International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

A Review of Diverse Pitch & Detection Methods

Shikha Rani[1], Nitin Jain[2]

[1]M.Tech (CSE), Hindu College of Engineering, Sonipat, Haryana, India

[2]Assistant Professor (CSE), Hindu College of Engineering, Sonipat, Haryana, India

Abstract: ***Pitch is that auditory attribute of sound according to which sounds can be ordered on a scale from low to high. Some of the factors are influencing the fundamental frequency are such as body size, laryngeal size, vocal fold length. In this paper we are discussing about the pitch detection algorithms designed to estimate the pitch or fundamental frequency of a periodic signals such as time domain, frequency domain and hybrid domain.***

Keywords: Pitch, Fundamental Frequency, Harmonics, Detection Techniques.

1. Introduction

According to Webster dictionary "The property of a sound and especially a musical tone that is determined by the frequency of the waves producing it: highness or lowness of sound". The ANSI definition of psycho acoustical terminology says that "pitch is that auditory attribute of sound according to which sounds can be ordered on a scale from low to high [1] ."The Pitch is represented by the fo (fundamental frequency) of sound wave vibration .Pitch is the reciprocal of pitch period i.e. The time duration of one glottal cycle. The sensation of a frequency can be referred as a pitch of the sound .pitch quality and loudness are the important characteristics of sound. The sound of only one frequency is pure tone such as given by a electronic signal generator. The fundamental frequency of vibration is referred as the pitch because it has a larger intensity. Loudness depends on the amplitude of the sound wave .The larger the amplitude the more energy the sound wave contains therefore the louder the sound .Loudness is a physiological sensation it depends on sound pressure but also on the spectrum of the harmonics and the physical duration .This is used to describe the quality of the waveform as it appears to the listener. Therefore the quality of a sound depends upon the waveform.

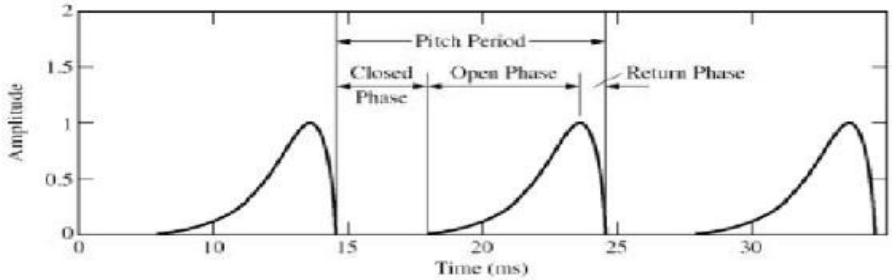

Figure 1: Representation of Pitch Period [14]

In this Fig 1.represent the pitch period and x -axis represent the time and y –axis represents the amplitude. The pitch period consist of 3 phases i.e. closed phase, open phase and return phase.

2. Factors Influencing the Fundamental Frequency (F0)

1. Body Size

The most obvious influence on pitch that comes to mind is the size of the sound producing apparatus; we can observe from the instruments of the orchestra that little objects tend to make higher-pitched sounds and big ones produce lower-pitched sounds. Therefore, it is logical to assume that shortest people would make high pitch, and longest people would make low pitch. And this assumption is borne out by the facts, at least to an extent. The fundamental frequency(referred to as fo) of baby cries from around 500 Hz. Child speech ranges from 250-400 Hz, adult females tend to speak at around 200 Hz on average and adult males around 125 Hz. Thus, the body size is one of the factors related to f0.

2. Laryngeal Size

Perhaps, a factor more relevant to the voice source is the size of the larynx. Men, on average, have a larynx about 40% taller and longer (measured along the axis of the vocal folds) than women, as seen in Fig 2. In this, there is not completely explain the difference between male and female fundamental frequency f0; there is a size difference inside the larynx, which fully explains the difference in f0.

