Pitch Estimation and Analysis of speech signal

Mahesh Manohar Kamble, Prof. (Mrs.) M.R Dixit

Abstract- Speech is the principal form of human communication since it began from day one when human beings start to communicate. The rate of vibration produce by the vocal cords is called a fundamental frequency (F0) or pitch period. Consequently, the pitch period estimation is to determinate the fundamental frequency for used in speech signal processing applications. The fundamental frequency range for a person is about 20 to 20 kHz, and the frequency of a sound wave will determine the human tone and pitch. The resultant of spikes in the correlation of voice data is to determine the period and therefore the pitch of the signal. Numerous pitch determination algorithms (PDAs) have been proposed in the literature. In general, they can be categorized into three classes: Time-domain, frequency-domain, and time–frequency domain Algorithms. The pitch tracking techniques using autocorrelation method and AMDF (Average Magnitude Difference Function) method involving the preprocessing and the extraction of pitch pattern.

Keywords: **Pitch, Pitch Detection Algorithm, Autocorrelation function, Speech Recognition System, Center-clipping.**

I. INTRODUCTION

Pitch detection is very important for many speech processing algorithm. Speech recognition system of tonal language use pitch tracking for tone recognition, which is important in disambiguating the myriad of homophones. Pitch is also crucial for prosodic variations in text-to-speech systems and spoken language systems. The fundamental frequency (*F0*) is the main cue of the pitch. However, it is difficult to build a reliable statistical models involving fundamental frequency F0 because of pitch estimation errors and the discontinuity ofthe F0 space. Thus, a reliable pitch detection algorithm (PDA) is a very important component in many speech processing systems.

II. BACKGROUND

A. Autocorrelation Method and AMDF

Basically, pitch detection algorithms use short-term analysis techniques. For every frame x_m we get a score f(T| x_m) that is a function of the candidate pitch periods T. A commonly used method to estimate pitch is based on detecting the highest value of the autocorrelation function in the region of interest. Given a discrete time signal x(n),defined for all n , the auto-correlation function is generally defined in

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

A variation of autocorrelation analysis for measuring the periodicity of voiced speech uses the AMDF, defined by the relation in

$$D_m = \frac{1}{L} \sum_{n=1}^{L} | x(n) - x(n-m) |, \quad m = 0,1,..m_{\max}$$

Where x(n) are the samples of input speech and x(n-m), are the samples time shifted m seconds. The vertical bars denote taking the magnitude of the difference x(n) – x(n-m) Thus a difference signal Dm, is formed by delaying the input speech various amounts, subtracting the delayed waveform from the original, and summing the magnitude of the differences between sample values. The difference signal is always zero at delay = 0, and is particularly small at delays corresponding to the pitch period of a voiced sound having a quasiperiodic structure.

B. Preprocessing Technique

From above, we know the autocorrelation function and AMDF can be used to detect the pitch. However the speech signal include very rich harmonic components. The minimum F0 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 F0 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 preprocessing of signal is necessary. Since, the range of F0 is generally in the range of 80-500 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. Generally, we use the lowpass- filter with 900 Hz. 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.

$$Y(n)=C[x(n)]$$

One of the nonlinear technique is center-clipping of speech which is first introduced by *M. M. Sondhi* [3]. The relation between input *x(n)* and *y(n)* is:

Manuscript received November 18, 2013.

Mahesh Manohar Kamble, Department of Electronics & Telecomm Engineering, K.I.T.'s College of Engineering, Kolhapur, India. (e-mail: vickkyshree9@gmail.com).
Prof. (Mrs.) M.R Dixit, Department of Electronics & Telecomm Engineering, K.I.T.'s College of Engineering, Kolhapur, India.
(e-mail: mrdixit@rediffmail.com).

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

where *CL* is the clipping threshold. Generally *CL* is about 30% of the maximum magnitude of signal. In application the *CL* should be as high as possible. To get the high *CL*, 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 *CL*.

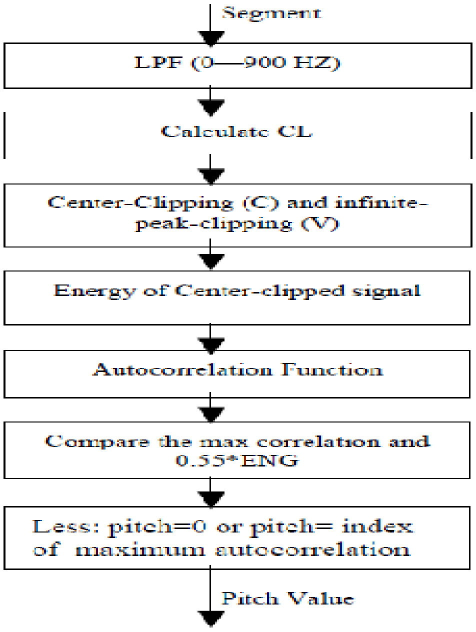
C. Post-processing

Generally, the pitch determination described above is still error-prone. The erroneous voiced/unvoiced decisions and inaccurate voiced pitch hypotheses can lead to noisy and undependable feature measurements. Then a smoothing stage is necessary in improving the performance of the system. The most common smoothing techniques includes: median filter, linear smoothing and dynamic programming technique. According to the reliability of pitch tracking algorithm, generally the median-filter is used. In the method of medianfilter, it uses a moving window with the length *L*. The value at point n is determined by the data from point *n-L* to point *n+L*. Then the median value in these *2L+1* points is chooses as the value the point.

