IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

Methods for Fundamental Frequency Extraction of Speech

Jyothi K, Dr. Hari R, Athira Shanmugan A

Department of Electronics and Communication Engineering,Government engineering College Wayanad, Kerala, India

ABSTRACT: A fundamental frequency extraction algorithm is an algorithm designed to estimate the fundamental frequency of oscillating signal, usually a speech or a musical tone. This can be done in both frequency and time domain or independently. Extraction of fundamental frequency from degraded speech is applicable in speech synthesis,voice conversion,speech coding, speaker recognition,gender recognition and speech recognition.

The range of fundamental frequency of speech is from 40Hz to 600Hz. There are frequency domain approach and temporal domain approach for calculating the fundamental frequency. In frequency domain the frequency spectrum of signal is obtained by using FFT(Fast Fourier Transform). In time domain the autocorrelation and spectral information from frequency domain are utilized.

Fundamental frequency estimation is a popular topic in many fields. There are several methods used for the fundamental frequency extraction.

KEYWORDS: Fundamental frequency, Autocorrelation, Fast Fourier Transform

I. INTRODUCTION

Fundamental frequency (f_0) estimation has been a popular research topic for many years. This topic coming under Acoustics, Speech and Signal Processing. In many areas the estimation of fundamental frequency is also known as pitch estimation. Since fundamental frequency is a low frequency it is very difficult to compute. Fundamental frequency is inverse of pitch period. The estimation of f_0 will be affected by degradations in speech signal.

There are several methods used for the computation of f_0. The methods are Subharmonic-to-Harmonic Ratio (SHR), A Pitch Estimation Filter (PEFAC), Sawtooth Waveform Inspired Pitch Estimator (SWIPE), YIN method and Single Frequency Filtering (SFF). In SHR the ratio of magnitude of subharmonic to the harmonic is calculated. In PEFAC the normalized and the signal is then passed through a filter. The filter output is used to calculate the f_0. In SWIPE method the frequency of maximum peak to valley distance of the signal is computed and weighted. YIN method uses normalized square difference of speech signal and its shifted version to compute f_0. SFF considers high signal to noise ratio (SNR) components of the signal at some frequencies.

Estimation of fundamental frequency is difficult since choosing a f_0 estimator is a difficult task. Some detectors works well for music and less for speech. So we have to design a detector that work well for both music and speech. Thus the f_0 estimation becomes more reliable.

The male voice have a fundamental frequency range from 85Hz to 180Hz. The fundamental frequency of female voice is more as compared to male voice and is in the range of 165Hz to 255Hz. The frequency range of clean speech signal is 300Hz to 4000Hz. Telephone speech signal has a frequency range from 80Hz to 14KHz and the cellphone speech has frequencies in MHz range.

There are many methods for computing the fundamental frequency in which each of them has its own advantages and disadvantages. The different methods are compared by using a comparison parameter called Gross Error (GE). Gross error is the percentage of deviation of estimated f_0 to the reference f_0. This parameter is computed for distant speech signals and for different noises. The SFF method gives better performance as compared to the other methods.

II. DIFFERENT METHODS

A. *Subharmonic to Harmonic Ratio (SHR):*

This method is used to estimate the fundamental frequency of a speech signal. Signal whose frequency is an integral multiple of frequency of reference signal is known as har- monics of a signal. Ratio of magnitude of

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

subharmonic and the harmonic components of a frequency is the SHR.When the ratio is small, the observed pitch remains the same. As the ratio increases above certain threshold, the subharmonics become clearly visible on the spectrum. Subharmonic is the midpoints between the harmonics. When the ratio is smaller than 0.2, the subharmonics do not have effects on observed fundamental frequency. As the ratio increases above 0.4, the f_0 is observed lower than that of the lowest subharmonic frequency. When SHR is between 0.2 and 0.4, the pitch lacks clarity. SHR is used to estimate the voice quality. Let A(f) represent the amplitude spectrum, and let f_0, and fmax be the fundamental frequency and the maximum frequency of A(f), respectively. The sum of harmonic amplitude is defined as

