[International Journal of Computer and Communication Technology](https://www.interscience.in/ijcct)

Volume 5 | Issue 3 Article 4

July 2014

Efficient Method of Pitch Estimation for Speech Signal Using MATLAB

G. Gandhimathi
Department of Electronics and Communication Engineering, Periyar Maniammai University, Vallam-613403, ggandhimathi@gmail.com

S. Jayakumar
.School of Computing Science and Engineering, Periyar Maniammai University, Vallam-613403, sjayakumar@gmail.com

Recommended Citation

Gandhimathi, G. and Jayakumar, S. (2014) "Efficient Method of Pitch Estimation for Speech Signal Using MATLAB," *International Journal of Computer and Communication Technology*: Vol. 5 : Iss. 3 , Article 4.
DOI: 10.47893/IJCCT.2014.1239
Available at: https://www.interscience.in/ijcct/vol5/iss3/4

Efficient Method of Pitch Estimation for Speech Signal Using MATLAB

G. Gandhimathi [1] & S. Jayakumar [2]

[1]Department of Electronics and Communication Engineering, .
[2]School of Computing Science and Engineering, Periyar Maniammai University, Vallam-613403

Abstract - In this paper, we are estimating the pitch of telephone speech signal. We use different types of methods such us, Burg, Covariance, Fast Fourier transform, Modified Covariance, Multiple Signal Classification (MUSIC) algorithm or Eigen vector, Multi Taper method (MTM), Welch, and Yule Auto Regressive (Yule AR) , to estimate the PSD using signal processing tool box of MATLAB. The spectrum was constructed and the pitch ,amplitude , and slope of the speech signal were calculated and their performance were analyzed.

I. INTRODUCTION

1.1 Speech Spectrum Analysis

Generally the human speech spectrum is less than 4000Hz. According to Nyquist theory, the minimum sampling rate for speech should be 8000samples/second. Due to our system is voice-controlled safety system; it is very helpful to analyze the speaker's voice before our actual design.

Our design is based on the recorder program installed in Windows XP and FFT function in MATLAB. After we speak one word, the recorder program will store the word in a .wav file. Notice this file is sampled at 16000 samples/second, 16bit/sample, so we need to convert it into 8000samples/second, 8bits/sample.

1.2 Power spectral Density (PSD) Estimation

The various methods such us, Nonparametric methods , Parametric methods, and Subspace methods of spectrum estimation [1], available in the Signal Processing Toolbox

1.2.1 Non Parametric Method (NPM)

Nonparametric methods are those in which the PSD is estimated directly from the signal itself. The simplest such method is the periodogram. An improved version of the periodogram is Welch's method [2]. A more modern nonparametric technique is the Multi Taper Method (MTM) [5].

Mathematical Model:

The periodogram estimate of the PSD of a length-L signal $X_L[n]$ is

$$\hat{P}_{xx}(f) = \frac{|XL(f)|}{fsL} \qquad (1)$$

Where

$$X_L(f) = \sum_{n=0}^{L-1} x(n)\; e^{-2\pi j f n/f} \qquad (2)$$

1.2.2 Parametric Method (PM)

Parametric methods are those in which the PSD is estimated from a signal that is assumed to be output of a linear system driven by white noise. Examples are the Yule-Walker autoregressive (AR) method and the Burg method. These methods estimate the PSD by first estimating the parameters (coefficients) of the linear system that hypothetically "generates" the signal

Mathematical Model :

All AR methods [6] yield a PSD estimate given by

$$\hat{P}_{AR}(f) = \frac{\epsilon p}{\left| 1 + \sum_{k=1}^{p} ap(k)\; e^{-\frac{2\pi jkf}{fs}} \right|} \qquad (3)$$

The different AR methods estimate the AR parameters ap(k) slightly differently, yielding different PSD estimates . The Yule-Walker AR method produces the same results as a maximum entropy estimator. The Yule-Walker equations can be solved efficiently via Levinson's algorithm, which takes advantage of the Toeplitz structure of the autocorrelation matrix. The

Burg method for AR spectral estimation is based on minimizing the forward and backward prediction errors while satisfying the Levinson-Durbin recursion [6] . The Burg method avoids calculating the autocorrelation function, and instead estimates the reflection coefficients directly. The Covariance method for AR spectral estimation is based on minimizing the forward prediction error. The Modified covariance method is based on minimizing the forward and backward prediction errors.

