Simplistic Implementation of Realtime Pitch Estimation using Pulse Modelling Approach

Jason Fernandes
Department of Electronics & Telecommunication
Don Bosco Institute of Technology
Mumbai, India.

Kelvin D'souza
Department of Electronics & Telecommunication
Don Bosco Institute of Technology
Mumbai, India.

ABSTRACT
In the realm of speech processing, pitch estimation is considered to be a very important tool for various applications in speech processing, thus many algorithms and methods have already been developed for pitch estimation. This paper presents a new and relatively simple method for estimating the pitch of a speech signal as compared to other algorithms and methods which have been applied. In this method of pitch estimation, the speech signal is filtered and then passed as a trigger to a switch whose one input is a pulse generator and the other is a reference potential. The number of pulses generated will be proportional to the frequency of the speech signal.

Keywords
Pitch estimation, Pulse Converter, Real-time Processing

1. INTRODUCTION
In voiced speech, pitch can be defined as the vibration of the vocal folds. Pitch Estimation is a very essential tool in various applications such as speaker identification and verification, speech coding, speech synthesis and speech instructions to the hearing impaired. Pitch Estimation of speech signals can be described as the accurate estimation of the fundamental frequency of the signal. However, estimation of the pitch of a speech signal is challenging task due to the quasi-periodic nature of speech and the mixed nature of excitation. There are numerous methods of pitch estimation which have been developed by a number of researchers, and these methods are mainly categorized in terms of functional domain viz. time, frequency and cepstrum methods.
For the time domain estimation of pitch, the methods used operate directly on the given speech waveform. In these pitch estimation methods the algorithm used are generally peak and valley measurements, zero-crossing and auto-correlation. For Since the speech signal is a quasi-periodic waveform, all these methods assume that the quasi-periodic nature of the signal is suitably processed to minimize the effects in the formant structure and hence simple time domain measurements will provide good estimates of the period.[6].
The most commonly used time-domain pitch estimation methods include: threshold-crossing analysis methods, parallel processing method, envelope analysis, auto-correlation and Average Magnitude Difference Function (AMDF) methods.
The existing methods for pitch estimation lack robustness, and moreover some of the methods are quite complicated. This paper presents a straight-forward way of finding the pitch of the given speech signal in the time-domain.

2. PROPOSED SYSTEM
The proposed system is given in block diagram in Fig. 1.1. The input is a speech signal whose pitch is required to be estimated. The blocks are explained as follows.

2.1 Filter
The filter used is a low pass filter kept at a stop-band frequency of 1 KHz. The filter is used to filter out the highfrequency signals and noise that would interfere with the speech signal.

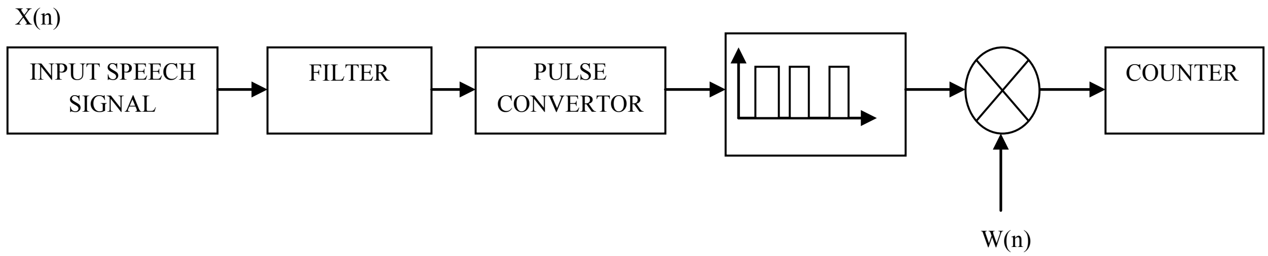

Fig.1.1:Block Diagram of the Proposed System

2.2 Pulse Converter

The block diagram of the pulse converter is shown in Fig.1.2. The speech signalwhich is passed through the low pass filter is the threshold for the switch.

