AUTOCORRELATION ANALYSIS FOR PITCH TRACKING

By

DHINESH.R

Reg. No. 1020106004

of

KUMARAGURU COLLEGE OF TECHNOLOGY

(An Autonomous Institution affiliated to Anna University, Coimbatore)

COIMBATORE - 641049

A MINI PROJECT REPORT

Submitted to the

FACULTY OF ELECTRONICS AND COMMUNICATION ENGINEERING

In partial fulfillment of the requirements
for the award of the degree

of

MASTER OF ENGINEERING

IN

APPLIED ELECTRONICS

MAY 2011

BONAFIDE CERTIFICATE

Certified that this project report entitled "**AUTOCORRELATION ANALYSIS FOR PITCH TRACKING**" is the bonafide work of **DHINESH.R** [Reg. no. 1020106004] who carried out the mini project under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Project Guide

Dr. Rajeswari Mariappan .

Head of the Department

Dr.Rajeswari Mariappan.

The candidate with university Register no. 1020106004 is examined by us in the mini project viva-voce examination held on ...23.05.11........

Internal Examiner

23/5

External Examiner

ACKNOWLEDGEMENT

I express my profound gratitude to our chairman **Padmabhusan Arutselvar Dr.N.Mahalingam B.Sc.,F.I.E.** for giving this opportunity to pursue this course.

At this pleasing moment of having successfully completed the project work, I wish to acknowledge my sincere gratitude and heartfelt thanks to our beloved principal **Dr.S.Ramachandran Ph.D.,** for having given me the adequate support and opportunity for completing this project work successfully.

I express my sincere thanks to **Dr.Rajeswari Mariappan Ph.D.,** the ever active, Head of the Department of Electronics and Communication Engineering, who rendering us all the time by helping throughout this project and also I extend my heartfelt thanks for her ideas and suggestion, which have been very helpful for the completion of this project work.

In particular, I wish to thank and everlasting gratitude to the project coordinator **Ms.R.Hemalatha M.E.,** Assistant professor,(SRG), Department of Electronics and Communication Engineering for her expert counseling and guidance to make this project to a great deal of success. With her careful supervision and ensured me in the attaining perfection of work.

Last, but not the least, I would like to express my gratitude to my family members, friends and to all my staff members of Electronics and Communication Engineering department for their encouragement and support throughout the course of this project.

ABSTRACT

One of the most time honored methods of detecting pitch is to use autocorrelation analysis on speech which has been appropriately preprocessed. The goal of the speech preprocessing in most systems is to whiten, or spectrally flatten, the signal so as to eliminate the effects of the vocal tract spectrum on the detailed shape of the resulting autocorrelation function. The purpose of this project is to spectrally flatten the speech signal. By appropriate adjustment of the threshold levels, the center clipping and peak clipping autocorrelation can be obtained

Autocorrelation to the speech segment is computed and then, the maximum of the autocorrelation is normalized to one. Extracting the positive part of the correlation and finding the maximum peak at zero delay. After finding maximum peak the index of the samples are converted to seconds. If the maximum peak exceeds threshold of the autocorrelation value at zero delay, then section is classified as voiced and the location of the maximum is the pitch period. Otherwise, the section is classified as unvoiced.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Pitch is a fundamental property of voiced speech. For voiced speech, the glottis opens and closes in a periodic fashion, imparting a periodic character to the excitation. The pitch period, T_0, is the time span between sequential openings of the glottis. The pitch frequency, F_0, is the reciprocal of the pitch period ($F_0 = 1/T_0$). The range of fundamental frequencies for human speakers is 50 to 500Hz. Pitch period estimates from the acoustic waveform can vary, because the voiced excitation of the vocal tract is only quasi-periodic.

1.1 PROJECT GOAL

The goal of the project is pitch tracking techniques using autocorrelation method, involving pre-processing and the extraction of pitch pattern. After pre-processing, DC content of speech segment is removed by subtracting from the mean value. Then finding maximum and minimum samples of the speech segment and centre clip to 75% to the samples.