1.2.3 Subspace methods (SM)

Subspace methods, also known as high-resolution methods or super-resolution methods, generate frequency component estimates for a signal based on an eigen analysis or eigen decomposition of the correlation matrix [3] . Examples are the Multiple Signal Classification (MUSIC) method and the Eigen Vector (EV) method.

Mathematical Model :

$$\hat{P}_{music}(f) = \frac{1}{e(f)H\ \left(\sum_{k=p+1}^{N} Vk\ Vk\ H\right) e(f)} \qquad (4)$$

$$= \frac{1}{\sum_{k=p+1}^{N} |\ Vk\ H\ .e(f)|2} \qquad (5)$$

where N is the size of the eigenvectors and e(f) is a vector of complex sinusoids

$$e(f) = [\,1 \quad \exp(j2\pi f) \quad \exp(j2\pi f.2) \quad \exp(j2\pi f.4)\ .\ .\ .\ .\ .\ .\ .\ .\ .\ \exp(j2\pi(n-1))\,] \qquad (6)$$

The EV method weights the summation by the eigen values of the correlation matrix:

$$\hat{P}_{ev}(f) = \frac{1}{\left(\sum_{k=p+1}^{N} |\ Vk\ H\ .e(f)|2\right)/\lambda k} \qquad (7)$$

II. PROPOSED METHOD

In this paper we record the speech signal the in .wav file , then treat the problem by importing the file in to the signal processing toolbox of MATLAB . different methods are implemented to estimate the power spectral density of the speech, analysis the performance of methods and the signal is then reconstructed by taking IFFT, as shown in Fig 1.

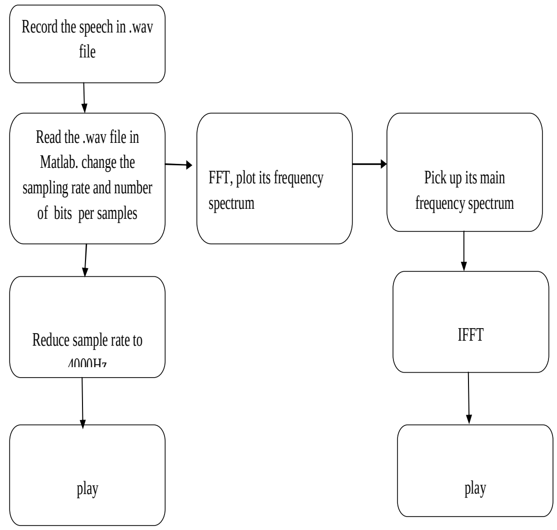

Fig. 1 : Speech Spectrum Analysis using MATLAB

III. DESIGN AND IMPLEMENTATION

3.1 Filter Design

From the above analysis result, we select the sampling rate in our system is 4000 sample /second, 8bits/sample. The cutoff frequency for LPF and HPT is 50Hz, 1500Hz respectively. In order to get the accurate fingerprint of the code, we use seven filters, their working range are: LPF: [0-50Hz] , BPF_1: [50-350Hz] ,BPF_2: [350-500Hz] , BPF_3: [500-750Hz] , BPF_4: [750-1000Hz] , BPF_5: [1000-1500Hz] , HPF: [> 1500Hz]

For the LPF and HPF filter, we take Bruce's sample code as a reference. However, we made a little change. The fingerprint of the speech is the accumulation of the square of the output of each filter. So we combine the calculation the square and accumulation in one filter function. For the fourth-order BPF. We duplicate the second-order filter but using different coefficients. After finishing our code, first, using an Impulse sequence to test it and compare the result with Matlab. This case is to test whether our filter correction is correct. Here we used sample impulse sequence xn=[16 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] . Finally, we tested the running time of our filter. Table1 shows the result.

Table 1 : Performance of the Filter

Types of the filter	**Running Time (cycles)**
LPF	213
HPF	213
BPF*5	395*5
Total	2401
System Requirement	4000

The sample interval is 1/4000*16M = 4000 cycles, which is much longer than processing time of all filters. So our filter design can meet the real time requirement of speech processing.

