Study of Pitch Detection Methods

Veena S[1] Nilashree S. Wankhede[2] Milind S. Shah[3]
[1,2,3]Department of Electronics and Telecommunication Communication
[1,2,3]FCRIT, Vashi

Abstract— Speech is produced by vibrations of the vocal cords and the rate of vibration of the cords is called pitch. Pitch is a feature of vocal speech which may contain speaker specific information. Pitch gets affected by various factors one of which is length of vocal folds. Various methods have been developed for pitch estimation which can be classified as time domain method, frequency domain method and hybrid method. Pitch estimate can be evaluated in time-domain by identifying periodicity features within the sound wave. The most commonly used time-domain methods include: parallel processing method, autocorrelation and Average Magnitude Difference Function (AMDF) methods. Frequency-domain pitch estimation can be evaluated by identifying certain features within the short-term spectra of speech signal. Hybrid Methods have features of both time and frequency domain. This paper describes some of the time domain and frequency domain methods for pitch estimation and its applications. In this paper autocorrelation method under time domain has been implemented using MATLAB software and the results obtained are verified using PRAAT software.

Key words: Pitch Detection,

I. INTRODUCTION

Speech is a principal form of communication. The process of speech production can be modeled as a time varying linear system. For voiced sounds the system is excited by quasi-periodic source and for unvoiced sounds a noise like source is used. During voiced speech production a vibration occurs in vocal folds [1]. Vocal folds work by using Bernoulli principle. According to Tanvia et al. [2] the movement of vocal fold occurs in three phases namely closing phase, opening phase and the return phase. The time taken to complete one such phase is known as pitch period and the inverse of pitch period is called fundamental frequency. Pitch is different for different speakers. This variation is mainly due to difference vocal folds. An adult male has pitch frequency in the range of 85-180Hz.In case of females it varies from 160-300Hz. For children the pitch value is higher than both male and female that is around 300-500Hz.

Pitch detection has applications in many areas that involve processing of sound, speech coding, speech recognition, lie detection system, efficient transmission and recognition of speech signals, speech pathology, identifying state of the speaker. [3]

Section 2 discusses pitch estimation techniques and its applications. Section 3 presents the implementation of autocorrelation method. Section 4 gives the results obtained by using MATLAB and the validation of these obtained results is done by using PRAAT. And Section 5 makes some of the conclusions drawn.

II. PITCH ESTIMATION

Pitch can be related to the rate of the vibration of the vocal fold in voiced speech production. Many factors influence the pitch. Pitch depends on the length of vocal folds. Tanvina et al. [2] have observed that vocal fold length and pitch are inversely related to each other. For male vocal fold length is in the range of 17-25mm, for females it is 12.5-17.5 mm and for infants it is around 6-8mm.

$$F_0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}} \qquad (1)$$

'σ' denotes longitudinal stress.
'ρ' is tissue density.
'L' is length of vocal fold.

Tension of vocal folds also affects pitch. An increase in vocal cord tension increases pitch and decrease in vocal tension decreases pitch. Other factor by which pitch gets affected is the emotional state of a person. From a study carried out by Baghshree et al. [4] on eight different speakers it was observed that the pitch value differs for the same person while having different emotional states like neutral and anger. Pitch frequency was less for neutral speech than angry speech [5].

A. Pitch Estimation Techniques

The pitch Estimation Techniques can be classified as

1) Time domain Method

In this method pitch detectors operate directly on the speech waveform to estimate the pitch period .There are various techniques used in this method. A basic assumption made in case of all time domain methods is that the quasi periodic speech signal is to be processed to remove formant effects. And then any of the suitable methods are applied to obtain pitch values. One among

them is Parallel processing approach proposed by Gold and Rabiner. In this case the speech signal is processed for creation of impulse trains which discards irrelevant features of pitch detection. Other commonly used methods are Autocorrelation and AMDF method. Autocorrelation method is implemented in detail in this paper [5].

a) Autocorrelation Method

The most widely used method for detecting pitch is based on finding the autocorrelation of interested region. According to Rabiner [12] the autocorrelation function is transformation of the signal which is useful for displaying structure in the waveform. In case of discrete time signal x(n) is defined for all n as

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

'N' represents length of analyzed sequence.