1.2 OVERVIEW

In low bit rate speech coders, pitch is usually transmitted once per frame and, when needed, the intermediate pitch values are obtained by interpolation between two adjacent pitch values. Although pitch usually evolves slowly, sometimes it has nonlinear variation and the estimated pitch differs from the real one. This affects the synthesised speech quality. We propose to use a preprocessor, which modifies the residual speech signal such that the pitch period evolves more smoothly without distorting perceptual speech quality. Thus, the pitch and the voicing level can be determined correctly.

1.2 SOFTWARE USED

- MATLAB 7.8

1.4 ORGANIZATION OF THE REPORT

- **Chapter 2** discusses about the importance of pitch estimation.
- **Chapter 3** discusses about preprocessing technique
- **Chapter 4** discusses the autocorrelation method.
- **Chapter 5** discusses about implementation.
- **Chapter 6** discusses the simulation results.
- **Chapter 7** shows the conclusion of the project.

CHAPTER 2
IMPORTANCE OF PITCH ESTIMATION

Motivation for speech coding is to reduce the cost of operation of voice communication that involves development of various efficient coding algorithms and relating areas. One of the important areas of speech coding is pitch estimation. There is a significant number of speech coding algorithms, which are broadly classified into four categories, namely,phonetics, waveform, hybrid and voice vocoders.

Phonetics vocoders are more related to the acoustic characteristics of speech signals, whose investigation is beyond the scope of this thesis. Second form coders arewaveform coders, which are based on a simple sampling and amplitude quantization process.only concept behind these coder types is amplitude quantization, the compression rate of speech signals is limited to very large numbers. Even the most recently standardized waveform coders require a minimum of 16 kbits/sec.

The main objective of the current speech coders isto reduce the minimum compression rate to 1 - 4 kbits/sec or even lower. With the increasing demand for further compression (low bit rate coding), and increasing number of different 3 applications, simple amplitude quantization is not an efficient process for transmission of speech signals.

In contrast to waveform coders, vocoders consider the details in the nature of human speech. In their principles, there is no attempt to match the exact shape of the signal waveform.Vocoders generally consist of an analyzer and synthesizer. The analyzer attempts to estimate and then transmit the model parameters that represent the original signal. Speech is synthesized using these parameters to produce an often crude and synthetic constructed speech signal. These types of algorithms are called perceptual quality coders. In this type of coders, speech signals are synthesized with an excitation that consists of a periodic pulse train or white noise. The complete quality of the synthesized speech signal depends on the excitation signal.

The basis for estimating these parameters is the fundamental pitch period. This in turn causes an incorrect selection of excitation, therefore the final speech quality.From the above discussion, it is evident that the fundamental pitch period estimation is the deciding factor in the final quality of speech signal. In general, whether speech quality, communication quality, professional quality or synthetic quality, it all depends on the correct estimation of fundamental pitch.

In various speech coding algorithms, pitch estimation is an important parameter in the final quality of synthesized speech signal. For example, the effect of selecting multiples of correct pitch period, when modelling voiced speech, the spacing between the harmonics (fundamental frequency) is given by $2\pi/\tau$, where 'τ' is time period. Instead the second multiple of the pitch period is selected as a fundamental pitch, then the frequency spacing of the harmonics will be $2\pi / 2\tau$, i.e., the spectrum will contain twice as many harmonics and it produces very rough output speech in the vocoder. On the other hand, if the second sub-multiple is selected as correct pitch, then the fundamental frequency will be $2\pi / (\tau / 2)$. In this case speech sounds thin, e.g., male voice will sound similar to female voice.

The glottal excitation waveform is not a perfect train of periodic pulses. Although finding the period of a perfectly period of periodic waveform is straightforward, measuring the period of a speech waveform, which varies both in period and in the detailed structure of the waveform within a period, can be quite difficult.

In some instances, the formants of the vocal track can alter significantly the structure of the glottal waveform so that the actual pitch period is difficult to detect. Such interactions generally are most deleterious to pitch detection during rapid movements of the articulators when the formants are also changing rapidly..Another difficulty in pitch detection is distinguishing between unvoiced speech and low level voiced speech. In many cases, transitions between unvoiced speech segments and low level voiced speech segments are very subtle and thus are extremely hard to pinpoint .In practical applications, the background ambient noise can also affect the performance of the pitch detector. This is especially serious in mobile communication environments where a high level of noise is present.

