INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1681)
Available online at: https://www.ijariit.com

Fundamental Frequency estimation and analysis of speech signal

Mahesh M. Kamble
vickkyshree9@gmail.com
PVPIT, Budhgaon, Sangli (affiliated to Dr. Babasaheb Ambedkar Technological University) Maharashtra

Tejal S.Bandgar
tejalbandgar.111@gmail.com
PVPIT, Budhgaon, Sangli (affiliated to Dr. Babasaheb Ambedkar Technological University) Maharashtra

ABSTRACT

Fundamental frequency is a critical component in speech signal processing analysis. The fundamental frequency (fo) is the rate at which the vocal cords vibrate, and the fundamental frequency range for a person is 120 to 400 Hz. This basic frequency varies depending on the size and form of the vocal cords, and it might differ for males, females, and children. Different time domain and frequency domain pitch detection techniques are utilised. The time domain methods include autocorrelation and AMDF (Average Magnitude Difference Function), whereas the frequency domain algorithm is Cepstrum. The fundamental frequency may be determined by pitch preprocessing and extraction.

Keywords— *Autocorrelation function, Speech Recognition System, Center-clipping Pitch.*

1. INTRODUCTION

Signals that are audible to humans are referred to as audio signals. Audio signals are generated by a sound source that vibrates at audible frequencies. The vibrations cause pressure waves to develop in the air, which travel at a speed of roughly 340 metres per second. These pressure signals are received by our inner ears, which then convey them to our brain for further analysis. Pitch, also known as fundamental frequency, is a key characteristic of audio transmissions. It represents the vibration frequency of the sound source. In other words, pitch is the reciprocal of the fundamental period, which is the fundamental frequency of audio signals. The time domain and the frequency domain are used to estimate pitch or fundamental frequency.In time domain Autocorrelation method and AMDF(Average Magnitude Difference Function) method can used whereas in frequency domain Cepstrum method is used. The human voice production system involves following steps as open and close of vocal cords which is also called as glottis which create a vibration in the air flow, a resonance is provided by Oral cavity, pharyngeal cavity, and nose cavity; there is a approach of a voiced signal or unvoiced signal .Due to the glottis pressure and the pushed air.The vocal chords get air from the lungs, open and close rapidly and further modulated by the resonances of pharyngeal, oral and nasal cavities. Pitch is determined by the vibration frequency by the vocal cords. The fundamental frequency is different for male, female and children category.

2. BACKGROUND

A. *Autocorrelation Method and AMDF*

The auto-correlation function (ACF) for pitch tracking is very basic method. This method is a time domain method which is used to determine the similarity between a frame and a shifted or delayed version of the frame.

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

Where N is the frame length, n is the time index in a frame, m is the shift in the signal, x (n) is the sampled signal.Here process is to shift the delayed version m times and compute the inner product of the overlapped parts to obtain n values of ACF.

MADF is very close to ACF except that it estimates the distance instead of similarity between a frames and its delayed version using following formula

$$D_{\mathrm{m}} = \frac{1}{L} \sum_{n=1}^{L} \left| (X(n) \; - X(\mathrm{n-m}) \right|$$

Where The samples of input voice are x (n), while the samples time shifted by m seconds are x (n-m)..

B. *Cepstrum*

The Cepstral method gives analyzer which is designed primarily for use in speech analysis. The logarithm of each successive amplitude spectrum so produced can be fed into a second spectrum analyzer of the same type. The cepstrum, or power spectrum, of the logarithm spectrum is then produced by this analyzer. A speech example with X referring to the spectrum of the speech signal, F to be the excitation components for instance, the glottal pulse train and V to the vocal tract shaping of the excitation spectrum.

C. *Feature Extraction*

Frame blocking is the process of converting a stream of audio signals into a sequence of frames. The signal s(n) is multiplied by a fixed length analysis window w(n) to extract a particular segment at a time. This is called windowing. Speech signal includes very rich harmonic components. The fo varies in the range about 80 Hz to 500 Hz as per the age and gender of person. Pitch detection is ineffective over 500 Hz. Pitch detection is implemented using a low pass filter with a pass band of up to 900 Hz. One specific shape nonlinear processing is commonly employed centre-clipping of speech to decrease the influence of formant structure. Energy is a characteristic that may be utilised to distinguish between voiced and unvoiced communication. It takes into account the limited time energy. Short time means the energy of analysis period.

$$E = \sum_{k=-\infty}^{\infty} s(n)^{2}$$

Table 1. PDAs with formulae

Sr. No	*PDA's*	*Formulae*
1	*Autocorrelation function*	$Rx(m) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x(n+m)$
2	*AMDF function*	$D_m = \frac{1}{L} \sum_{n=1}^{L} \lvert X(n) - X(n-m) \rvert$
3	*Cepstrum function*	$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \lvert X(\omega) \rvert e^{j\omega n} d\omega$