A. Average Magnitude difference Function(AMDF Method) [6]

AMDF method is known to be the variation of autocorrelation method. A difference signal is obtained between the delayed speech and original rather than correlating them as in case of ACF.

$$D_r = \frac{1}{L} \Sigma_{j=1}^{L} \left| S_j - S_{j-\tau} \right| \qquad (3)$$

S_j are samples of input speech

$(S_j) = (S_1, S_2, \ldots ..S_L)$

In such extractors, separation of fine structure from spectral envelope is a challenge which could be overcome by using prior knowledge of voicing and thereby making pitch decision more reliable.

B. Frequency domain Method

According to Rabiner *et al.* in this type of detectors a series of impulses are present in the frequency spectrum of periodic signal at fundamental frequency (f_0) and at its harmonics $2f_0$, $3f_0$. Hence the period of signal could be found by using algorithms like Harmonic product spectrum and Cepstrum method [5].

a) Harmonic Product Spectrum

The multiplication of original spectrum is carried with the compressed spectra and resulting peaks at fundamental frequency gets magnified. Further measurement of the maximum coincidence of harmonics for each spectral frame is carried out.

$$Y(e^{j\omega}) = \prod_{r=1}^{R} X\left|\left(e^{j\omega r}\right)\right|$$

$$p = \arg\max \left[Y(\omega)\right] \qquad (4)$$

Maximum value of periodic correlation array is obtained and the frequency corresponding to it becomes pitch frequency [7].

b) Cepstrum Method

The word Cepstrum is formed by swapping spectrum. Cepstum is defined as the inverse DFT of the log magnitude of the DFT of a signal

$$C[n] = F^{-1}\{\log |F[x(n)]|\} \qquad (5)$$

'F' is the DFT and

'F^{-1}' is IDFT.

The block diagram of cepstral pitch detector is shown in fig. Cepstrum of each block is computed by making a voiced/unvoiced decision. This is carried by fixing a threshold value. If the peak value exceeds the threshold then the section is classified as voiced. Otherwise a zero crossing count of block is made and if it exceeds threshold value then block is classified as unvoiced [5].

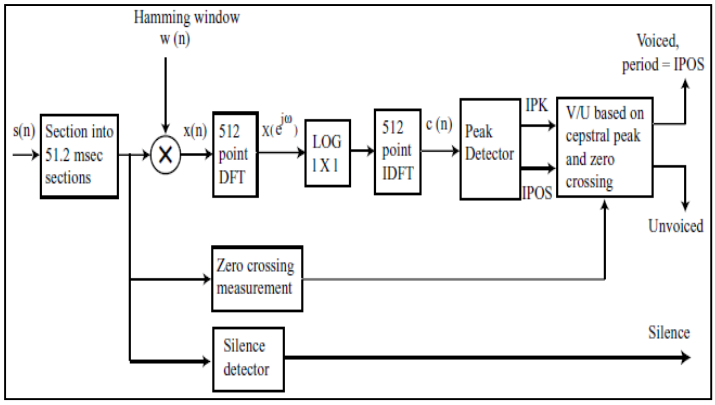

Fig. 1: Block diagram of Cepstrum pitch detector

2) Hybrid Method

Hybrid method uses features of both time domain and frequency domain methods. It uses frequency domain to obtain a spectrally flattened waveform and then uses time domain method to find the pitch period. Some examples of this method involve SIFT and spectral equalization.

a) Simplified Inverse Filtering Technique (SIFT)

In this method initially the speech samples are low pass filtered and then the decimation of the samples is done. To find the coefficients of inverse filter the autocorrelation method of LPC analysis is done. Next, inverse filtered speech signal is auto correlated to obtain pitch period by interpolating the function with the nearest peak [5].

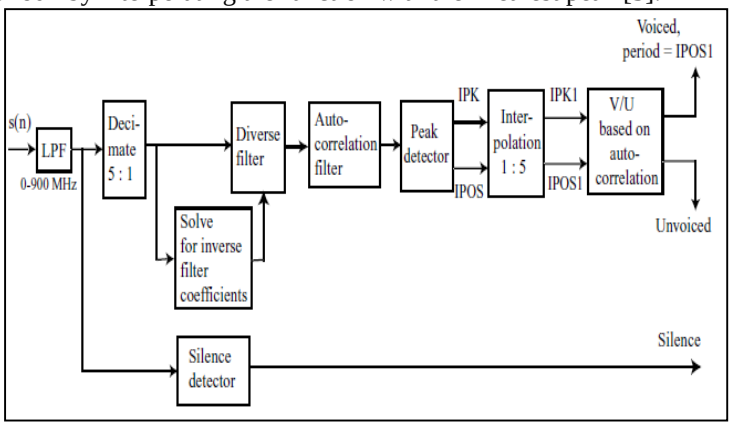

Fig. 2: Block diagram of SIFT pitch detector

3) Spectral Equalization LPC Method:

Here pattern recognition technique is used to classify speech signal as voiced or unvoiced. For voiced part further processing is performed (LPF and decimation).