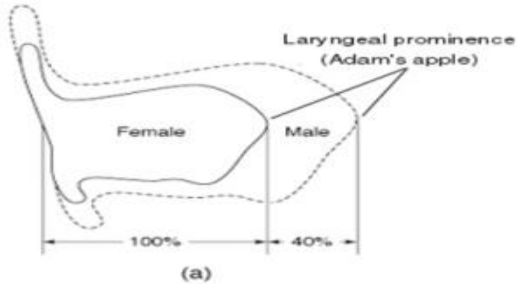

Figure 2(a): Laryngeal shape of female and male speaker

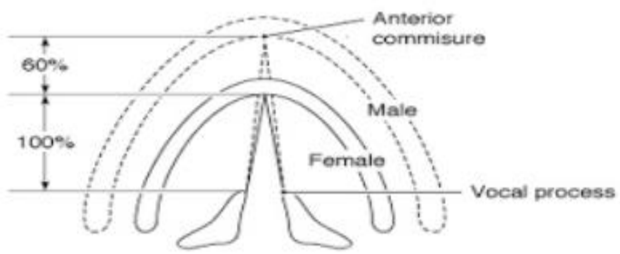

Figure 2(b): Relative sizes of the laryngeal

3. Vocal Fold Length

In this factor is discussed about the length of the vocal fold. If it is assumed that the vocal folds are 'ideal strings' with uniform properties, their fundamental frequency f0 is governed by Equation 1.

$$F_0 = \frac{1}{2L}\sqrt{\frac{\sigma}{\rho}} \quad \ldots.(1)$$

Where,

L: Length of vocal folds, σ: Longitudinal stress, ρ: Tissue density.

Briefly, some other factors that influence the fundamental pitch period are:

4. Difference between languages.
5. Specifics of different applications.
6. Emotional state of different person.
7. Environmental conditions under which speech is produced.

3. Pitch Detection

Fundamental frequency estimation (fo) is refers to as pitch detection. Pitch detection algorithms designed to estimate the pitch or fundamental frequency of periodic signals. Pitch detection is of interest whenever a single quasiperiodic sound source is to be studied or modeled, specifically in speech and music [2][3].

Pitch detection methods can be classified into various approaches.

1. Time domain detection
2. Frequency domain detection
3. Hybrid domain detection.

1 .Time Domain Detection

A group which utilizes principally the time-domain properties of speech signals. Time domain is the analysis of mathematical function, time series of economic data with respect to time. A time domain graph shows how a signals changes with time. Time-domain pitch detectors operate directly on the speech waveform to estimate the pitch period The assumption is if quasiperiodic signals is to reduce the effect of formant structure then simple time-domain measurements will provide good estimates of the period. A time domain feature detection method the signal is usually preprocessed to accentuate some time domain feature, then the difference between the time occurrences of that feature is calculates as the period of the signal[4,5,6].

Some methods are:

a)Zero-Crossing Rate-.zero crossing rate is a simple technique that consists of counting the no of times signal crosses at zero reference .It is easy technique but cannot give accurate results. Its dealing with the noisy signals where the fundamental is lesser then the partials, This results gives poor result.

b) Autocorrelation Function.-In this pitch detection technique mainly focused on the time domain. In this, when a segment of a signal correlated with itself. It is most useful when the frequency is low to mid .in this basically focused on the distance between the positions of maximum and second maximum correlation .This detection is mainly used in speech recognition application where the range of pitch is limited .It is very cheaper way to calculate the fft13].

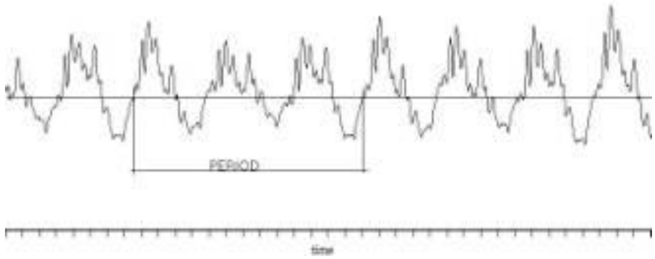

Figure 3: (a) Harmonic signal in time domain showing that the zero crossing technique could lead to unclear results[5].