3. IMPLEMENTATIONS

The approach requires that the speech signal be low-passed filtered to 900 Hz, according to the discussion above. For processing, the filtered signal is digitised at a sampling rate of 10-kHz and divided into overlapping 30-ms (300 samples) portions. Because all pitch detectors compute the pitch period 100 times per second, or every 10 milliseconds, neighbouring portions overlap by 20 milliseconds or 200 samples. The determination of a clipping threshold CL for the 30-ms portion of speech is the second stage of processing. The clipping threshold in the beginning and last 10-ms parts of the section is set to 68 percent of the smaller of the peak absolute sample values. Following the clipping level determination, the 30-ms segment of speech is centre clipped, followed by infinite peak clipping. Then the energy of the signal is calculated. After computing the ACF/AMDF according to peak values the fundamental frequency is estimated.

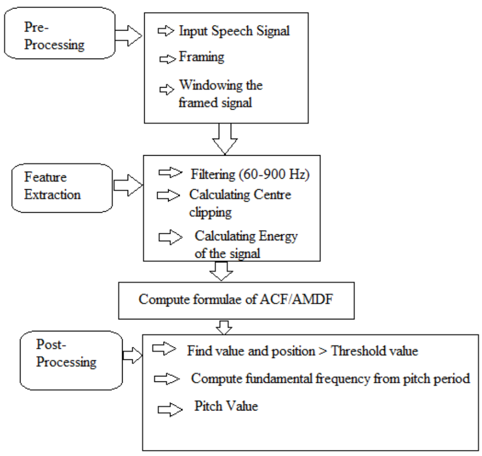

Fig 1. ACF/AMDF technique for fundamental frequency estimate (fo) on the blockchain

4. RESULTS

Table 2. Comparison table of three pitch detection algorithm for adult male samples

Subject/Age	Autocorrelation	AMDF	Cepstrum
M1/42	162.12	162.11	205.29
M2/38	152.98	152.65	166.19
M3/30	166.57	172.40	198.45
M4/48	171.61	173.58	203.73
M5/35	164.29	167.04	183.83
M6/28	195.54	199.68	255.88

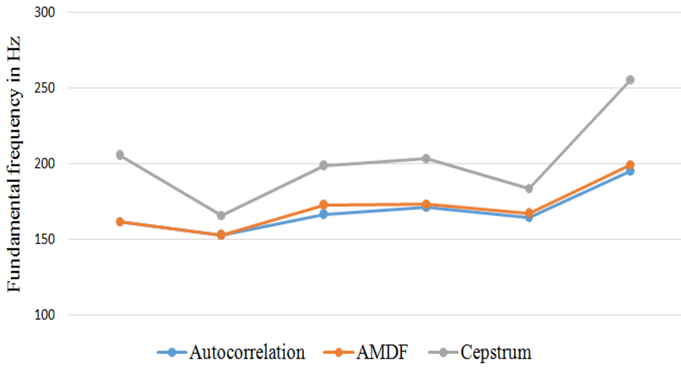

Fig 2. Comparison chart of three pitch detection algorithm for adult male samples

Table 3. Comparison table of three pitch detection algorithm for adult female samples

Subject/Age	Autocorrelation	AMDF	Cepstrum
F1/38	232.92	233.15	260.82
F2/34	250.96	250.34	255.78
F3/30	280.40	280.92	279.88
F4/38	234.50	235.11	240.67
F5/40	241.7	241.46	258.14
F6/38	244.67	244.54	274.32

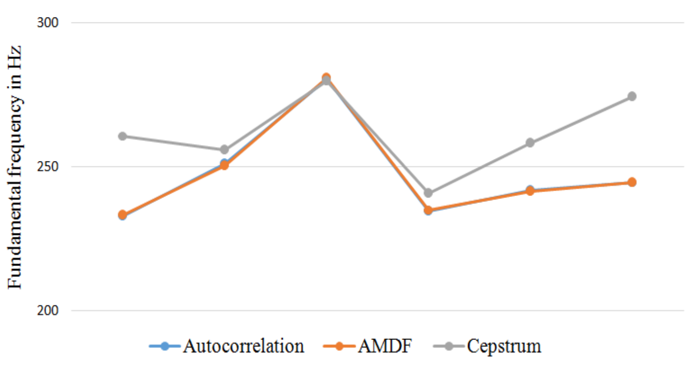

Fig 3. Comparison chart of three pitch detection algorithm for adult female samples

Table 4. Comparison table of three pitch detection algorithm for children's samples