$$SH = \sum_{1}^{N} A(nf_0) \quad (1)$$

where N is the maximum number of harmonics contained in the spectrum, A(f) = 0 if f > fmax and n varies from 1 to N. The sum of subharmonic amplitude is given by

$$SS = \sum A((n\text{-}1/2)f_0) \quad (2)$$

The SHR is the ratio of SS and SH. The SHR [1] valueis calculated in logarithmic scale so we transform the linear frequency to logarithmic scale

$$SHR = SS/SH \quad (3)$$

The advantage of SHR method is that this method effectively reduces the gross error rate resulting from subharmonics. The disadvantage is that the average gross error is higher for cellphone, telephone and clean speech signals. SHR is usedin applications such as voice quality estimation.

B. Pitch Estimation Filter (PEFAC)

Here the extraction of fundamental frequency is through filtering. So it is not applicable for telephone signals sinceit is a band limited signal. The pitch estimation in PEFAC
[2] can be done in three ways. First in time domain wherethe peaks in the autocorrelation function is calculated. In frequency domain,the harmonic peaks in the power spectrumis estimated. Third method is the time- frequency domain analysis on the outputs of a bank of bandpass filter. In this method first we estimates the fundamental frequency of each frame. This is done by convolving its power spectral densityin the log-frequency domain with a filter that sums the energy of the pitch harmonics. This will reject the additive noisethat has a smoothly varying power spectrum. Before filtering, amplitude compression is done to attenuate narrowband noise components. The location of peak in the filter output is estimated. This method perform well for non uniform noises and poor performance in clean speech signals. In application such as speech recognition this method is employed.

C. Sawtooth Waveform Inspired Pitch Estimator (SWIPE)

In this method the maximum average peak-to-valley dis- tance across the signal harmonics is calculated and it is used to derive the weight function based on frequency . In the harmonic spectrum the amplitude decay is inversely proportional to frequency. Thus the determination of the pitch of a sawtooth waveform is not easy because its spectrum has more than one component. The pitch of a sawtooth waveform
[3] corresponds to its fundamental frequency. The fundamental frequency in this situation is the component with the highest energy. Thus the pitch will be the largest peak in the spectrum.This method overcomes the errors in the clean speech and poorperformance in cellphone and telephone signals. In speech andmusic processing this method is used.

D. YIN Method

This method is based on the autocorrelation method with several modifications and which prevent errors. The algorithmhas the feature that the error rates three times lower than the other methods, as evaluated with some signals. This algorithm is suited for high-pitched voices and music since there is no upper limit on the frequency. The algorithm is simple and can be implemented efficiently with low latency, and it involves few parameters for tuning. It is based on a periodic signal that may be modified in different ways to handle various forms of

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

aperiodic signals that necessary in different applications. The first step in this method is to find the autocorrelation of the discrete signal. The second step is to evaluate the difference function. This choose a higher-order peak with low error. The difference function is immune to problems, such as changesin amplitude causes period-to-period dissimilarities. In third step the cumulative mean normalized difference function is calculated. Then the absolute threshold is evaluated and finallyobtains the best estimation. YIN method [4] performes well inclean speech signal and poor performance in non uniformly distributed noises. In speech coding and music information retrieval, this method is used.

III. SINGLE FREQUENCY FILTERING (SFF)

This method is based on single frequency filtering (SFF) [5], where the amplitude envelope of the signal is obtained at each frequency with high time resolution and frequency resolution. This results higher signal-to-noise ratio (SNR) regions in time and frequency. The amplitude envelope at a particular frequency can be used to extract the pitch period information. The most robust envelope across different frequencies can be selected for f_0 extraction. Since the envelope is calculated at a single frequency, the fundamental frequency information is notaffected by the vocal tract resonance. There are mainly four steps involved in this method shown in fig.1. The steps are computation of envelopes, noise compensation, determination of dominant frequency and fundamental frequency estimation.

