World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

# Spectrum estimation for voiced speech using average weighted linear prediction

Md Arifour Rahman [1,*], Md Masudur Rahman [1] Arifuzzaman Joy [1] and Md. Najmul Hossain [2]

[1] *Department of Electrical and Electronic Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh.*

[2] *Department of Electrical, Electronic and Communication Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh.*

World Journal of Advanced Research and Reviews, 2024, 22(02), 1495–1503

Publication history: Received on 11 April 2024; revised on 18 May 2024; accepted on 21 May 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.2.1569

## Abstract

In many different areas, such as neurophysics, geophysics, and communication, it is vital to effectively handle signals when there is unwanted disturbance. Linear prediction (LP) methods provide precise modeling capabilities for various parametric models and are used in a wide range of fields such as predicting future data, compressing speech, analyzing spectra, filling in missing signals, and decreasing noise. LP is especially beneficial in speech processing systems. Speech analysis and synthesis greatly relies on spectral envelopes, which are closely connected to models of speech production and perception. However, the performance of the LP method may decline when limited information is involved in pitch synchronous analysis. In order to overcome this constraint, we developed the implementation of the Average Weighted Linear Prediction (AWLP) technique, which does not require the use of a window. The AWLP utilizes a signal of weighted prediction error acquired from the autocorrelation process of LP. By performing this action, it provides a more precise estimation of the power spectrum and coefficients used for vocal tract parametric models. Importantly, the effectiveness of the AWLP can be seen in both pitch synchronous and asynchronous analyses, which means it is appropriate for situations where segments are chosen randomly. We confirm the effectiveness of the AWLP technique by employing it on both synthetic voice and real speech. Additionally, we assess its effectiveness by comparing it to the direct autocorrelation technique and a modified autocorrelation (MA) approach. To sum up, the AWLP method presents a hopeful solution for reliable spectral analysis that addresses the divide between pitch synchronous and asynchronous situations.

**Keywords:** LP; Autocorrelation; Pitch Synchronous; Pitch Asynchronous; PEF; WPEF

## 1. Introduction

Speech communication between humans and machines is becoming more common as mobile communication and computers become more advanced. In speech communication system, the representation of the voice signal in a suitable manner and the preservation of the message contents are the major concerns. The acoustic features, *i.e.*, vocal tract shape, formant frequencies, coefficients, power spectrum, bandwidths and pitch are associated with the representation of the speech signal and have deep applications in speech recognition, synthesis and coding [1-3]. Therefore, the transmission, recording and modeling of the speech information with accurately and efficiently is a challenging issue. In the field of neurophysics, geophysics, communication and other areas of applications, analysis of spectral properties of signals are also required [4][5]. Linear Prediction (LP) is a very effective tool that has been used in a wide range of signal processing applications, particularly in speech processing system [4][7]. LP is one of the robust speech analysis techniques and has the ability to model the voiced speech accurately and efficiently by a small set of parameters closely related to the speech production transfer function. To be specific, LP technique can model the vocal tract by an all-pole filter. Accurate and efficient power spectrum estimation of speech signal has a great attention in signal analysis and synthesis because of their close connection with the speech production and perception model [6]. Development of the

* Corresponding author: Md Arifour Rahman

parametric model with accurate estimation is required so that the model can be used in different application, such as prediction, control and data compression.

The basic formulation of the LP analysis seeks to find an optimal fit to the envelop of the speech spectrum. The fit is obtained by solving for the prediction coefficients that minimize the sum of squares of the prediction error. The LP technique suffers from various sources of limitations [4][7][8]. These limitations are mostly manifested during voiced segments of speech. The auto-correlation method and the covariance method, also called stationary and non-stationary respectively [8], are two of the most commonly used LP methods. The Levinson-Durbin algorithm [9][10], which is used in the auto-correlation process, is known for ensuring the stability of the resulting autoregressive (AR) model [8]. The autocorrelation process, which requires less computation, is covered in this paper.