The rapid advancement in the efficiency of digital signal processors and digital signal processing techniques has stimulated the development of speech coding algorithms. These trends are likely to continue, and speech compression most certainly will remain an area of central importance as a key element in reducing the cost of operation of voice communication systems.

CHAPTER 3
PREPROCESSING TECHNIQUE

Pre-processing of speech signal is very crucial in the applications where silence or background noise is completely undesirable. Applications like speech and speaker recognition needs efficient feature extraction techniques from speech signal where most of the voiced part contains speech or speaker specific attributes. Silence removal is a well known technique adopted for many years for this and also for dimensionality reduction in speech that facilitates the system tovbe computationally more efficient. This type of classification of speech into voiced or silence/unvoiced sounds finds other applications mainly in fundamental frequency estimation, formant extraction or syllable marking and so on.

The fundamental frequency is an important parameter in the speech analysis and synthesis. It plays an eminent role in the speech production and perception. In application areas such as speech enhancement, analysis and prosody modeling, low-bit rate coding, and speaker recognition, a reliable pitch estimation is required.

The speech signal includes very rich harmonic components. The minimum F_0 is about 80 Hz and the maximum is about 500 Hz. Most of them are in the range of 100-200 Hz. Thus the signal may involve 30-40 harmonic components. And the F_0 component is often not the strongest one. Because the first formant usually is between 300-1000 Hz. That is, the 2-8 harmonic components usually stronger than fundamental component. The rich harmonic components let the pitch tracking become very complex. It usually has the harmonic errors and sub harmonic errors.

To improve the reliability some pre-processing of signal is necessary. Since, the range of F_0 is generally in the range of 50 Hz to 500 Hz, then the frequency components above 500 Hz is useless for pitch detection. Thus a low-pass filter with pass-band frequency above 500 Hz would be useful in improving the performance of pitch detection.

Also to reduce the effects of the formant structure on the detailed shape of the short-time autocorrelation function, the nonlinear processing is usually used in pitch tracking. One of the nonlinear technique is centre-clipping of speech .

$$y(n) = clc[x(n)] = \left\{ \begin{array}{ll} (x(n) - C_L), & x(n) \geq C_L \\ 0, & |x(n)| < C_L \\ (x(n) + C_L), & x(n) \leq -C_L \end{array} \right\}$$

Where C_L is the clipping threshold. Generally C_L is about 30% of the maximum magnitude of signal. In application the C_L should be as high as possible. To get the high C_L, we can catch the peak value of the first 1/3 and the last 1/3 of signal and use the less one to be the maximum magnitude. Then we set the 60-80% of this maximum magnitude to be C_L.

CHAPTER 4
AUTOCORRELATION METHOD

A discrete time signal *x (n),* defined for all *n,* the auto-correlation function is generally defined as:

$$R_X(m) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} x(n) \times x(n+m)$$

The autocorrelation function of a signal is basically a (noninvertible) transformation of the signal that is useful for displaying structure in the waveform. Thus, for pitch detection, if we assume *x(n)* is exactly periodic with period *P*, i.e., *x(n) = x(n + P)* for all *n*, then it is easily shown that

$$Rx(m) = Rx(\mathrm{m} + \mathrm{P})$$

i.e., the autocorrelation is also periodic with the same period . Conversely, periodicity in the autocorrelation function indicates periodicity in the signal.

For a non stationary signal, such as speech, the concept of a long-time autocorrelation measurement as given above is not really meaningful. Thus, it is reasonable to define a short-time autocorrelation function, which operates on short segments of the signal as:

$$R_x(m) = \frac{1}{N}\sum_{n=0}^{N-1} [x(n+l)W(n)]\,[x(n+l+m)W(n+m]$$

where *w(n)* is an appropriate window for analysis, *N* is the section length being analyzed, *N'* is the number of signal samples used in the computation of *R(m), Mo* is the number of autocorrelation points to be computed, and *l* is the index of the starting sample of the frame

$$N' = N - m$$

So that only the N samples in the analysis frame (i.e., *x (l), x (l+1)* . . . *x (l + N - 1))* are used in the autocorrelation computation. Values of 200 and 300 have generally been used for *Mo* and *N,* respectively; it is corresponding to a maximum pitch period of 20 ms (200 samples at a 10 kHz sampling rate) and a 30 ms analysis frame size.