3.2 Spectrum of PSD

We record the sentence as 'power spectral density' in female voice, in the Matlab environment as a .wav file, convert this as an audio data in a vector form and import this file in to signal processing toolbox speech_signal_ IN is shown in SPTool startup.spt browser , shown in fig. 2.

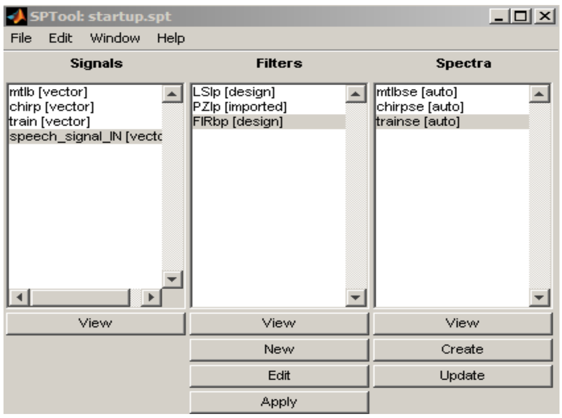

Fig. 2 : SPTool startup.spt browser

The speech signal is processed by the FIR band pass filter. The signal is transformed in to the frequency domain by taking 1024 point FFT .

3.2.1 Spectrum of Non Parametric Method for PSD

We use , non parametric methods such us, Welch , Multi Taper method (MTM) to estimate the PSD . The corresponding spectrum is viewed by the spectrum viewer. This is shown in Fig 3, and Fig 4.

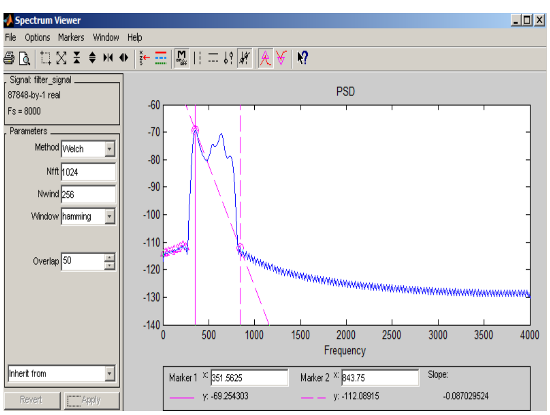

Fig. 3 : PSD using WELCH

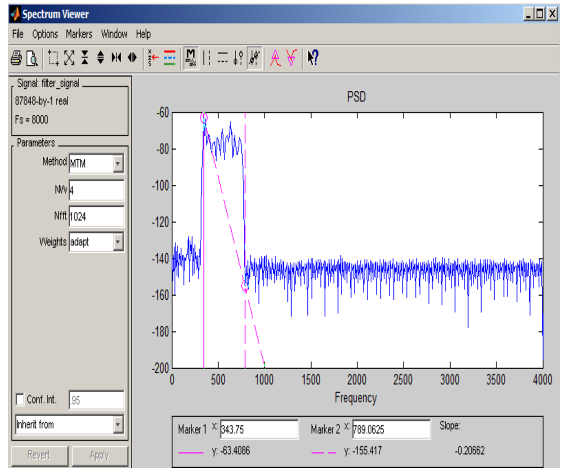

Fig. 4 : PSD using MTM

From the fig 3, and 4 , the span of pitch is almost similar in both WELCH and MTM but slope is steep in MTM

3.2.2 Spectrum of Parametric method for PSD

In parametric methods , Yule Auto Regressive (Yule AR) , Burg, Covariance, Modified Covariance, Multiple Signal Classification (MUSIC) algorithm and Eigen vector, to estimate the PSD and the Corresponding spectrum is viewed by the spectrum viewer. This is shown in Fig 5, Fig 6, Fig 7, and Fig 8.