In pitch asynchronous analysis of voiced speech, these two methods (autocorrelation and covariance) show almost same result. In pitch synchronous analysis, the autocorrelation method has a poor efficiency, but it ensures the stability of the approximate all-pole filter. Whereas, the covariance method is more efficient and better assumption than autocorrelation method, but it does not guarantee the stability of the estimated speech production transfer function [8]. Pitch synchronous: The analysis portion is less than or equal to one pitch period data length, and Pitch asynchronous: The analysis portion is greater than one pitch period data length. A modified autocorrelation (MA) method [11] of LP has been established to mitigate this issue in pitch synchronous analysis. Hence, the analysis segment is used as exactly equal to one full pitch period data length. If the analysis segment is less than one full pitch period in pitch synchronous analysis, this modified method degrades its performance. For pitch asynchronous analysis, all of these three methods perform well. As a result, a method in pitch synchronous and asynchronous (both) analysis, *i.e.*, can be called random analysis, of voiced speech that performs better than autocorrelation and modified autocorrelation in pitch synchronous analysis which does not need exactly one full pitch period analysis segment and ensures the stability of the approximate all-pole filter, and also performs well in pitch asynchronous analysis is needed. Random analysis means the analysis segment is selected randomly. One such approach has been suggested in this paper.

LP by autocorrelation also suffers from amplitude distortion in ill-conditioned environment [4]. A pre-emphasis is often used to mitigate the ill-conditioning problem. A number of techniques have been mentioned in [12] to mitigate ill-conditioning problem. The proposed method also helps to mitigate ill-conditioning issue. In this Paper, we present an Average Weighted Linear Prediction (AWLP) method for LP by using autocorrelation method to deal with all of these mentioned issues. This paper is covered with power spectrum, coefficients and spectral bias value B estimation and comparing the AWLP method over direct autocorrelation and MA method.

## 2. Proposed Method

The proposed method uses autocorrelation procedure to estimate better prediction coefficients as well as power spectrum. This AWLP method, basically an extended version of autocorrelation method, performs well for both in pitch synchronous and asynchronous analysis by using a weighted prediction error filter. In this section, the algorithm for the proposed AWLP method is described in detail first. Later, the properties of the proposed method are described.

### 2.1. Algorithm

This section will clarify AWLP algorithm. If we consider a speech signal with analysis segment $N$ as $s(n)$, where $n = 0,1,2,.....,N-1$, firstly we have to calculate the autocorrelation function of $s(n)$ as

$$R_n(m) = \sum_{n=0}^{N-1} s(n)s(n+m) \dots \dots \dots (1)$$

By considering a Toeplitz matrix with a size of $p+1$, we obtain

$$\begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p) \\ R_n(1) & R_n(0) & \dots & R_n(p-1) \\ R_n(2) & R_n(1) & \dots & R_n(p-2) \\ \dots & \dots & & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ R_n(p) & R_n(p-1) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \dots \\ \dots \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(0) \\ R_n(1) \\ R_n(2) \\ \dots \\ \dots \\ \dots \\ R_n(p) \end{bmatrix} \dots \dots \dots (2)$$

where, $p$ is the prediction order, α are the prediction coefficients and $\alpha_0 = 1$ is assumed. Then the obtained solution to find prediction coefficients $\alpha_k$, where $k = 0, 1, 2, ...., p$, is as

$\alpha = R^{-1}r$ … … … (3)

**R** is the $(p+1) \times (p+1)$ Toeplitz matrix, *i.e.*, it is symmetric and all the elements along a given diagonal are equal. We implement Levinson-Durbin recursive algorithm [9][10] to solve this equation and obtained the prediction coefficients $\alpha_k$ .

Secondly, we define a weighted prediction error (WPE) signal $e_w(n)$ by using a weighted prediction error filter (WPEF). The weighted prediction error signal is defined as

$$e_w(n) = s(n) - \mu \sum_{k=0}^{p} \alpha_k \, s(n-k) \text{ … … . . } (4)$$

where, $\mu = 0.12$ and $\alpha_0 = 1$ are assumed. The values of $\alpha_k$ are previously calculated. The transfer function of the given signal $e_w(n)$ can be represented as

$H_{WPEF}(z) = 1 - \mu \sum_{k=0}^{p} \alpha_k \, z^{-k}$ … … … (5)

The prediction error filter (PEF) is realized as a finite impulse response (FIR) filter. The proposed AWLP method weight the prediction error filter and that filter type is similar to that of the pre-emphasis filter, but the filter coefficients are changing here. Each LP coefficient $(\alpha_k)$ is multiplied with a constant value $\mu$ to obtain WPE signal. If we consider $\mu\alpha_k = \beta_k$, then the WPE signal can be written in more compactly as

$$e_w(n) = s(n) - \sum_{k=0}^{p} \beta_k \, s(n-k) \text{ … … … } (6)$$

where, $\beta_0 = -1$ is assumed. So, now the output signal of the WPEF is $e_w(n)$ corresponding to the analysis speech segments$(n)$. Then we have to calculate weighted prediction error coefficients $\beta_k$ using autocorrelation method.