$$E(\tau) = R(0) - R(\tau)$$

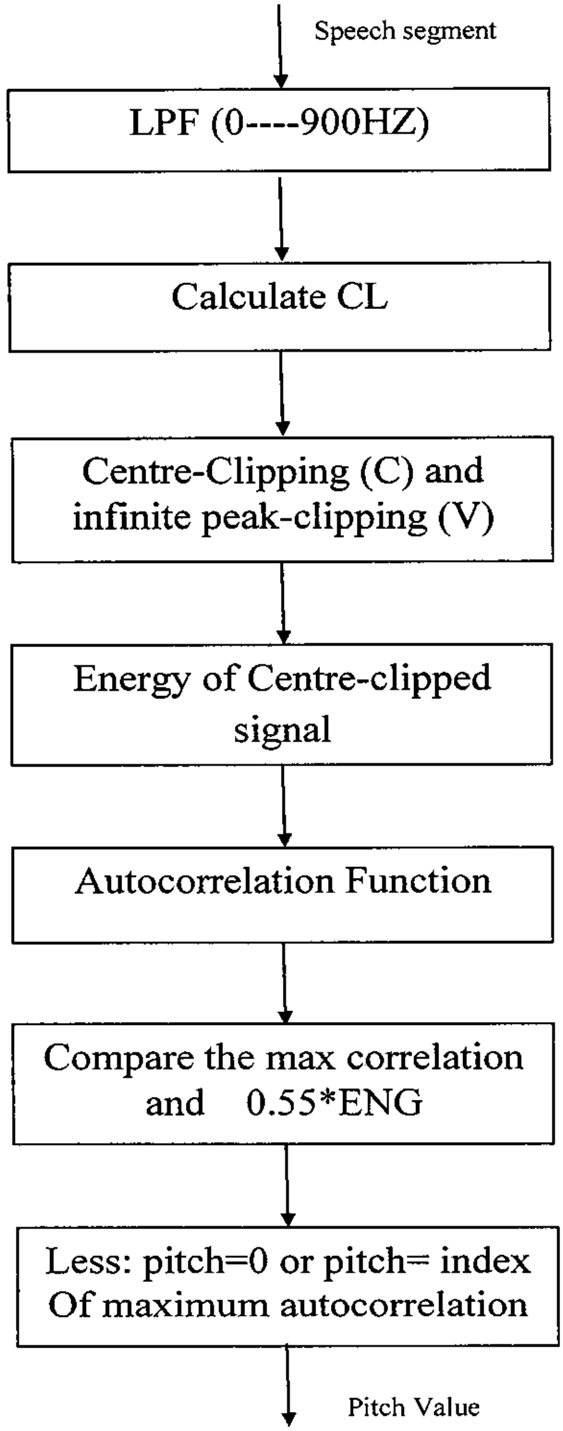
The minimization of the estimation error, E(τ), is equivalent to maximizing the auto-correlation R(τ). The variable τ is called lag or delay and the pitch is equal to the value of τ, which results in the maximum R (τ).

CHAPTER 5
IMPLEMENTATION

Autocorrelation pitch detector based on the centre-clipping method and infinite-clipping is used in our implementation. shows a block diagram of the pitch detection algorithm. The method requires that the speech be low-passed filtered to 900 Hz. The low-pass filtered speech signal is digitized at a 10-kHz sampling rate and sectioned into overlapping 30-ms (300 samples) sections for processing. Since the pitch period computation for all pitch detectors is performed 100 times/sec i.e., every 10 ms, adjacent sections overlap by 20 ms or 200 samples.

The first stage of processing is the computation of a clipping threshold C_L for the current 30-ms section of speech. The clipping level is set at a value which is 68 percent of the smaller of the peak absolute sample values in the first and last 10-ms portions of the section. Following the determination of the clipping level, the 30-ms section of speech is centre clipped, and then infinite peak clipped.