2. Frequency Domain Detection

A group which utilizes principally the frequency-domain properties of speech signals. .Frequency domain is the analysis of mathematical function or signals with respect to frequency .Frequency stands for no. of cycle(s) per unit time .A frequency domain graph shows how much of the signals lies within each given frequency band over a range of frequencies .The property that are use in frequency domain pitch detector, if the signal is periodic then the frequency spectrum consist of a series of impulses at the f0 and its harmonics i.e. 2f0, 3f0.... Frequency domain methods call for the signal to be frequency transformed, then the frequency domain representation is examine for the first harmonic, the biggest common divisor of all harmonics, or other such implication of the period. Window’s signal is confirm to avoid spectral marking, and depending on a window's type, a minimum number of periods of the signal must be study to enable accurate location of harmonic peaks [7][8] Thus simple measurements can be made on the frequency spectrum of the signal (or a nonlinearly transformed version of it as in the cepstral pitch detector [9]) to estimate the period of the signal.

Some methods are:

a) ***Harmonic Product spectrum***- The fundamental frequency can be determined by measuring the frequencies of its top harmonic components [10] and computing the biggest common divisor of these harmonic frequencies[11]. The biggest common divisor can be determined by making an entry to a frequency histogram for each harmonic frequency and at integer divisions of the harmonic frequency. The frequency at the peak of the scatter diagram i.e. histogram represents the biggest common divisor, and hence the fundamental frequency. Some nice features of the method include: low cost, unaffected to additive and repetition noise, changeable to different kind of inputs (changing the number of harmonics to consider or replacing multiplication by ads).

b) *Cepstrum*- A cepstrum (/ˈkɛpstrəm',ˈsɛpstrəm'/) is the result of taking the Inverse Fourier transform (IFT) of the logarithm of the estimated spectrum of a signal. It may be pronounced in the two ways given, the second having the advantage of avoiding confusion with 'kepstrum' which also exists .The power cepstrum is useful to finds the particular application in the analysis of human speech .The name "cepstrum" was derived by reversing the first four letters of "spectrum". Operations on cepstra are labeled quefrency analysis, liftering, or cepstral analysis. The term cepstrum [12] is formed by reversing the first four letters of spectrum. In this basically focused on this idea to take the Fourier transform to the log-magnitude Fourier spectrum. Thus, if the original spectrum belongs to a harmonic signal, it is moving to be periodic in the frequency representation, and taking the Fast Fourier Transform again it will show a peak corresponding to the period in frequency, thus we can confine the fundamental period. It can be also interpreted as a de-convolution process.

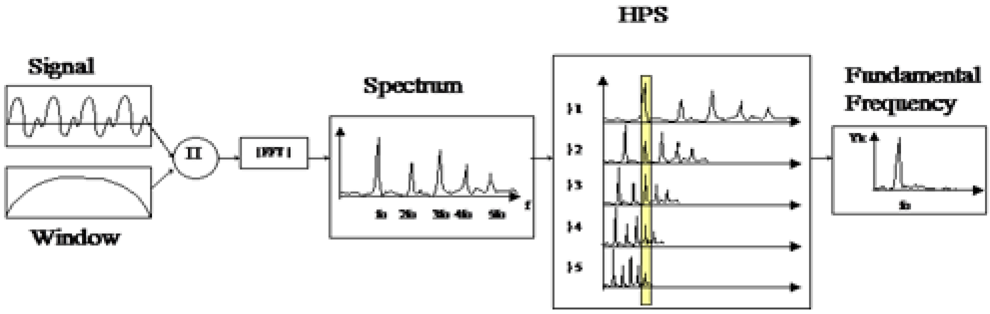

Figure 3: (b) Harmonic Product Spectrum Implementation [10]

3. Hybrid Domain Detection

A group which utilizes both the time- and frequency-domain properties of speech signals. The class of hybrid pitch detectors incorporates features of both the time-domain and the frequency-domain approaches to pitch detection. For example, a hybrid pitch detector might use frequency-domain techniques to provide a insignificant level time waveform, and then use autocorrelation measurements to estimate the pitch period. A combination of time and frequency domain.