Finally, we have to average the coefficients. The predicted AWLP coefficients are then

$\gamma_k = \frac{\alpha_k + \beta_k}{2}$ … … … (7)

The power spectrum of AWLP method is estimated using $\gamma_k$ as

$$P_{AWLP}(\omega) = \frac{\sigma^2_{AWLP}}{\left|1 + \sum_{k=1}^{p} \gamma_k \, e^{jk\omega}\right|^2} \text{ … … . . } (8)$$

where, $\gamma_k$ are the prediction coefficients and $\sigma^2_{AWLP}$ is prediction error power of AWLP method.

To more simplify the AWLP process:

- Develop a frame for LP analysis from a speech signal.
- Estimate the prediction coefficients from frame using LP by autocorrelation method.
- Estimate a weighted prediction error signal from frame where weighting factor is a constant value $\mu$ = 0.12 multiplied with the prediction coefficients.
- Find out the LP coefficients from weighted prediction error signal by autocorrelation method.
- Then average these two types of coefficients and find the resulting coefficients.
- Finally, estimate the power spectrum using resulting coefficients.

Figure 1 shows the block diagram of power spectrum estimation with AWLP method.

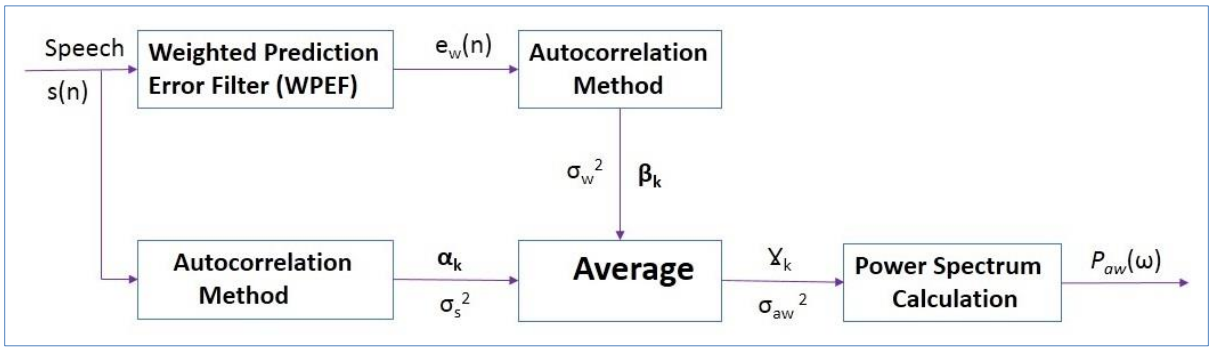


**Figure 1** Power Spectrum Estimation with AWLP Method

### 2.2. Properties

The algorithm of AWLP described in section 2.1 is an extended version of autocorrelation method using weighted prediction error signal. The properties of the proposed AWLP method are given below:

The proposed AWLP method does not use any kind of window function, like Hamming, Hanning, etc., to avoid unwanted signal distortion in autocorrelation process.

It estimates better assumption in both pitch synchronous and asynchronous analysis. The frame length of analysis segment should be less than 25 ms. Because, speech is highly non-stationary signal. Speech information remain constant at 20-25 ms.

It helps to reduce ill-conditioning problem. The method uses a weighted prediction error filter which is as like as a pre-emphasis filter, often use in ill-conditioned environment, where filter coefficients are changed here instead of a constant coefficient.

It is stable and reliable in pitch synchronous analysis.

## 3. Experimental Result

In this section, we show the performance, *i.e.*, power spectrum estimation, prediction coefficients calculation, spectral bias measurement, of the AWLP method with that of the direct autocorrelation method and MA [11] methods. As the true values of spectral parameters of synthetic vowel signals are previously known, we use synthetic vowel signals, generated by the Liljencrants-Fant (LF) model [13], for analysis and accuracy in estimating the power spectrum first, and then investigate these over real speech. This section evaluate the performance of these three methods on the basis of pitch synchronous and asynchronous. For autocorrelation method, *i.e.*, direct method, we use a Hamming window function before its analysis. But any kind of window function was not used for other methods, *i.e.*, modified autocorrelation and AWLP methods.