In order to spectrally flatten the speech waveform Newton transformation is used. A peak picker is used to determine the pitch period at the 2-kHz rate and a simple interpolation network is used to obtain higher resolution in the value of the pitch Period.

The voiced-unvoiced pattern recognition algorithm uses a training set which provides a statistical description of the measurements used in the algorithm for each of classes. The success of this method of making a voiced-unvoiced decision depends heavily on how well the training set of data characterizes the different speech classes [5].

C. Applications

1) Speech Coding

Determination of correct pitch is very essential in noisy to enhance quality of environment plays an essential role for quality speech coding. The input voice signal of communication systems is often distorted by background noise and hence there is a need of reliable Pitch detection algorithm (PDA), which performs well under any reasonable noise condition. Verteletskaya et.al [15] have carried a research work in which a new pitch detection algorithm named "Spectro –temporal-autocorrelation (STA)" proposed by Kondoz was tested along with the autocorrelation, cepstrum and AMDF method. This new method was found to work well under different noise conditions. Evaluation of pitch detection algorithms were based on following parameters Gross error it is the percentage of voiced frames with an estimated F0 value that deviates from the reference value by more than 20% Classification error is the percentage of unvoiced frames classified as voice and voiced frames classified as unvoiced.

Evaluation of PDA's were carried out on pure speech signal results and were noted. Pitch detection algorithm evaluation on adverse conditions with SNR 20dB, 10dB and 5dB.STA algorithm dominated other algorithms and was found suitable to be implemented in speech coding systems working with noise corrupted speech.

2) Voice Pathology Diagnosis

Modification of three basic pitch detection algorithms namely autocorrelation, HPS and wavelet transform were carried out by Cosme et al. [8].This was done for diagnosing three voice pathologies that is roughness, hoarseness and breathiness. Database of healthy voice was compared with the pathological voice obtained by using the software "Dr. Speech".

3) Speech Enhancement

One main challenge being faced in speech processing is the presence of background noise. Xin et al. have described about spectral subtraction method for speech enhancement. For validating the efficiency of proposed method four types of noise that is white , pink ,factory, babble were added to speech signal. A good speech enhanced effect was observed. [9]

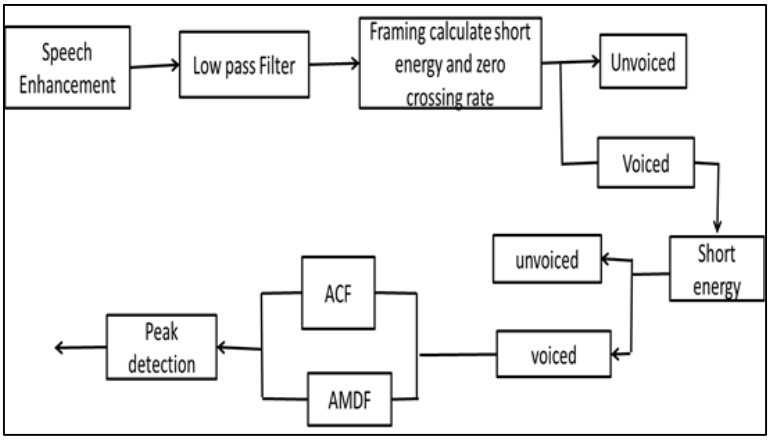

Fig. 3: Flow diagram of pitch detection [9]

III. IMPLEMENTATION OF AUTOCORRELATION METHOD

Autocorrelation method of detecting pitch is one of the most reliable and robust method. Autocorrelation function correlates the signal with itself. Here each element is multiplied by a shifted version of the input and the results are added to obtain the autocorrelation. From literature survey it was observed that for non-stationary signals like speech the short time autocorrelation is useful rather than long- time autocorrelation[1] [5] [10] [11] .

