.

[6] Tamas Bardi," Speech F0 estimation with enhanced voiced-unvoiced classification", Hiradastechnika, LIX. Volume 2004/12.

[7] G. Gandhimathi and S. Jayakumar," Efficient Method of Pitch Estimation for Speech Signal Using MATLAB", Volume- 3, Issue-1.

[8] Li Tan and Montri Karnjanadecha," Pitch Detection Algorithm: Autocorrelation Method and AMDF", ISCIT, 2003.

[9] Fredrik Bajers Ve, "Bandwidth Expansion of Narrowband Speech using Linear Prediction", Aalborg University, 2004.

[10] Leonard Janer, Juan Jose Bonet and Eduardo Lleida-Solano, "Pitch Detection and Voiced/Unvoiced Decision Algorithm based on Wavelet Transform", Spanish Ministry of Education and Sciences.

[11] M. M. Sondhi, "New methods of pitch extraction," IEEE Trans. Audio Electroacoustics (Special Issue on Speech Communication and Processing—Part Ii) vol. AU-16, pp. 262—266, June 1968.

[12] John R. Deller, John H.L. Hansen, John G.Proakis," Discrete-Time Processing of Speech Signal", IEEE edition, 2000.

[13] Fabien Gouyon, François Pachet, and Olivier Delerue," On the Use of Zero-Crossing Rate for an Application of Classification of Percussive Sounds", Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.

[14] David Gerhard," Pitch Extraction and Fundamental Frequency: History and Current Techniques", University of Regina, November 2003.