Following clipping the autocorrelation function for the30-ms section is computed over a range of lags from 20 samples to 160 samples (i.e., 2-ms-20-ms period). Additionally, the autocorrelation at zero delay is computed for voiced/unvoiced determination. The autocorrelation function is then searched for its maximum value. If the maximum exceeds 0.55 of the autocorrelation value at zero delay, the section is classified as voiced and the location of the maximum is the pitch period. Otherwise, the section is classified as unvoiced.


Speech segment
LPF (0----900HZ)
Calculate CL
Centre-Clipping (C) and infinite peak-clipping (V)
Energy of Centre-clipped signal
Autocorrelation Function
Compare the max correlation and 0.55*ENG
Less: pitch=0 or pitch= index Of maximum autocorrelation
Pitch Value

P- 3469

Figure 5.1 Implementation flow diagram

CHAPTER 6
RESULTS AND DISCUSSION

The simulation of this project has been done using MATLAB 7.8

MATLAB (matrix laboratory) is a numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.

MATLAB is a high performance interactive software package for scientific and engineering computation. It is an easy-to-use environment where problems and solutions are expressed just as they are written mathematically.

It is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages

Here original speech signal in figure 6.1 and 6.3 are pitch plotted using autocorrelation method and their results were shown in figure 6.2 and 6.4

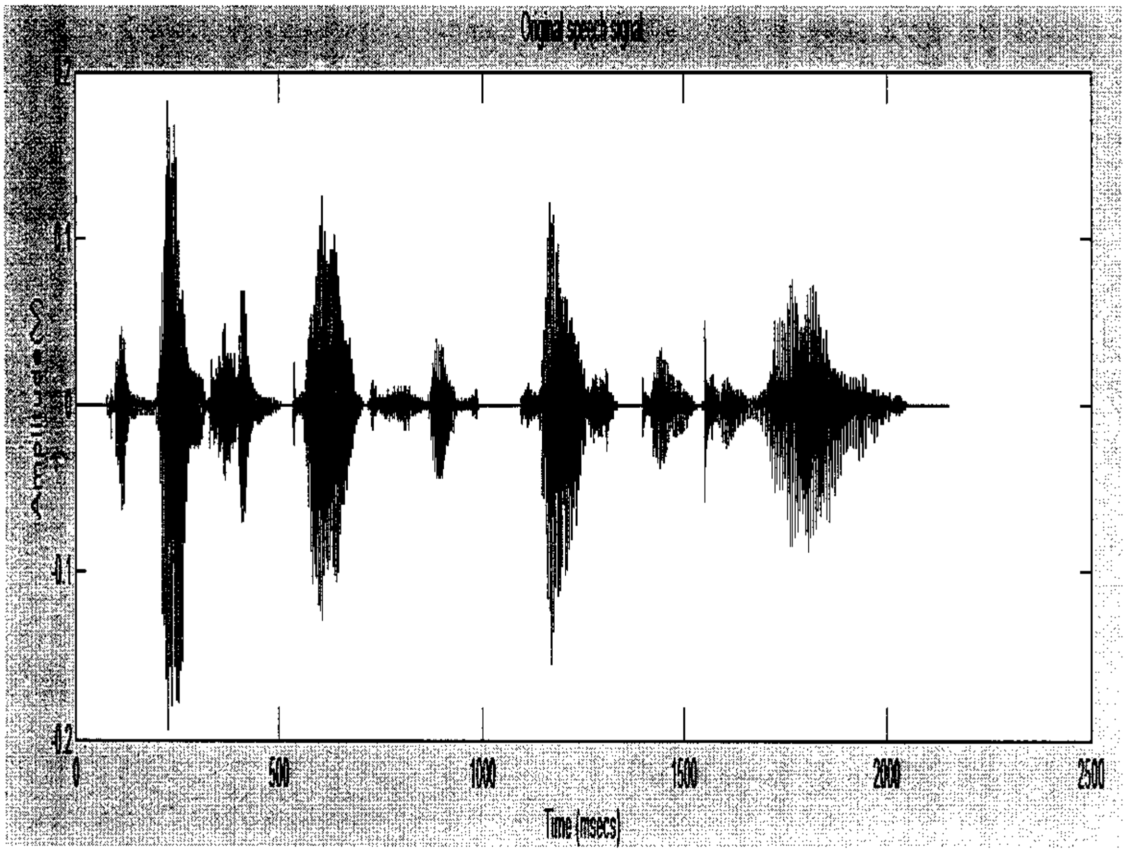


Figure 6.1 Original Speech Signal (i).

