Fundamental Frequency Estimation and Analysis of Speech Signal

Mahesh M. Kamble*, Tejal S. Bandgar
Electrical Department, Babasaheb Ambedkar Technological University, Lonere, PVPIT Budhgaon, Sangli, Maharashtra, India
****Corresponding Author***
E-Mail Id: vickkyshree9@gmail.com

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 domain of time and frequency pitch estimation techniques are utilized. 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, pitch detection algorithm, pitch detection algorithm, pitch detection algorithm, pitch detection algorithm, pitch*

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 meters 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 sound source's frequency of vibration. Pitch can be term as requite of the rudimentary period, or the rudimentary frequency of audio signals. The time discipline or domain and the frequency discipline are used to estimate pitch or fundamental frequency. In time discipline Autocorrelation process and AMDF (Average Magnitude Difference Function) process can used whereas in frequency discipline Cepstrum process 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 an 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.[1-7]

BACKGROUND

Autocorrelation Method and AMDF

Pitch tracking using the auto-correlation function (ACF) is a simple approach. 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.

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

N is the breadth of the frame, n is the index of time in a frame, m is the shifted version in the signal, x (n) is the signal which can be called as sampled. To acquire n values of ACF, shift the delayed version m times and compute the inner product of the overlapping portions. AMDF is quite similar to ACF, except that it uses the following formula to determine the distance between a frame and its delayed version, rather than similarity.

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

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

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 voice example, with X referring to the speech signal's spectrum, F referring to excitation components such as the glottal pulse train, and V referring to the excitation spectrum's vocal tract shaping.

Feature Extraction

The process of turning a stream of audio signals into a sequence of frames is known as frame blocking. To extract a single segment at a time, the signal s(n) is multiplied by a specified length analysis window w(n). This is alluded to as windowing. Speech signal includes very rich harmonic components. The f_o 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 center-clipping of speech to decrease the influence of formant structure. Energy is a characteristic that may be utilized to distinguish between voiced and unvoiced communication. It takes into account the limited time energy. Short time means the energy of analysis period.

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

PDA's	Formulae
Autocorrelation Function	$Rx(\mathrm{m}) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(\mathrm{n})\, x(\mathrm{n}+\mathrm{m})$
AMDF Function	$D_{\mathrm{m}} = \frac{1}{L} \sum_{n=1}^{L} \left\| (X(n) - X(\mathrm{n}-\mathrm{m}) \right\|$
Cepstrum Function	$c(\mathrm{n}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left\| \mathrm{x}\ (\omega) \right\| e^{j\omega \mathrm{n}} d\omega$

IMPLEMENTATIONS

According to the discourse above, the approach desire that the speech signal be low-passed filtered to 900 Hz. The filtered signal is digitized at a 10-kHz sampling rate and split into overlapping 30-ms (300 samples) chunks for refining. Because all pitch detectors estimate the pitch period

100 times per second, or every 10 milliseconds, adjacent sections of the waveform overlay by 20 milliseconds, or 200 samples. The second stage of refining involves identifying a clipping threshold CL for the 30-ms segment of speech. In the first and last 10-ms of the segment, the clipping threshold is set to 68 percent of the smaller of the peak absolute sample values. After determining the clipping level, the 30-ms speech segment is center clipped, followed by indefinite peak clipping.. Then the energy of the signal is calculated. After computing the ACF/AMDF according to peak values the fundamental frequency is estimated.