Subject/Age	Autocorrelation	AMDF	Cepstrum
c1/13	300.61	300.94	306.33
c2/11	302.42	302.80	339.49
c3/13	314.96	314.32	324.42
c4/9	306.93	306.93	316.84
c5/13	349.79	349.3	354.68
c6/15	338.20	337.66	345.52

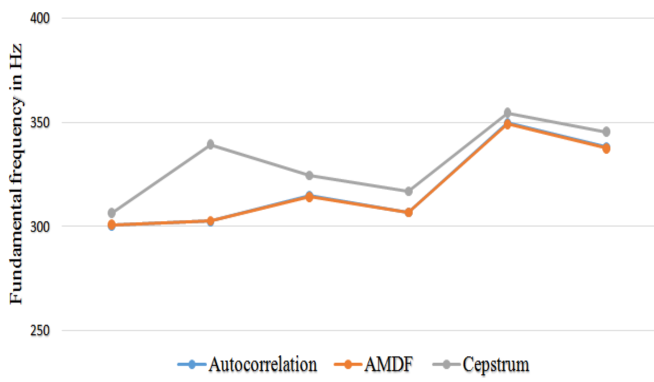

Fig. 4 Comparison chart of three pitch detection algorithm for children's samples

4. CONCLUSIONS

The fundamental frequency of a sound wave will determine the human tone and pitch. The method to ascertain the pitch period of voice signal is important not only in speech signal processing, but also in diagnosing vocal cord symptoms.

Pitch is the most distinctive difference between male and female speakers. Three PDAs based on the autocorrelation function, the average magnitude difference function and cepstrum analysis are introduced.The results show that children has a fundamental frequency near or above 300 Hz whereas female has fundamental frequency in the range 210-280 Hz whereas male fundamental frequency lies in between 150-220 Hz.

5. REFERENCES

[1] SPICE: Self-Supervised Pitch Estimation Beat Gfeller, Christian Frank, Dominik Roblek, Matt Sharifi, Marco Tagliasacchi, and Mihajlo Velimirovi´ IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 28, 2020

[2] C. Yeh, A. Roebel, X. Rodet, Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals, IEEE Trans. Audio Speech Lang. Process. 18 (6) (Aug. 2010) 1116–1126.

[3] E. Nakamura, E. Benetos, K. Yoshii, S. Dixon, Towards complete polyphonic music transcription: integrating multi-pitch detection and rhythm quantization, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 101–105.

[4] L. Rabiner; M. Cheng; A. Rosenberg; C. McGonegal A comparative performance study of several pitch detection algorithms IEEE Transactions on Acoustics, Speech, and Signal Processing (Volume: 24, Issue: 5, Oct 1976)

[5] Menglu Li, Zhijun Zhao, & Ping Shi Query by Humming Based on the Hierarchical Matching Algorithm 2015 IEEE International Conference on Computer and Communications（ICCC 2015）

[6] Smt. Sanjivani S. Bhabad, G. K. Kharate, Smita C. Shinde, 'Pitch detection in time, frequency and cepstral domain for articulatory handicapped people',978-1-4799-1626-9/$31.00 2013 IEEE.

[7] Savitha S Upadhya, 'Pitch Detection in Time and Frequency Domain', 2012 International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20, Mumbai, India, 978-1-4577-2078-9/12/$26.00©2011 IEEE.

[8] zhijun Cui, 'Pitch Extraction Based on Weighted Autocorrelation Function in Speech Signal Processing', 2012 2nd International Conference on Computer Science and Network Technology, 978-1-4673-2964-4/12/$31.00 ©2012 IEEE.

[9] Jeiran Choupan, SeyedGhorshi, Mohammad Mortazavi, FarshidSepehrband, 'PITCH EXTRACTION USING DYADIC WAVELET TRANSFORM AND MODIFIED HIGHER ORDER MOMENT', 978-1-4244-6871-3/10/$26.00 ©2010 IEEE.

[10] YANG Fan, LIU Ming-hui, XU Sun-hua, PAN Guo-feng, 'Research on aNew Method of Preprocessing and Speech Synthesis Pitch Detection', 2010 International Conference On Computer Design And Appliations (ICCDA 2010), 978-1-4244-7164-51$26.00 © 2010 IEEE

[11] Geliang Zhang, Simon Godsill, 'TRACKING PITCH PERIOD USING PARTICLE FILTERS', 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 978-1-4799-0972-8/13/$31.00 ©2013IEEE.

[12] Yuan Zong, Yumin Zeng, Mengchao Li, RuiZheng, 'Pitch Detection Using EMD-Based AMDF', 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP) June 9 – 11, 2013, Beijing, China, 978-1-4673-6249-8/13/$31.00 ©2013 IEEE.

[13] www.scribd.com

[14] www.es.scribd.com

