Estimation Of Formant Frequency Of Speech Signal By Linear Prediction Method And Wavelet Transform

[1]D. K. Sahoo, [2]Satyasis Mishra, , [3]G. Panda, [4]P. K. Dash, ,
[1,2] Centurion University, Bhubaneswar, Orissa
[3]IIT, Bhubaneswar,[4]SOA University, Bhubaneswar, Orissa

Abstract

This paper presents a new approach to estimate the formant frequency of the speech signal by using the linear predictive coder(LPC) method and Wavelet Transform. LPC filtering is used to obtain an estimate of vocal tract impulse response which is free from periodicity. Thus linear prediction of the resulting vocal tract impulse response is expected to be free from variations of fundamental frequencies. In this paper a detail study on the prospects of LPC prediction is shown as a formant tracking tool especially for any male speech voice signal to obtain accurate estimation. The solutions obtained by the current method are guaranteed to be stable which makes it superior for many speech analysis applications. Further wavelet transform is applied to denoise the signal and formant frequency of denoised signal is calculated.

***Keywords**: Linear prediction, Autocorrelation, cepstrum, Fundamental frequency effect, wavelet transform.*

1. Introduction

Wavelet Transforms, in particular the Continuous Wavelet Transform (CWT), expand the signal in terms of wavelet functions which are localized in both time and frequency. Thus the Wavelet Transform (WT) of a signal may be represented in terms of both time and frequency. For reconstruction of the signals from its discrete samples the Continuous Wavelet Transform is sampled with Nyquist criteria. In wavelet analysis the use of a fully scalable modulated window solves the signal cutting problem. As the matter of fact, the wavelet series is simply a sampled version of the Continuous Wavelet Transform (CWT), and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy on the other hand, requires a significant amount of computation time and resources.

The Discrete Wavelet Transform (DWT) [7], on the other hand, provides sufficient information both for analysis and synthesis of the original signal with a reduction in computation time. The Continuous Wavelet Transform (CWT) was computed by changing the scale of the analysis window, shifting the window in time, multiplied by the signal, and integrated over all times. In the discrete case, filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies.

It is well known to any scientist and engineer who work with a real world data that signals do not exist without noise, which may be negligible (i.e. high SNR) under certain conditions. However, there are many cases in which the noise corrupts the signals in a significant manner, and it must be removed from the data in order to proceed with further data analysis. The process of noise removal is generally referred to as signal denoising or simply denoising. Although the term "signal denoising" is general, it is usually devoted to the recovery of a digital signal that has been contaminated by additive white Gaussian noise (AWGN), The optimization criterion according to which the performance of a denoising algorithm is measured is usually taken to be mean squared error (MSE)-based, between the original signal (if exists) and its reconstructed version. This common criterion is used mostly due its computational simplicity. Moreover, it usually leads to expressions which can be dealt with analytically. However, this criterion may be inappropriate for some tasks in which the criterion is perceptual quality driven, though perceptual quality assessment itself is a difficult problem, especially in the absence of the original signal.

Formant frequencies are the principal analytical features in speech processing. This is because they are clearly related to the articulator act and the perception of speech . Formant information is used extensively in coding, analysis/synthesis applications, and recognition of speech .Linear predictive analysis [8] is one of the most powerful techniques to extract formant frequencies. The importance of this method lies in its ability to provide accurate estimates and its relative speed of computation The basic formulation of the linear prediction seeks to find an optimal fit to the envelope of the speech spectrum. Since the source of voiced speech is of a quasi-periodic nature with spiky excitations, those impulsive periodic innovations sometimes result in inaccuracy in spectrum estimation, especially, in case of high-pitched speech. In this paper, we briefly illustrate the cause of inaccuracy of formant frequency estimation in case of pitch-asynchronous autocorrelation method and propose a solution based on LPC filtering .In the conventional autocorrelation method when a finite segment is extracted over multiple pitch periods, the obtained autocorrelation sequence is actually an ‘‘aliased’’ version of the true autocorrelation of vocal tract system impulse response. This is because the replica of autocorrelation of vocal tract impulse response is repeated periodically with the periodicity equivalent to pitch period, which overlaps and distorts the underlying autocorrelation of the speech waveform. As the pitch period of high-pitched speech is small, the periodic replicas cause ‘‘aliasing’’ of the autocorrelation sequence.