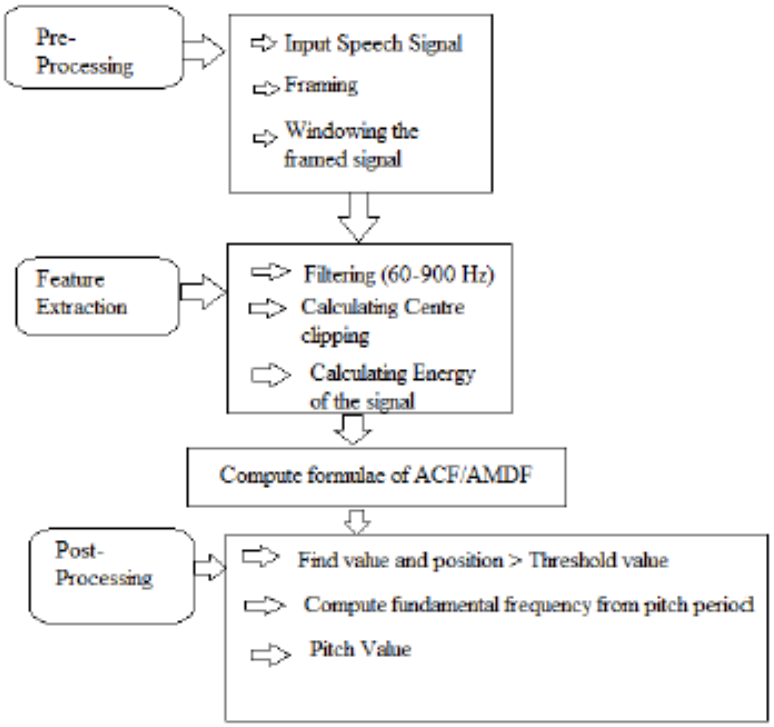

Fig. 1: *ACF/AMDF Technique for Fundamental Frequency Estimate (fo) on the Block Chain.*

RESULTS

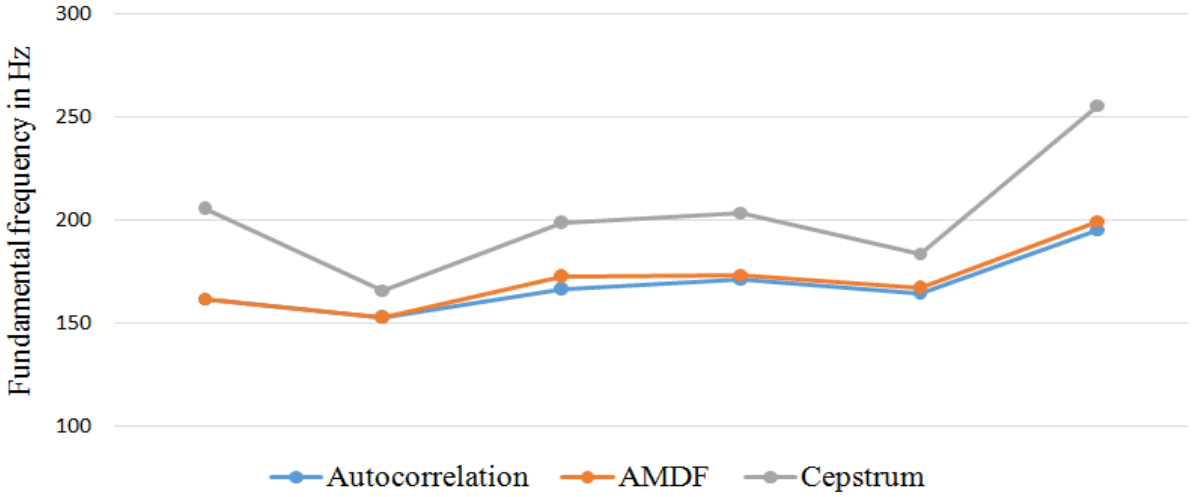

Fig. 2: *Comparison Chart of Three Pitch Detection Algorithm for Adult Male Samples.*

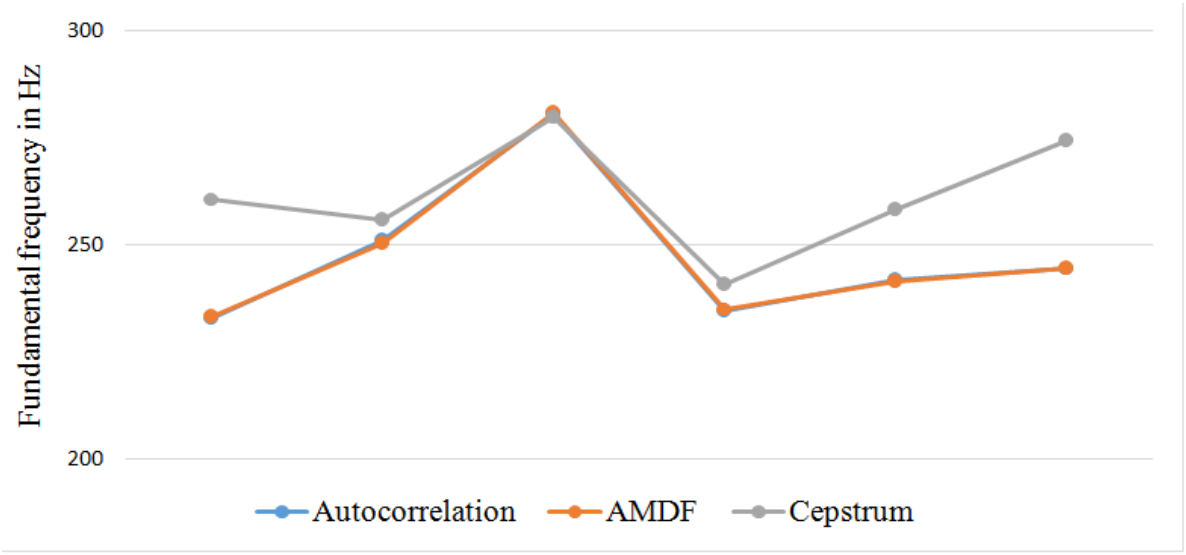

Fig. 3: *Comparison Chart of Three Pitch Detection Algorithm for Adult Female Samples.*