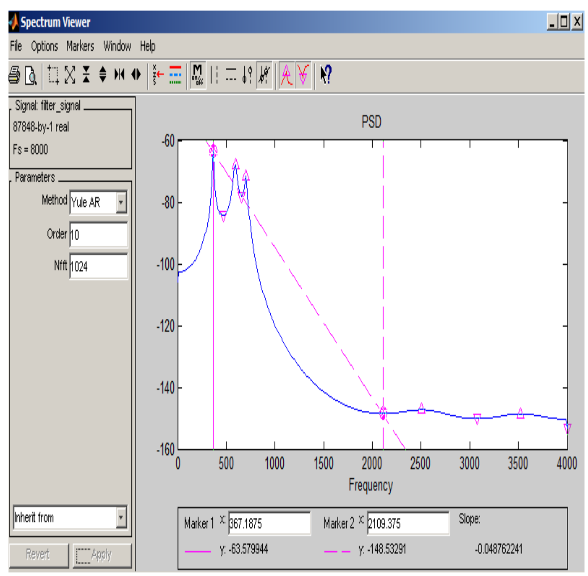

Fig. 5 : PSD using Yule AR

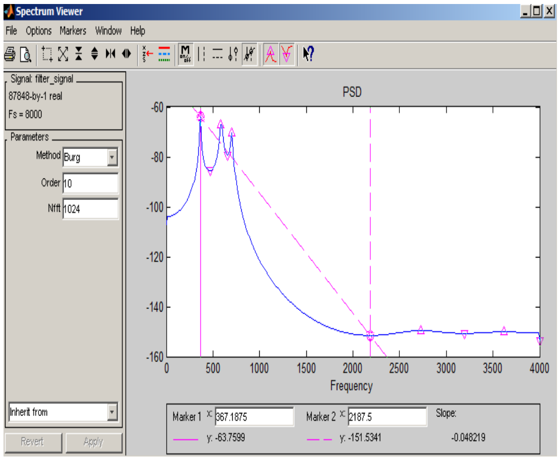

Fig. 6 : PSD using Burg

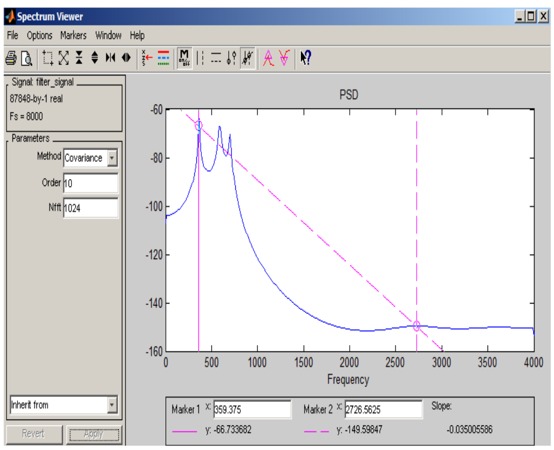

Fig. 7 : PSD using Covariance

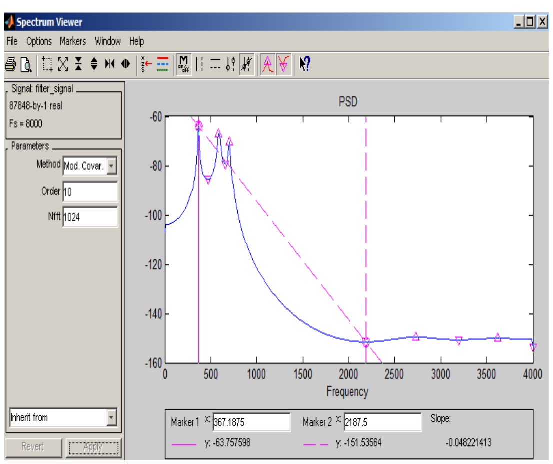

Fig. 8 : PSD using Modified covariance

In nonparametric, the span of pitch and slope is improved .

3.3.3. Spectrum of Subspace Method (SM) for PSD

In this method in fig 9 the amplitude of the pitch attains the maximum value in decibel is (-36.89 7959db) compared to NPM (-63.714286 db) and PM (-63.469388db)

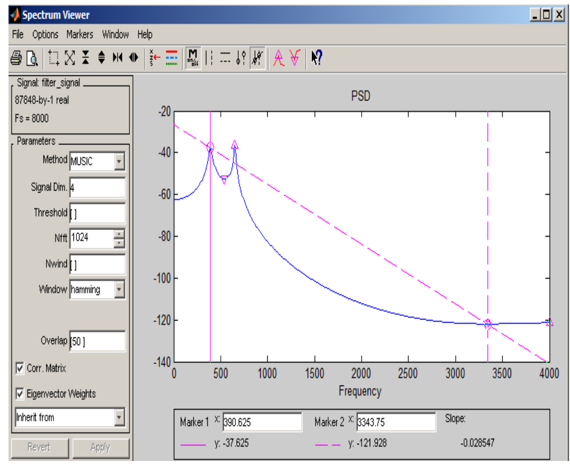

Fig 9 PSD using MUSIC

IV. RESULT AND DISCUSSION

The final view of the SPTool startup.spt browser window is shown in Fig 10. The spectral of all the methods can export in to the work space of Matlab or external disk for further processing .