### 3.1. Synthetic Vowel

By exciting an all-pole filter with a periodic train of pulses, we generate all the synthetic vowel signals. The transfer function of the all-pole filter is

$$H(z) = \frac{G}{1-\sum_{k=1}^{M} a_k z^{-k}} \dots \dots \dots (9)$$

Here, G is the gain function and $a_k$ are the filter coefficients. For our experiment, we assumed G = 0.1354 and the coefficients for different vowels are exhibited in tables. The fundamental frequency ($F_0$) for the source excitation is 80 *Hz*. The vowels are generated with a rate of 10 *kHz* sampling frequency $f_s$. We calculate power spectra using fast Fourier transform (FFT) where FFT point was 1024. Comparison is made by the visual inspection as well as by computing the spectral bias value *B* which is defined as

$$B = \frac{1}{\pi f_s} \int_0^{\pi f_s} \frac{|\hat{P}(\omega) - P(\omega)|}{P(\omega)} d\omega \dots \dots \dots (10)$$

Where $P(\omega)$ and $\hat{P}(\omega)$ denote the true spectrum and estimated power spectrum, respectively. We have illustrated the estimated power spectra from prediction coefficients and prediction error power using FFT. The true power spectra have been obtained by FFT of the filter coefficients $a_k$ which are used to generate synthetic vowel. The LP order for all vowel was 10.

If we consider the data sample of analysis segment is $L$ and the data sample of one full pitch period is $N$, then the pitch synchronous analysis is called for $L \leq N$ and asynchronous analysis is obtained for $L > N$. $N$ is 125 for all vowel and $L$ is selected on the basis of synchronization.

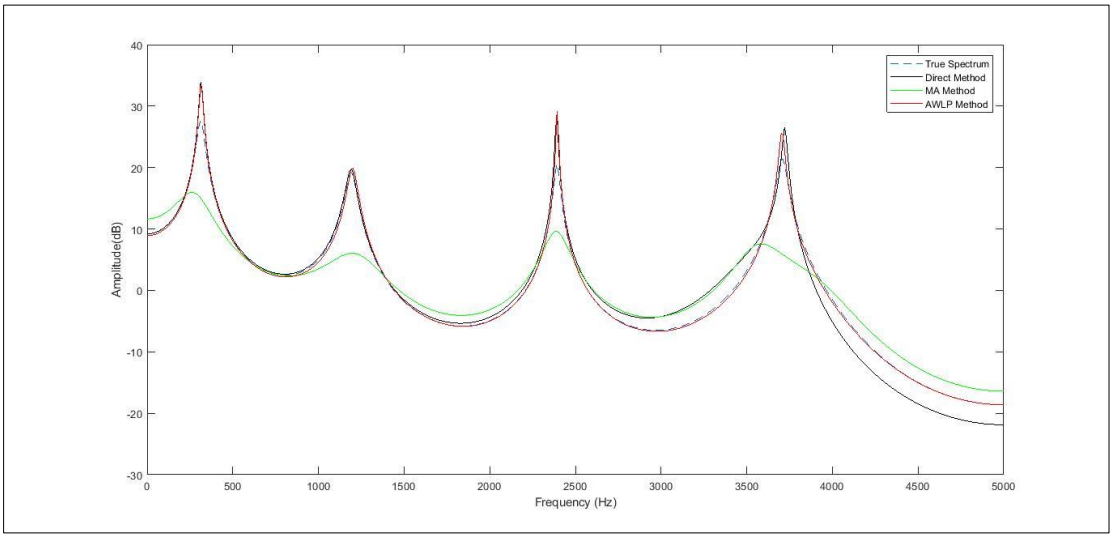


**Figure 2** Power Spectrum Analysis for Pitch Synchronous ($\boldsymbol{L < N}$) of vowel /u/

Power spectra for pitch synchronous and asynchronous of vowel /u/ have been shown in Figure 2 and in Figure 3. In Figures, the true power spectrum has been plotted in dotted line. The black colored spectrum reveals the direct method, the green color is for MA method and the red color marks the proposed AWLP method. In Figure 2, while $L < N$ ($L = 110$), the power spectrum of the proposed AWLP method is very close to the true power spectrum. The MA method losses its stability. The performance of the direct method has also degraded. Figure 3 also shows that the power spectrum for proposed method is very close to the true spectrum. Though visual inspection shows strong evident, we calculated spectral bias $B$ by the formula in Equation 10 shown in Table 1. The estimated coefficients of Figure 3 placed in Table 2 proves more accurate estimation of proposed method. After observing Table 1 and 2, we can strongly say that the proposed AWLP method can estimate more accurate coefficients than other two methods in pitch synchronous analysis.