This paper organizes as follows: section-2 shows the mathematical calculation of LPC and section-3 Shows the explanation of wavelet transform in section-4 the results followed by conclusion and reference.

2. LPC (Linear Predictive Coder)

LPC system which is predicting from the previous samples used for identification of numerical values of frequencies. In this section, the LPC method is described for formant estimation that is implemented using a set of digital resonators. Each resonator represents a formant in a segment in the frequency domain. The spectrum is divided into segments such that only one formant resides in each segment.

The linear prediction problem can be stated as finding the coefficients which result in the prediction of the samples q(n) in terms of past samples q(n-k).

$$q(n)=\sum_{k=1}^{p}a_p(k)q(n-k) \qquad (1)$$

$$q(z)=\sum_{k=1}^{p}a_p(k)q(n-k)q(z) \qquad (2)$$

$$q(z)=\sum_{k=1}^{p}a_p(k)q(z)z^{-k} \qquad (3)$$

$$q(z)=\frac{1}{1+\sum_{k=1}^{p}a_p(k)q(z)z^{-k}} \qquad (4)$$

To determine $a_p(k)$ of the model we use autocorrelation of $S_A(n)$ and is given by

$$\sum_{k=1}^{p}a_p(k)r_{ss}(m-n)=-r_{ss}(m). \qquad (5)$$

The normal equation for minimizing the error is given by

$$\sum_{k=1}^{p}a_p(k)r_{ss}(m-n)=0 \qquad (6)$$
$$a_p(0)=1$$

Where m=1,2,….p and $r_{ss}(m)$ is the autocorrelation and is defined as

$$r_{ss}(m)=\sum_{n=0}^{N}S_A(n)S_A(n+m) \qquad (7)$$

Where S_A is the signal sample
The linear equation(5) can be expressed in the matrix form as

$$R_{SS}a=-r_{ss} \qquad (8)$$

Where $R_{SS}a$ is a p × p autocorrelation matrix, r_{ss} is a p×1 autocorrelation vector and ‘ a ’ is a p×1 vector of the model.

$$\begin{bmatrix} R_{SS}(0) & R_{SS}(1)....... & R_{SS}(P-1) \\ R_{SS}(1) & R_{SS}(0)....... & R_{SS}(P-2) \\ \vdots & \vdots & \vdots \\ R_{SS}(P-1) & R_{SS}(P-2)....... & R_{SS}(0) \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}=\begin{bmatrix} r_{ss} \\ r_{ss} \\ \vdots \\ r_{ss} \end{bmatrix}. \qquad (9)$$

The above equation can be solved by using Levinson-Durbin algorithm recursively as

$$a_1(1)=\frac{-R_{SS}(1)}{R_{SS}(0)} \qquad (10)$$

The next step to solve for the coefficients $\{a_2(1),a_2(2)\}$ of the second order predictor and expressing the solutions in the terms of $a_1(1)$. The two equations are obtained from equation (6) as

$$a_2(1)R_{SS}(0)+a_2(2)R_{SS}(1)=-R_{SS}(1)$$

$$a_2(1)R_{SS}(1)+a_2(2)R_{SS}(0)=-R_{SS}(2)$$

By using the solution in equation (10) to eliminate $R_{SS}(1)$, we obtain the solution

$$a_2(2)=-\frac{R_{SS}(2)+a_1(1)R_{SS}(1)}{R_{SS}(0)\left[1-|a_1(1)|^2\right]} \qquad (11)$$

Similarly all the parameters can be calculated recursively.