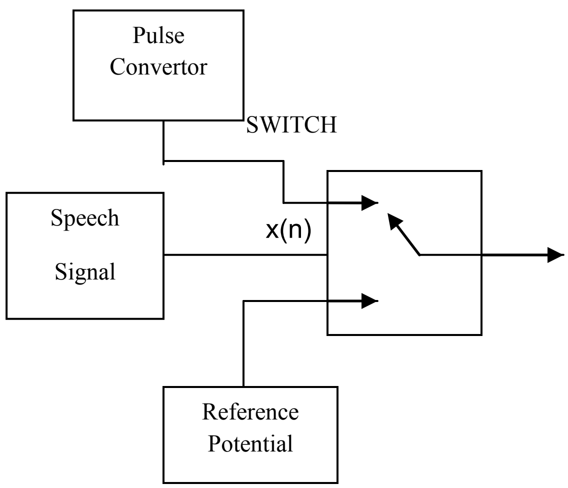

Fig.1.2:Pulse Converter

When the switch detects a threshold, it changes its value between the pulse generator and the ground potential. The number of pulses that are passed through the switch given by x(n), is directly proportional to the frequency of the speech signal. If the frequency of the speech signal is low the number of switching will be less and hence the pulses allowed to be passed will be more.

Accordingly, as the frequency of the signal increases the switching time increases hence the number of pulses passed will be less. The switching time period of the pulses generated is kept to a value of 0.001 secondswhich helpeliminate the possibility of any unwanted high frequency signals to be sampled by the system hence ensuring the formant frequencies only are sampled by the system while the higher frequencies are not converted to the required pulse train.

The switching action is also controlled by a threshold within the switch. The threshold is 0.03 of the normalized amplitude which will eliminate the pickup of low energy signals as well as noise and other interference. This threshold can be varied accordingly to get the pitch of speech signals at lower energy. This helps in detecting pitch of a speech signal which is attenuated.

2.3 Window Function

A frame of the pulse stream is extracted by multiplying it with a Rectangular window, given by W(n) of a 0.05 ms time period which allows the estimation of the pitch over a fixed time duration so the short time pitch of the signal can be extracted effectively.

2.4 Counter

The counter block is an up counter used to count the pulses within the specified window length and at the end of thewindow time period it will reset the counter and give a readout, which can be plotted to show the pitch of the waveform in that time instant.

The pitch can be calculated by the formula shown below:

$$Pitch = \frac{\frac{Number\ of\ Pulses}{Counted} \times Time\ Period\ of\ Window}{Time\ Period\ of\ Pulse}$$

3. RESULTS

The proposed system was simulated to test its efficiency and the results were compared to the pitch estimation method using autocorrelation.

To get an indicative of the pitch the total count within each window period is plotted in a graph as a function of time. The sample used here is the voiced waveform|ah| as spoken by a male.

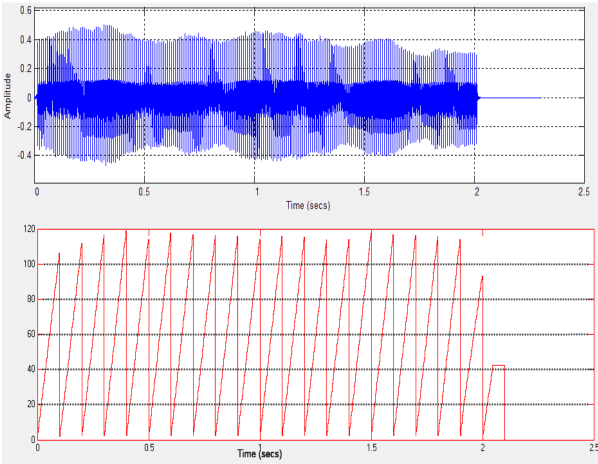

Fig. 2.1

Fig2.1 shows the time scale plot of the waveform as the sample is played out. To make the reading more accurate the peaks can be averaged over fixed time duration so as to achieve the readout of the frequency.

Since the sampling times are so chosen that high frequency noise will not be accepted by the system also as the threshold levels which can be appropriately set such that the system will not respond to any random pickup over the microphone.

The example of the same voiced speech sample with a white noise source as well as minor disturbance is shown below in Fig. 2.2.