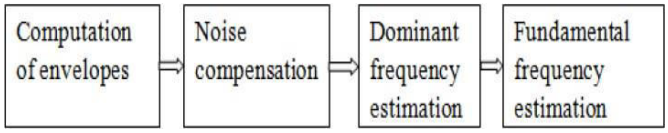

Fig. 1: Single frequency filtering method for f_0 estimation

A. *Computation of Envelopes*

The discrete-time speech signal s(n) is differenced, and the differenced signal is denoted by x(n). The sampling frequency is f_s. The signal x(n) is multiplied by a complex sinusoidal signal with a normalized frequency [6]. The resulting operationin the time domain is given by,

$$x_k = x(n)exp(-jw_k n) \tag{4}$$

where

$$w_k = 2\pi f_k / f_s \tag{5}$$

Since the input signal is multiplied with the exponential sequence, the resulting spectrum will be the shifted version of input signal. This signal is then passed through a single pole filter and the transfer function of the filter is given by,

$$H(z) = 1/1 + rz^{-1} \tag{6}$$

The single-pole filter has a pole on the real axis at a distanceof r from the origin. The the location of the root is at z =-rin the z-plane, which is equal to $f_s/2$. The output is given by,

$$y_k(n) = -ry_{n-1} + x_n \tag{7}$$

and the envelope of the output is,

$$e_k(n) = (y^2_{kr} + y^2_{ki})^{1/2} \tag{8}$$

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

The envelope is obtained from the real and imaginary partsof the output. The pole of the filter should be lie inside theunit circle for the system to be stable [7]. Fig.2.shows the envelopes for different frequencies. At r=0.95 the spectrum gives high temporal resolution and at r=0.995 spectrum gives high frequency resolution. Fig.2.d gives higher frequency resolution for r = 0.995 and fig.2.e shows lower frequency resolution for r = 0.95.

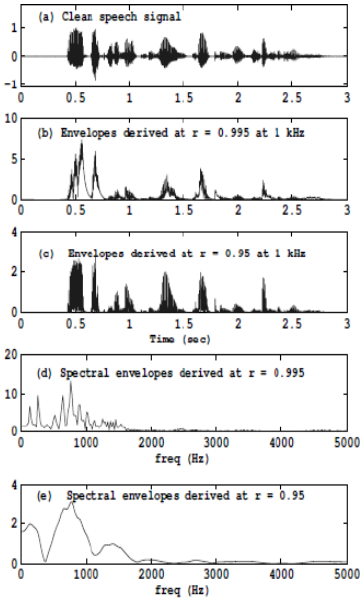

Fig. 2: Envelopes for different poles

B. Noise Compensation

To compensate for the effect of noise, a weight value w_k at each frequency is computed. This reduces the degradations in the signal caused by noises [8]. The weight value is then multiplied with the envelope,given by

$$c_k(n) = w_k e_k(n) \tag{9}$$

Figure 3 shows the noise compensation of clean signal and noise signal. Multiplying the envelope with a weight factor will reduce the effect of noise.In fig.3,(a) shows clean speech signal, (b) is the speech signal corrupted by noise, (c) is the envelopes as a function of time, (d) is corresponding weightedenvelopes and (e) shows envelopes as a function of time for clean speech shown in (a).

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

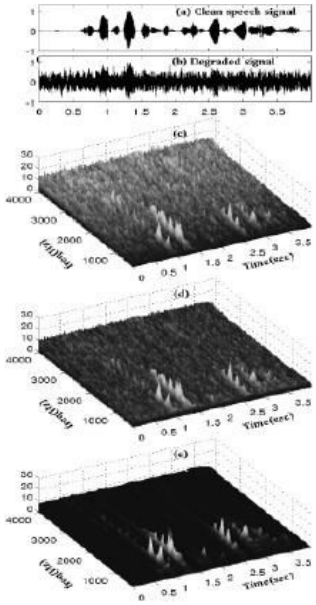

Fig. 3: Noise compensation

C. *Dominant Frequency Estimation*

Here the sequence is autocorrelated and this signal is nor- malized. The dominant peak in the normalized autocorrelationsequence for each frequency is calculated to extract the correctlocation of pitch period for the signal. If the estimated pitch value is same as that of the autocorrelated clean speech signal, the corresponding frequency is taken as the dominant frequency F_D

D. *Fundamental Frequency Estimation*

The autocorrelation sequence of each segment is normalizedby dividing the energy of the segment. The location of the maximum peak in the normalized autocorrelation sequence is taken as the pitch period and its inverse is the fundamental frequency (f_0). The fundamental frequency values for voiced regions are mostly in the range of 300 - 1200 Hz. Hence the envelopes are computed in the 300 - 1200 Hz.