Some methods are:

a) EMD-The Empirical Mode Decomposition (EMD) techniques is basically focused on the adaptive time-frequency domain technique. It has been proved quite versatile in a broad range of applications for extracting signals from data generated in noisy nonlinear and non-stationary processes. As useful as EMD proved to be, it ties to remove the difficulties unresolved. The limitation of the EMD is the mode Mixing , which is defined as a single Intrinsic Mode Function (IMF) IMF consist of signals of different scales or same scale that occupy in different IMF components. Scale separation i.e. mode mixing is often a consequence of signal intermittency.

b) EEMD- This technique is used to remove the scale separation problem without introducing a subjective intermittence test, a new noise-assisted data analysis (NADA) method is proposed, the Ensemble EMD (EEMD), which defines the true IMF components as the mean of an ensemble (The ensemble mean is treated as the true answer) of trials, each consisting of the signal plus a white noise of finite amplitude. This new approach is based on the intuition obtain from recent studies of the analytical properties of white noise, which showed that the EMD is effectively an adaptive dyadic filter Banka when applied to white noise The principle of the EEMD is simple: the added white noise would populate the whole time-frequency space uniformly with the constituting components of different scales. When signal is added to this uniformly distributed white background, the bits of signal of different scales are automatically projected onto proper scales of reference established by the white noise in the background definitely; each individual trial may produce very noisy results, for each of the noise- added decay consists of the signal and the added white noise. Since the noise in each trial is different in disparate trials, it is canceled out in the ensemble mean of enough trials.

4. Conclusion

Pitch is that auditory attribute of sound according to which sounds can be ordered on a scale from low to high. In this review paper, we have discussed about some of the factors are influencing the fundamental frequency are such as body size, laryngeal size, vocal fold length and pitch detection techniques. These pitch detection techniques is further classified into time frequency and hybrid detection domain.. The current investigation examines a method that parameterizes pitch in discourse as musical pitch interval *H* measured in units of cents and that disaggregates the sequence of peak word-pitches using tools employed in time-series analysis and digital signal processing. Before these imagined scenarios come true, a very large amount of work done in both researching process and implementation. Futuristic trends in pitch are towards the quantitative pitch distribution in spontaneous events.

References

[1] A very complete definition of pitch by Ernst Terhardt (terhardt@ei.tum.de) could be found in: http://www.mmk.e-technik.tu-muenchen.de/persons/ter/top/defpitch.html
[2] W. Hess, *"Pitch Determination of Speech Signals,"* Berlin: Springer Verlag, 1983.

[3] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal, *"A Comparative Performance Study of Several Pitch Detection Algorithms*," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 24, no. 5, pp. 399-418, 1976.
[4] T. V. Ananthapadmanabha and B. Yegnanarayana, *"Epoch Extraction of Voiced Speech*," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 23, no. 6, pp. 562-570, 1975.
[5] Y. M. Cheng and D. O'Shaughnessy, "*Automatic and Reliable Estimation of Glottal Closure Instant and Period*," IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. 37, no. 12, pp. 1805-1815, 1989.
[6] H. W. Strube, "*Determination of the Instant of Glottal Closure From the Speech Wave*," Journal of the Acoustical Society of America, vol. 56, no. 5, pp. 1625-1629, 1974.
[7] F. J. Harris, "On *the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform*." Proceedings of the IEEE, vol. 66, no. 1, pp. 51-84, 1978.
[8] A. H. Nuttall, "*Some Windows With Very Good Sidelobe Behavior*," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 1, pp. 84-91, 1981.
[9] A. M. Noll, *"Cepstrum pitch determination*," J. Acoust. Soc. Amer., vol. 41, pp. 293-309, Feb. 1967.
[10] Many attempts have been done in this line, see for instance: http://mtg.upf.edu/biblio/author/Serra
[11] Book by Schroeder, 1968
[12] The idea was first proposed by Noll ,"Cepstrum pitch determination", 1967.
[13] Hui-Ling Lu, *"A Hybrid Fundamental Frequency Estimator for Singing Voice"*, center for computer research in music and acoustics(CCRMA),Aug1999.
[14] Lawrence R. Rabiner, Ronald W,"Introduction to Digital Speech Processing", First edition, 1978S