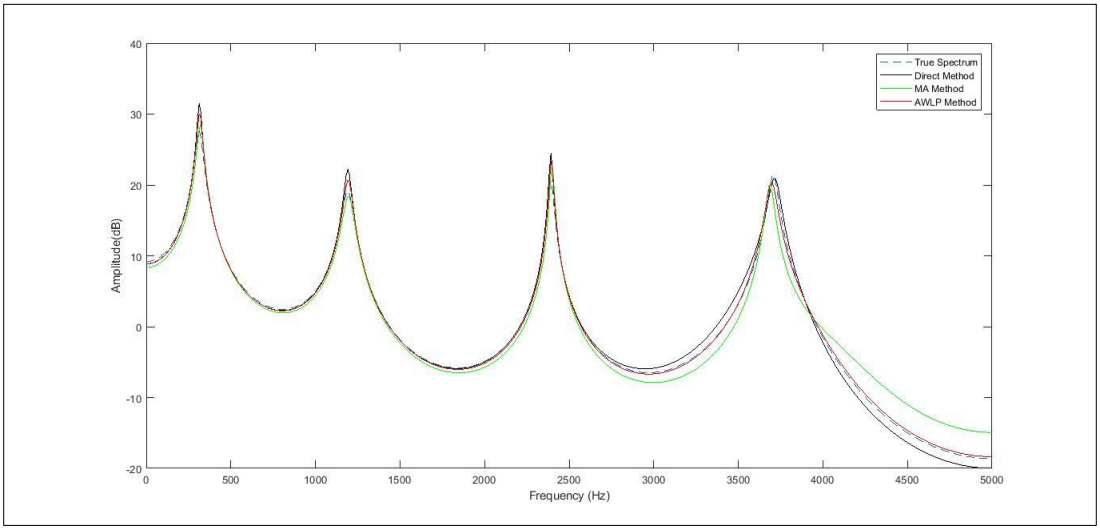


**Figure 3** Power Spectrum Analysis for Pitch Asynchronous ($L > N$) of vowel /u/

**Table 1** Spectral Bias Value B for the vowel /u/ for different Methods in Random Analysis

| Analysis Segment | Direct Method | Modified Autocorrelation Method | AWLP Method |
|---|---|---|---|
| L<N | 0.42 | 0.48 | 0.07 |
| L=N | 0.38 | 0.59 | 0.15 |
| L>N | 0.28 | 0.36 | 0.18 |
| Average | 0.36 | 0.47 | 0.13 |

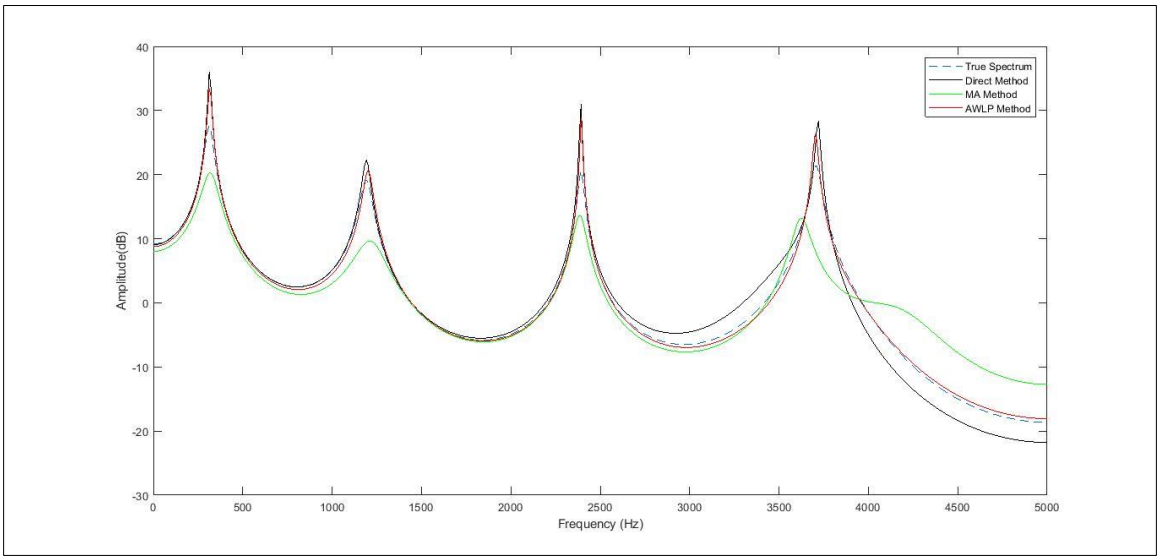


**Figure 4** Power Spectrum Analysis for Pitch Synchronous ($\boldsymbol{L = N}$) of vowel /u/

**Table 2** Coefficients Estimation by Different Methods for Vowel /u/ in Pitch Synchronous Analysis ($\boldsymbol{L < N}$)