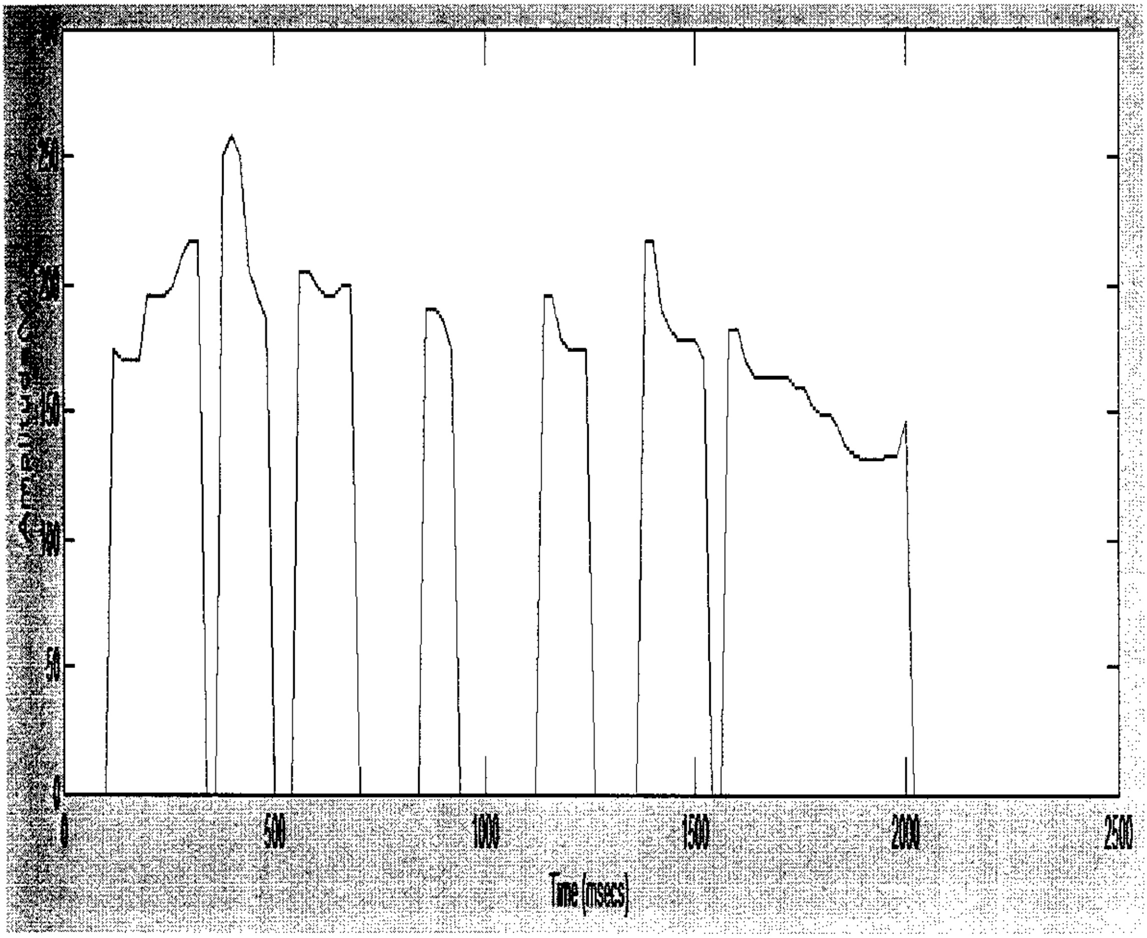


Figure 6.2 Simulation result of Pitch plot using autocorrelation method for Original Speech Signal (i).