The parameter used to evaluate the performance of different methods is the Gross Error (GE) parameter. It is the percentage of frames whose estimated f_0 values deviate from the reference f_0 values by more than 20 percent. The scores are given in percentage error. Performance of the f_0 estimation methods is also evaluated using the measure of average GE obtained by averaging the scores of the GE across low and high SNR regions. This parameter gives the overall performance of a method without reference to the level of degradation. SFF method shows improvement in comparison with other methods[9]. GE value for different types of noises and distant speech can be done for performance evaluation of different methods over SFF method.

IV. CONCLUSION

Single Frequency Filtering (SFF) output contains segments Single of high signal-to-noise ratio (SNR) components which are utilized for the extraction of fundamental frequency from degraded speech. To obtain good temporal resolution of the envelope at a given frequency, the bandwidth of the single frequency filter should be varied . The method utilizes the speechat high SNR regions of different frequencies and at different times. The variance of speech across frequency is higher than that for noise, after compensating for spectral characteristics for noise. The spectral characteristics of noise are computedby the SFF approach. SFF method shows improvement incomparison with other methods for different types of noises in speech, including the distant speech [10]. The method does not involve estimation of noise characteristics. It may be possible to improve the performance of the SFF method by using suitable post processing techniques and filtering is done using a pole close to the unit circle in the Z plane.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207207 |

REFERENCES

[1] Xuejing Sun, "A pitch determination algorithm based on subharmonic-to-harmonic ratio", Department of Communication Sciences and Disorders,Northwestern University 2299 n. Campus Dr. Evanston, IL 60208, USA

[2] Manish P. Kesarkar and Professor. Preeti Rao,"Feature Extraction ForSpeech Recognition",Electronic Systems Group, EE. Dept, IIT Bom-bay,November2003

[3] Arturo Camacho "SWIPE: A Sawtooth Waveform Inspired Pitch Estima-tor for Speech and Music", University of Florida 2007

[4] Mirza A. F. M. Rashidul Hasan," Correlation Based Fundamental Fre-quency Extraction Method in Noisy Speech Signal", IJCSEIT, Vol.7,No.1, February 2017

[5] Vishala Pannala, G.Aneeja, Sudarsana Reddy Kadiri,andB.Yegnanarayana,"Robust Estimation of Fundamental Frequency
using Single Frequency Filtering Approach",INTERSPEECH 2016 September 8–12, 2016, San Francisco, USA

[6] Raju Alluri K.N, R.K.Sivananda, Sudarsanareddy, Suryakanth V,Gangashetty, Anil Kumar Vuppala," Detection of Repla Attacks using Single Frequency Filtering Cepstral Coefficients",Interspeech 2017

[7] G. Aneeja and B. Yagnanarayana," Extraction of fundamental frequency from degraded speech using temporal envelopes at high SNR fre-quencies", IEEE/ACM Trans.on Audio,Speech,and Language Process-ing,pp.2329,2017

[8] G. Aneeja and B. Yegnanarayana, "Single frequency filtering approach for discriminating speech and nonspeech"IEEE/ACM Trans. on Audio,Speech, Language Process., vol. 23, no. 4, pp. 705–717, April 2015.

[9] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg and C. A. McGonegal,"A comparative performance study of several pitch detection algo-rithms",IEEE Trans. Acoust., Speech and Signal Process,vol. 24, no. 5,pp. 399-418, 1976.

[10] David Gerhard,"Pitch Extraction and Fundamental Frequency- Historyand Current Techniques",Technical Report TR-CS,2003-06 November,2003

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details