$$R_n(k) = \sum_{m=-\infty}^{\infty} \left[x(m)w(n-m)x(m+k)w(n-k-m)\right] \qquad (6)$$

w(n) is appropriate window length

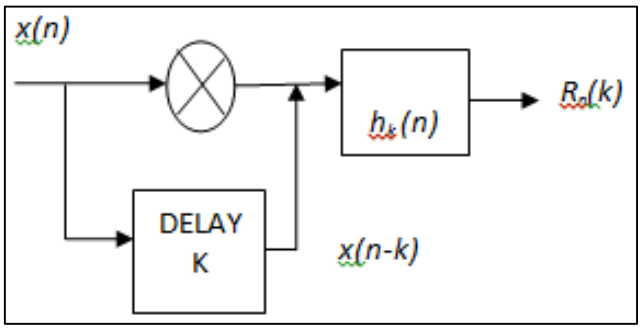

Fig. 4: Block Diagram Representation of short time autocorrelation[1]

A. Algorithm

Initially recording of speech signal was carried out six different speakers including two men,two women and two children(a boy and a girl).For adults recording was carried out at 12KHz and for children it was done at 32KHz.Later a windowed signal was obtained. and autocorrelation function was calculated using the formula mentioned above.Pitch period was obtained between the highest peak and second largest peak.And then pitch frequency was also found.

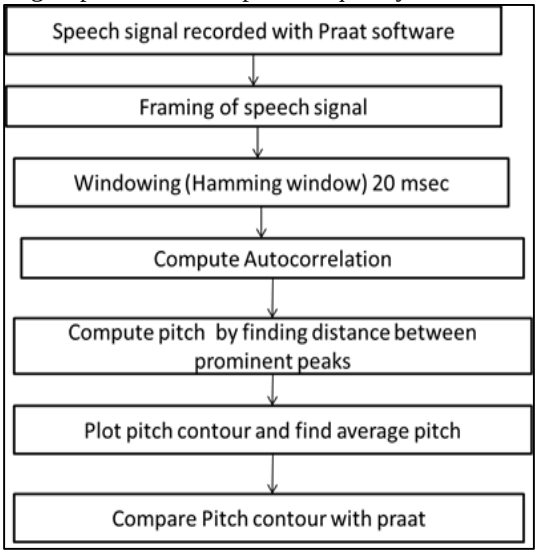

Fig. 5: Algorithm of Autocorrelation

B. Results

Five vowels /a/,/i/,/e/,/o/,/u/ uttered by two men,two women and two children were used to calculate the pitch values making use of autocorrelation method. Fig.6 shows experimental results of autocorrelation method of vowel 'a' spoken by male speaker 1. Speech waveform of recorded vowel is shown in part (a) and part (b) presents the wideband spectrogram. The part (c) of figure pitch contour implemented algorithm which is similar to pitch contour from Praat software as shown in part (d).