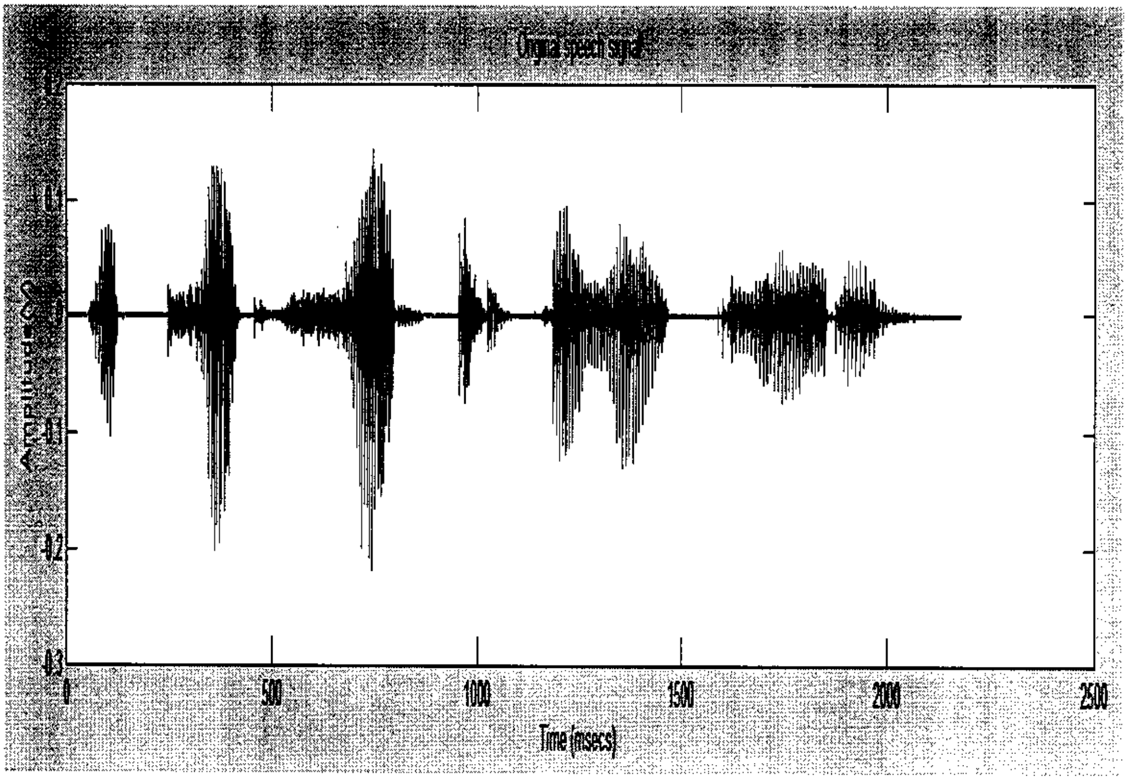


Figure 6.3 Original Speech Signal (ii)