III. IMPLEMENTATIONS

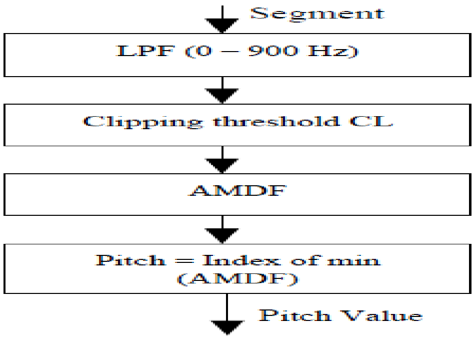
A. Modified Autocorrelation Method

According to the discussion above, the modified autocorrelation pitch detector based on the center-clipping method and infinite-clipping is used in our implementation. 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/s 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 *CL* 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 center 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).


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

B. AMDF

We only implement a coarse quantization. We leave the voice/unvoiced detection and the decision logic as the further work. Fig. 3 shows a block diagram of the AMDF pitch detector. The speech signal, is initially Then the signal pass a low-pass filter (0-900 Hz) and set the first 20 samples to be zero. The clipping threshold is then calculated and the center-clipping is done on the signal. Then average magnitude difference function is computed on the center-clipped speech signal at the lag (20—140 samples) through the signal from 20 to 160 samples. The pitch period is identified as the value of the lag which the minimum AMDF occurs. Thus a fairly coarse quantization is obtained for the pitch period.sampled at 10 kHz.



C.Correntropy

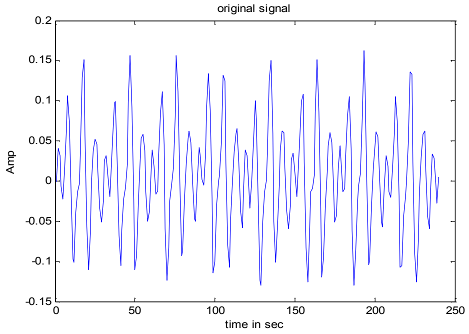
A new generalized similarity measure

Define correntropy of a stationary random process $\{x_t\}$ as

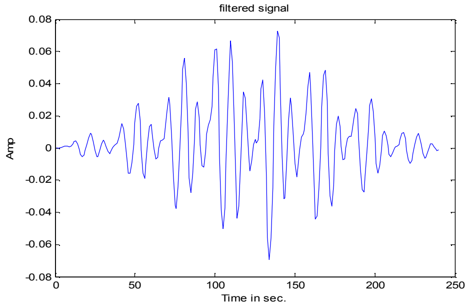
$$V_x(t,s) = E(\kappa(x_t - x_s)) = \int \kappa(x_t - x_s)p(x)dx$$

The name correntropy comes from the fact that the average over the lags (or the dimensions) is the information potential.

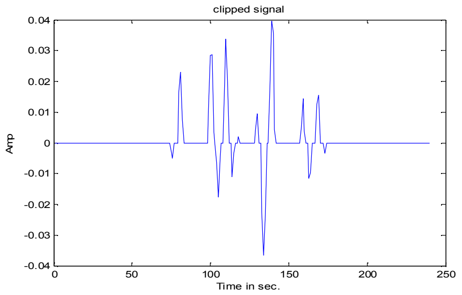
IV. EXPERIMENTS



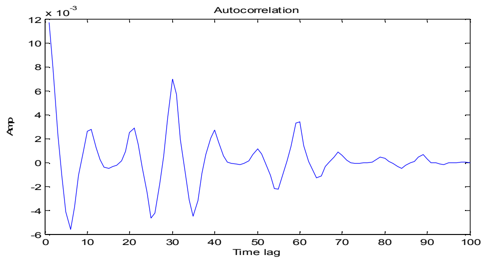
Input voice signal.



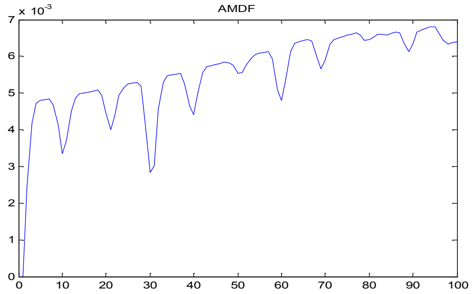
Filtered signal.



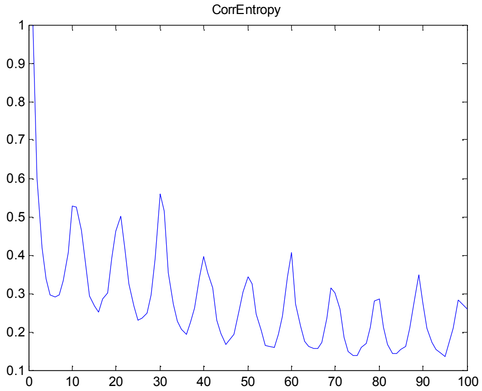
Clipped signal.



Autocorrelation signal.



AMDF Signal



Correntropy signal

V. CONCLUSIONS

The work that we described here are pitch detection algorithms and the related techniques including preprocessing post-processing and extraction of pitch pattern. According to our observing of the experiments. We found that both autocorrelation method, AMDF and correntropy algorithm can provide the accepted results.

REFERENCES

[1]. 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 Audio, Signal, and Speech Processing 24, 399-417 1976.

[2]. M. M. Sondhi, “New methods of pitch extraction,” IEEE Trans.Audio Electroacoust., vol. AU-16, pp. 262-266, June1968.

[3]. Yi Kechu, Tian Fu, Fu Qiang, “YU YIN XIN HAO CHULI”, China Machine Press, BeiJing, 2000.