In Table 2 , the pitch, amplitude of the pitch, and slope for each method are tabulated.

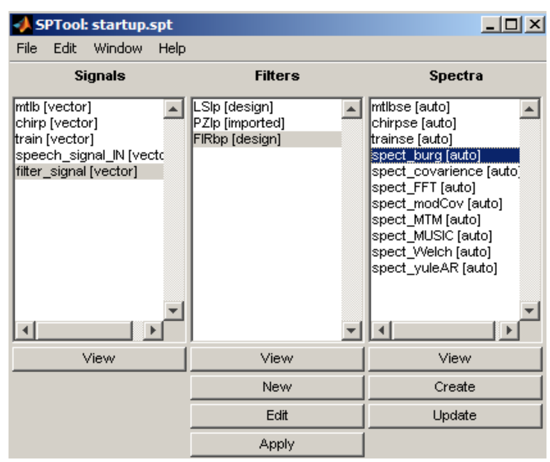

Fig. 10 : Final SPTool startup.spt browser

Table 2 : Spectral analysis in different methods of PSD.

Method	Type	Pitch (Hz)	Amplitude of the pitch (db)	slope
Non Parametric	WELCH	351.5625	-69.306122	-0.087029524
	MTM	343.75	-63.714286	-0.20662
Parametric	Yule AR	367.1875	-63.469388	-0.048762241
	Burg	367.1875	-65.510204	-0.048219
	Covariance	367.1875	-65.102041	-0.035005586
	Modified.cov	367.1875	-64.693878	-0.048221413
Surface	MUSIC	390.625	-36.897959	-0.028547

V. CONCLUSION

For measuring the pitch of the speech signal , the parametric method of estimation gives best performance. They tend to produce better results than classical nonparametric methods when the data length of the available signal is relatively short. The surface methods are best suited for line spectra that is, spectra of sinusoidal signals and are effective in the detection of sinusoids buried in noise, especially when the signal to noise ratios are low.

REFERENCES

[1] Hayes, M.H. Statistical Digital Signal Processing and Modeling. New York: John Wiley & Sons, 1996.

[2] Kay, S.M. Modern Spectral Estimation. Englewood Cliffs, NJ: Prentice Hall, 1988.

[3] Marple, S.L. Digital Spectral Analysis. Englewood Cliffs, NJ: Prentice Hall, 1987.

[4] Orfanidis, S.J. Introduction to Signal Processing. Upper Saddle River, NJ: Prentice Hall, 1996.

[5] Percival, D.B., and A.T. Walden. Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques. Cambridge: Cambridge University Press, 1993.

[6] Proakis, J.G., and D.G. Manolakis. Digital Signal Processing: Principles, Algorithms, and Applications. Englewood Cliffs, NJ: Prentice Hall, 1996.

[7] Stoica, P., and R. Moses. Introduction to Spectral Analysis. Upper Saddle River, NJ: Prentice Hall, 1997.

[8] G. Miet, A. Gerrits, and J. C. Valiere, "Low-band extension of telephone-band speech," Pro. IEEE Int. Conf. On Acoustics, Speech, Signal Processing, vol. 3, pp. 1851-1854, Jun 2000.

[9] S. Chennoukh, A. Gerrits, G. Miet, and R. Sluitjer, "Speech enhancement via frequency bandwidth extension using line spectral frequencies," Pro. IEEE Int. Conf. On Acoustics, Speech, Signal Processing, vol. 1, pp. 665-668, May 2001.

[10] D.A. Heide, G.S. Kang, "Speech enhancement for bandlimited speech", Proceedings of the ICASSP, Vol. 1, Seattle, WA, USA, May 1998, pp. 393-396. 20dB.

◈◈◈