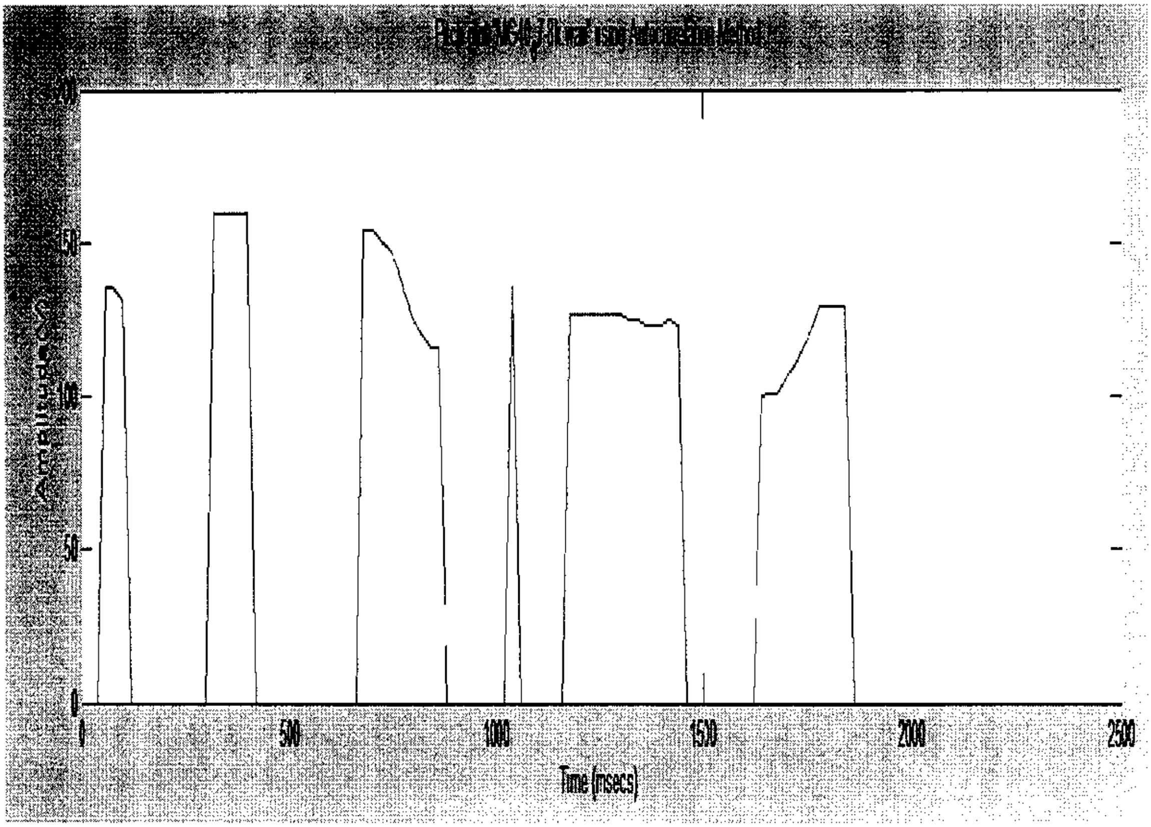


Figure 6.2 Simulation result of Pitch plot using autocorrelation method for Original Speech Signal (ii).

CHAPTER 7
CONCLUSION AND FUTURE SCOPE

In this project the pitch detection algorithm using autocorrelation method, techniques including pre-processing and extraction of pitch pattern were shown. Also the nonlinearities provide some degree of spectral flattening, thereby enhancing the périodicity peaks in the correlation function, and reducing the correlation peaks due to the formant structure of the waveform were described.

This method should be appropriate for any frame-by-frame speech analysis system in which pitch is extracted.

FUTURE SCOPE

For pitch detection the voice/unvoiced determination and the segmenting of pitch contour are the future works.

BIBLIOGRAPHY

[1] Yang Fan ,Liu Ming,Xu Sun;Pan Guo-Feng. Hebei Univ of Technology,China "Method of Preprocessing and speech Detection" Pages V1 399 – 401 ,Vol:1 AUG 2010.

[2] Lawrence R.Rabiner ,Member, IEEE" On the Use of Autocorrelation Analysis for Pitch Detection" IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING"VOL ASSP 25,NO 1,FEB 77

[3] Goldberg, Randy G. "A practical handbook of speech coders", CRC Press LLC.

[4] Milan Jelínek, Member, IEEE, and Redwan Salami," IEEE Wideband Speech Coding Advances in VMR-WB Standard," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 4, MAY 2007

[5] B. Bessette, R. Salami, R. Lefebvre, and M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, K. Järvinen, " The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)," IEEE Trans. On Speech and Audio Processing

[6] 3GPP TS 26.193 "AMR Wideband speech codec; Source Controlled Rate operation," 3GPP Technical Specification.

[7] 3GPP TS 26.190 "Adaptive Multi-Rate wideband speech transcoding," 3GPP Technical Specification

[8] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, J.-P. Adoul, "GSM Enhanced Full Rate Codec," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Munich, Germany, 20–24 April 1997, pp. 771– 774

[9] P. C. Loizou, Speech Enhancement, Theory and Practice , CRC Press, 2007.

[10]M. M. Sondhi, "New methods of pitch extraction," IEEE Trans.Audio Electroacoust., vol. AU-16, pp. 262-266, June 1968.