| Coefficient No | True Value | Direct Method | Modified Autocorrelation Method | AWLP Method |
|---|---|---|---|---|
| 1 | -0.9912 | -1.1434 | -0.8285 | -1.0039 |
| 2 | 0.5255 | 0.905 | 0.4251 | 0.5435 |
| 3 | -0.9272 | -1.4212 | -0.8319 | -0.9324 |
| 4 | 1.1236 | 1.6173 | 0.7850 | 1.1131 |
| 5 | -1.0991 | -1.5971 | -0.7800 | -1.1065 |
| 6 | 1.1472 | 1.6637 | 0.7763 | 1.1621 |
| 7 | -0.8114 | -1.3211 | -0.4801 | -0.8140 |
| 8 | 0.0587 | 0.5170 | 0.0572 | 0.0367 |
| 9 | -0.2630 | -0.5445 | -0.2543 | -0.2278 |
| 10 | 0.5839 | 0.6701 | 0.3931 | 0.5879 |

The proposed method was also applied on other vowels. And spectral bias value was estimated for the proposed method and the other two methods shown in table 3. It is observable that the average spectral bias value of the proposed method is minimum than the other two methods.

**Table 3** Spectral Bias Value B for five vowels for different Methods in Random Analysis

| Vowel | Direct Method | Modified Autocorrelation Method | AWLP Method |
|---|---|---|---|
| /a/ | 0.32 | 0.62 | 0.21 |
| /e/ | 0.56 | 0.08 | 0.34 |
| /i/ | 0.12 | 0.15 | 0.06 |
| /o/ | 0.44 | 0.39 | 0.28 |
| /u/ | 0.42 | 0.48 | 0.07 |
| Average | 0.37 | 0.34 | 0.19 |

*3.1.1. Real Speech*

In this section, the applications of AWLP method in real speech have been discussed. After getting better results in synthetic vowel experiments, we investigated the performance of the proposed method for both male and female speaker in the real speeches and observed that the proposed method provides better performance than the existing other methods. Some of our experimental results in the real world have been shown in this section. Continuous real speech signals of speakers, which were recorded in a sound-isolated room, were utilized with a sampling frequency of 8 *kHz*. These speeches were recorded using a Sony ECM-J3M microphone.

Figure 5 show the examples of power spectra of real vowel /o/ by male speaker. The true spectra in black colored line are obtained by FFT where the FFT point is 1024. The other spectra show as the direct method in green colored, the MA method in dotted blue colored and the proposed AWLP method in red colored line. The order of LP was 10 in the time of investigation. The proposed method was also applied on the continuous speech pronunciation. Figure 6 shows the power spectrum of continuous speech signal from a randomly selected voiced portion.