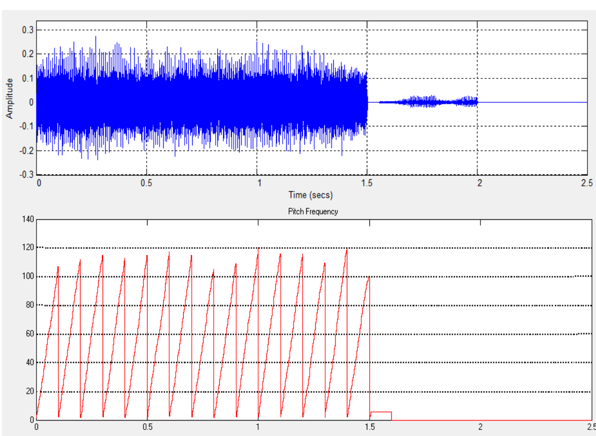

Fig. 2.2

As seen from the graphs the system is capable of estimating the pitch even in very noisy environments and at the same time it ignores any sort of minor sounds and successfully detects the silence in the waveform as shown in the Fig. 2.2. The signal picked up after the time of 1.5 seconds in Fig. 2.2 is random noise which has not been detected in the pitch estimator.

Table 2.1

Input Waveform	Pitch With Autocorrelation	Pitch with proposed method
|ah|	120.3 Hz	124 Hz
|ah| with white noise	131 Hz	121 Hz

Table 2.1 shows the results of autocorrelation and the proposed methodology be very similar highlighting the accuracy of the system and its ability to perform even in a noisy environment. The biggest advantage is that the pitch is being estimated on a real-time basis in the proposed system.

4. CONCLUSION

The proposed system is highly accurate and also reduces the computational load on the processorby decreases the number of calculations required to achieve a pitch. The implementation of this algorithm will allow for a real-time detection of the pitch of waveform which will enhance the development of learning aids for the hearing impaired by givingreal-time feedback. The development of speaker recognition systems with much higher efficiency at lower computational loads can also be achieved.

5. REFERENCES

[1] Dominik Joho, MarenBennewitz, and Sven Behnke"Pitch Estimation Using Models of voiced speech on three levels", Proceedings of 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[2] MarekSzczerba&AndrzejCzyżewski "Pitch Estimation Enhancement Employing Neural Network- Based Music Prediction", Sound & VisionEngineeringDepartmentTechnicalUniversity of Gdańskul.Narutowicza.

[3] Sergio Roa, MarenBennewitz, and Sven Behnke "Fundamental Frequency Estimation based on Pitch-Scaled Harmonic Filtering", University of Freiburg Department of Computer Science.

[4] Hess W.,"PitchDetermination of Speech signals: Algorithms and Devices", Springer Verlag, Berlin Heidelberg, New York, Tokyo, 1983.

[5] Dushyant Sharma, Patrick. A. Naylor "Evaluation of Pitch Estimation in Noisy Speech for Application in Non-Intrusive Speech Quality Assessment" Department of Electrical and Electronic Engineering, ImperialCollege, London, UK

[6] Lawrence R. Rabiner, Michael J. Cheng, Aaron E. Rosenberg, Carol A. McGonegal "AComparative Performance Study of Several PitchDetection Algorithms" IEEE Transactions on Acoustics Speech and Signal Processing, VOL. ASSP-24,NO. 5, OCTOBER 1976

[7] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-22, pp. 353—362, Oct. 1974.

[8] Yipeng Li and DeLiang Wang “Pitch Detection in Polyphonic Music Using Instrument Tone Models” Department of Computer Science and Engineering&Center for Cognitive ScienceThe OhioStateUniversity.

[**9**] M. Wu, D. L.Wang, and G. Brown, “A Multipitch Tracking Algorithm for Noisy Speech,” IEEE Transactions on Speech andAudio Processing, vol. 11, no. 3,pp. 229–241, 2003.

[10] J. D. Market, "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio Electroacoust., vol. AU-20, pp. 367—377, Dec. 1972.

[11] J. D. Markel and A. H. Gray, Linear Prediction of Speech. NewYork: Springer, 1976.