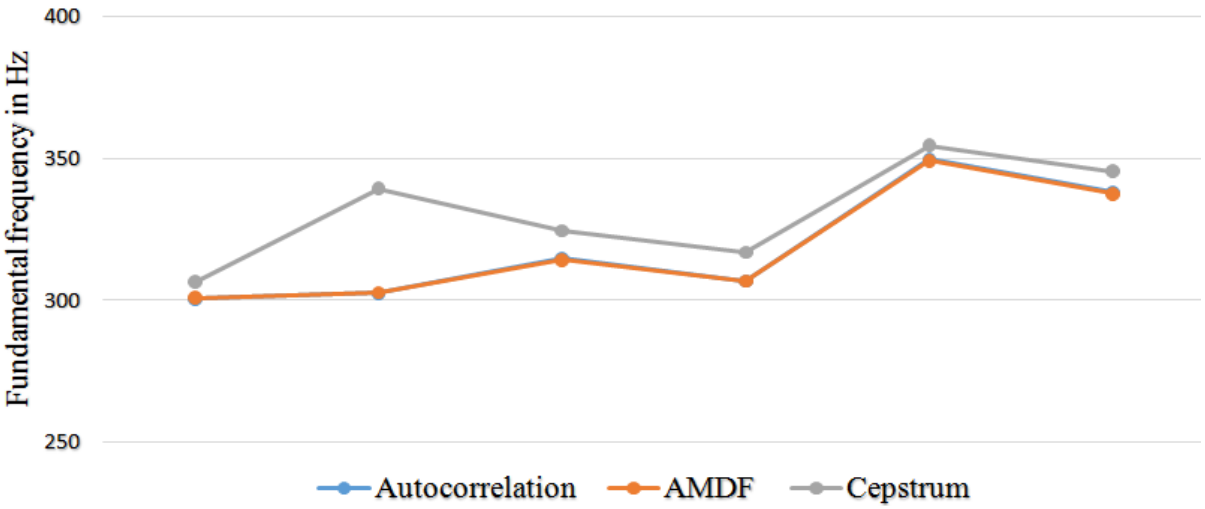

Fig. 4: *Comparison Chart of Three Pitch Detection Algorithm for Children Samples.*

CONCLUSIONS

The tone of human and pitch are influenced by the fundamental frequency of a sound wave. 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. The autocorrelation function, the average magnitude difference function, and cepstrum analysis are three PDAs that have been 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.

REFERENCES

1. Tripicchio, P., Satler, M., Dabisias, G., Ruffaldi, E., & Avizzano, C. A. (2015, July). Towards smart farming and sustainable agriculture with drones. In *2015 International Conference on Intelligent Environments* (pp. 140-143). IEEE.
2. Davidson, M. W., Le Toan, T., Mattia, F., Satalino, G., Manninen, T., &

Borgeaud, M. (2000). On the characterization of agricultural soil roughness for radar remote sensing studies. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 630-640.

3. Spoorthi, S., Shadaksharappa, B., Suraj, S., & Manasa, V. K. (2017, February). Freyr drone: Pesticide/fertilizers spraying drone-an agricultural approach. In *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)* (pp. 252-255). IEEE.
4. Einicke, G. A., & White, L. B. (1999). Robust extended Kalman filtering. *IEEE Transactions on Signal Processing*, *47*(9), 2596-2599.
5. Rusu, R. B., & Cousins, S. (2011, May). 3d is here: Point cloud library (PCL). In *2011 IEEE international conference on robotics and automation* (pp. 1-4). IEEE.
6. Fukao, T., Nakagawa, H., & Adachi, N. (2000). Adaptive tracking control of a nonholonomic mobile robot. *IEEE transactions on Robotics and Automation*, *16*(5), 609-615.
7. Huang, Y., Hoffman, W. C., Lan, Y., Fritz, B. K., & Thomson, S. J. (2015). Development of a low-volume sprayer for an unmanned helicopter. *Journal of Agricultural Science*, *7*(1), 148.