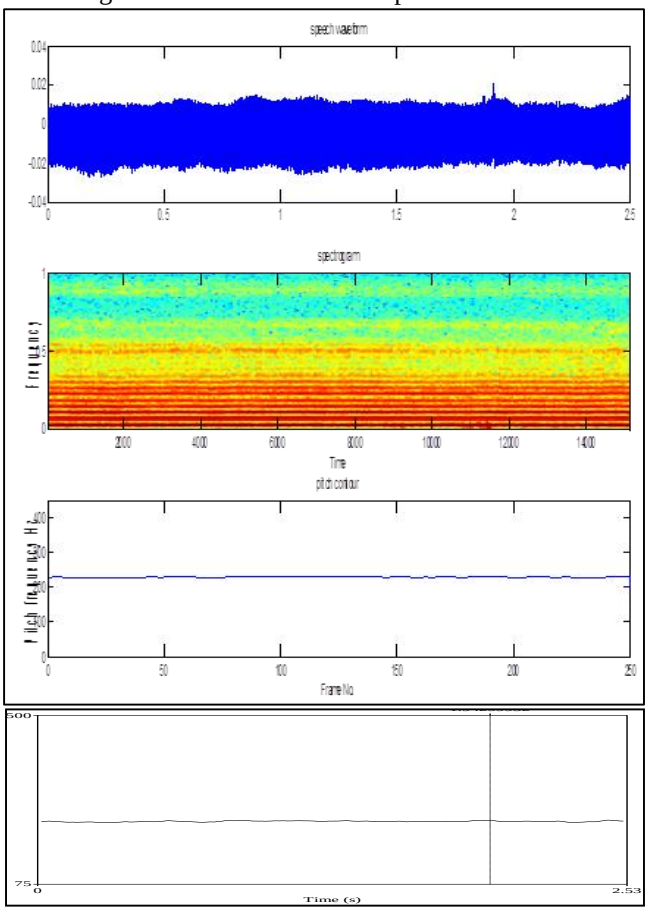

Fig. 6: Pitch contour from autocorrelation algorithm and Praat software for /a/ vowel. (a)Speech waveform (b) spectrogram of speech signal (c) Implemented autocorrelation pitch estimation (d) Pitch estimation by Praat software

Table I gives a comparison of pitch values of vowels uttered by Male speaker 1 using the Matlab and the PRAAT software. It can be noted that the pitch values obtained are almost similar with slight variations. Similarly results for other speakers were obtained.

Vowels uttered by male 1	Pitch frequency value from MATLAB (Hz)	Pitch frequency value from PRAAT (Hz)
/a/ (arm)	99.173	100.2
/i/(see)	122.449	123.9
/u/(foot)	133.33	133.5
/o/(go)	123.71	125
/e/ (met)	120	121.9

Table - 1. Pitch values of vowels uttered male speaker1

IV. CONCLUSION

In this paper a brief review on factors affecting pitch and various method of detecting pitch are presented.Autocorrelation method of finding pitch has been implemented.From the results For men the results of pitch estimation were in the range of 80-140Hz.For women the results of pitch estimation were in the range of 200-300Hz.And for chidren the results of pitch estimation were in the range of 250-350Hz.These values were found to be very much matching with the actual values.Results obtained were validated using the PRAAT software.

REFERENCES

[1] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, NJ, 1978.
[2] T. B. Patel and H.A. Patil "Novel Approach for Estimating Length of the Vocal Folds using Fujisaki Model" IEEE 9th International Symposium on Chinese Spoken Language Processing (ISCSLP),pp:308-312,Sep2014.
[3] S. Kumar, S. Bhattacharya, P. Patel. "A New Pitch Detection Scheme Based on ACF and AMDF".IEEE Conference on Advanced Communication Control and Computing Technologies (ICACCCT), pp: 1235 -1240, May 2014.

[4] V. Bageshree, S. Pathak, A.R. Panat "Extraction of Pitch and Formants and its Analysis to identify 3different emotional states of a person", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012.
[5] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. Mc Gonegal, "A comparative performance study of several pitch detection algorithms", IEEE Trance. Acoust., Speech Signal Processing(ASSP), vol. ASSP-24, pp. :399-418, Oct. 1976.
[6] M J. Ross "Average Magnitude Difference Function Pitch Extractor", IEEE Trans ASSP, vol. 77, pp.353 -362, 1974 .
[7] N. Sripriya, T. Nagarajan, "Pitch Estimation Using Harmonic Product Spectrum derived from DCT", IEEE International conference of IEEE Region 10, Oct 2013.
[8] L. Cosme, L. lvarez, D. Ayllo "Pitch detection in pathological voices driven by three tailored classical pitch detection algorithms", Recent Advances in Signal Processing, Computational Geometry and Systems Theory. In Proc 11th WSEAS internantional conference on signal processing, computational geometry and artificial vision, pp.: 113-118, 2011.
[9] X. Xu, T. Zhang, Y. Zhang "An Improved "Pitch Detection of Speech Combine With Speech Enhancement", IEEE International Conference on Image and Signal Processing", 2014.
[10] Li Tan and M. Karnjanadecha, "Pitch Detection Algorithm: Autocorrelation Method and AMDF".Proceedings of the 3rd International Symposium on Communications and Information Technology, pp.:551-556, 2003.
[11] S.S. Upadhya, "Pitch Detection in Time and Frequency Domain", IEEE International Conference on Communication, Information and Computing Technology, pp.:1-5, Oct. 2012.
[12] L.R.Rabiner, "On the use for autocorrelation analysis for pitch detection". IEEE Transaction on Acoustics, Speech Signal Processing, pp.: 24-33, Feb 1977.