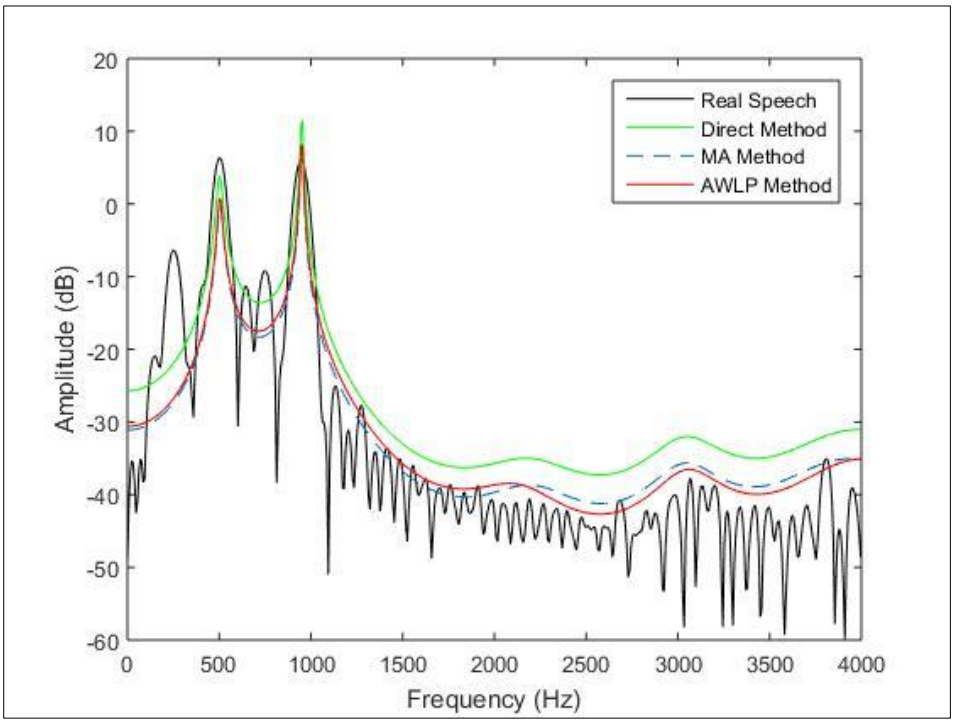


**Figure 5** Power Spectrum of Real Vowel /o/ (Male Speaker)

From the view of Figure 6, it is observed that the spectrum by the AWLP method is commonly closer to that of the true power spectrum obtained by FFT from the original speech. This excellent performance of the estimated AWLP method

is anticipated from the experimental performance of synthetic vowels as earlier section discussed. These results validate that the AWLP method is useful on the real speeches as well.

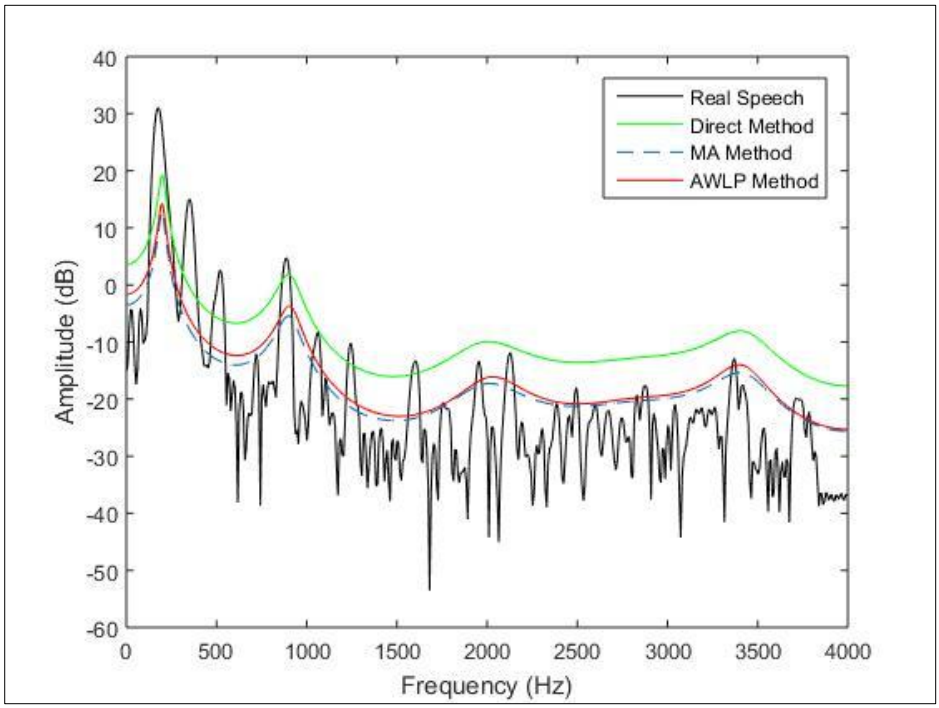


**Figure 5** Power Spectrum of Real Speech (Female Speaker)

## 4. Conclusion

In this research, we suggested an accurate power spectrum and coefficients estimation technique (AWLP) in the autocorrelation domain utilizing a WPEF. It is a fresh use of the widely recognized PEF idea. Here, we have conducted an experimental comparison of the AWLP method's accuracy in expressing the spectrum envelop of synthetic and actual voiced speech with that of direct autocorrelation and the MA approach. The experimental results demonstrate the accuracy and efficiency with which the suggested AWLP approach in pitch synchronous analysis calculates the power spectrum and coefficients. Even in pitch synchronous analysis, when the analysis segment is less than a pitch period, it maintains its stability. Additionally, it functions well in pitch asynchronous analysis. Therefore, if we need to analyze both synchronous and asynchronous pitches simultaneously, this would be the best option.

## Compliance with ethical standards

### *Acknowledgment*

The authors would like to thank the authorities of the Ministry for Science and Technology, Bangladesh for inspiring by selecting as an M.Sc. thesis fellow under Bangabandhu National Science and Technology.

### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Lawrence. R. Rabiner and Ronald W. Schafer, Digital Processing of Speech Signals, Prentice Hall, Inc., Englewood Cliffs, New Jersy 07632, 1978, P-(38-106).

[2] G. Fant, Acoustic Theory of Speech Production, Mouton, The Hague, 1970.

[3] J. L. Flanagan, Speech Analysis,Synthesis, and Perceptions, 2nd ed.1em plus 0.5em minus 0.4emSpringer-Verlag, New York,1972.

[4] J. Makhoul, Lineart prediction : A Tutorial Review, Proc. IEEE, Vol. 63, No. 4, 1975.

[5] L. -C. Wood and S. Treitel, Seismic Signal Processing, Proc. IEEE, Vol. 63, No. 4, pp. 649-661. 1975.

[6] J. Makhoul, Spectral Analysis of Speech by Linear Prediction,, IEEE Trans. Audio Electroacoust., Vol. AU-21, 1973, pp. 140-148.

[7] B. S. Atal and S. Hanauer, Speech analysis and synthesis by linear prediction of the speech wave, J.Acoust. Soc. Amer., Vol. 50, No. 2, 1974, pp. 637-655.

[8] S. Chandra and W. C. Lin, Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis, IEEE Trans. on Acoust., Speech, Signal Processina, Vol. ASSP-22, No. 6, 1974, pp.403-415.

[9] N. Levinson, The wiener RMS (root mean square) error criterion in filter design and prediction, J. Mathematical Physics, vol. 25, no. 4, pp. 261-278, 1947.

[10] J. Durbin, The fitting of time-series models, Int. Statistical Review, vol. 28, no. 3, pp. 233-243, 1960.

[11] K. K. Paliwal and P. V.S. Rao, A modified autocorrelation method of linear prediction for pitch synchronous analysis of voiced speech, Speech and Digital Systems Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India, vol. 3, Issue no. 2, pp. 181-185, 1981.

[12] P. Kabal, Ill-conditioning and bandwidth expansion in linear prediction of speech, Proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing Acoust, 2003.

[13] G. Fant, J. Liljencrants and Q. G. Lin., A four parameter model of glottal flow, Quart. Progress and Status Rep., Speech Transmission Lab, Royal Inst. Technol., 1985, pp. 1-4

## Author's short biography

**Md. Arifour Rahman** was born in Naogaon, the People's Republic of Bangladesh in 1979. He received his B.Sc. and M.Sc. degrees in Applied Physics and Electronic Engineering (Presently named Electrical and Electronic Engineering) from the University of Rajshahi, Bangladesh, in 2003 and 2004, respectively. In 2018, he received his PhD in Digital Signal Processing at the Graduate School of Science and Engineering, Saitama University, Saitama, Japan. His current research interests include Digital Signal Processing, Machine Learning and Artificial Intelligence.

**Md. Masudur Rahman** was born in Rajshahi, the People's Republic of Bangladesh in 1995. He received his B.Sc. and M.Sc. degrees in Electrical and Electronic Engineering from the University of Rajshahi, Bangladesh, in 2017 and 2018, respectively. His current research interests include Digital Signal Processing, Image Processing, and Machine Learning.

**Arifuzzaman Joy,** was born in Dhaka, the People's Republic of Bangladesh in 1997. He received his B.Sc. degrees in Electrical and Electronic Engineering from the University of Rajshahi, Bangladesh, in 2020. His current research interests include Digital Signal Processing, Bioinformatics, Machine Learning, and Artificial Intelligence.

**Md. Najmul Hossain** (Member, IEEE and IEEE ComSoc) was born in Rajshahi, the People's Republic of Bangladesh in 1984. He received his B.Sc. and M.Sc. degrees in Applied Physics and Electronic Engineering (Presently named Electrical and Electronic Engineering) from the University of Rajshahi, Bangladesh, in 2007 and 2008, respectively. In 2020, he received his PhD in Advanced Wireless Communication Systems at the Graduate School of Science and Engineering, Saitama University, Saitama, Japan. His current research interests include computer vision, antenna and wave propagation, advanced wireless communications, and corresponding signal processing, especially for OFDM, orthogonal time frequency space (OTFS), orthogonal chirp division multiplexing (OCDM), MIMO, and future-generation wireless communication networks.