So $'a'$ can be written as $a=\lfloor a_1,a_2..........a_p\rfloor$

Now the difference equation from (4) can be written as

$$q(z)=\frac{S_0(Z)}{S_A(Z)}=-a(2)z^{-1}-a(3)z^{-1}.........a(n+1)z^{-p}$$

$$S_0(Z)=-a(2)S_A(Z)z^{-1}-a(3)S_A(Z)z^{-1}.........a(n+1)S_A(Z)z^{-p}$$

$$S_0(Z)=-a(2)S_A(n-1)-a(3)S_A(n-2).........a(n+1)S_A(n-p)$$

Where $'p'$ is the number of poles or the order of the filter and $S_0(Z)$ is the predicted value.

Choosing the order of the predictor and finding the roots of the parameter 'a' greater than 0.01 we will have frequency greater than zero hertz.

$$\text{Predictor order} = \mathrm{P_o} = \frac{\text{sampling frequency}}{\text{number of samples}}$$

The parameter 'a' is obtained from the LPC predictor[7].

$$a = lpc[S_A, P_o]$$

Calculating the roots R$_0$ of parameter 'a' and the frequency content is given by

$$F_{Content}=\left[\frac{\arctan 2(im(Ro), re(Ro))\times F_S}{No.of\ samples\ per\ cycle}\right] \qquad (12)$$

Where F$_S$=Sampling frequency

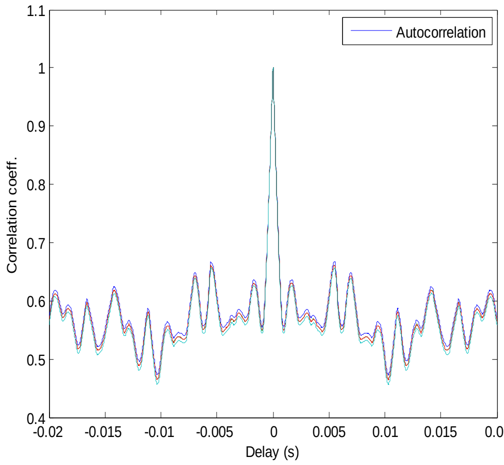

Fig.1 Autocorrelation function of the speech signal **" I LIKE DIGITAL SIGNAL PROCESSING".**

Figure-1 shows a means to estimate fundamental frequency from the waveform directly is to use *autocorrelation*. The autocorrelation function for a signal shows how well the waveform shape correlates with itself at a range of different delays. The autocorrelation approach works best when the signal is of low, regular pitch and when the spectral content of the signal is not changing too rapidly. The autocorrelation method is prone to pitch halving errors where a delay of two pitch periods is chosen by mistake.

we can see that the autocorrelation function peaks at zero delay and at delays corresponding to ± 1 period, ± 2 periods, etc. We can estimate the fundamental frequency by looking for a peak in the delay interval corresponding to the normal pitch range in speech, say 2ms(=500Hz) and 20ms (=50Hz).

Linear prediction models the signal as if it were generated by a signal of minimum energy being passed through a purely-recursive IIR filter.

3. Wavelet Transform

In this section, the wavelet transform and its implementation for discrete signals are reviewed briefly. This review is not intended by any means to be rigorous, and its sole purpose is to describe the tools. A wavelet is a wave-like oscillation with an amplitude that starts out at zero, increases, and then decreases back to zero. Unlike the sines used in Fourier transform for decomposition of a signal, wavelets are generally much more concentrated in time. They usually provide an analysis of the signal which is localized in both time and frequency, whereas Fourier transform is localized only in frequency.

In particular, the Wavelet Transform (WT) known as the “Mathematical Microscope” in engineering allows the changing spectral composition of a nonstationary signal to be measured and presented in the form of a time-frequency map and thus, it is suggested as an effective tool for non stationary signal analysis. It was first introduced by Morlet (Morlet et al. [1982]) in describing the Continuous Wavelet Transform (CWT) using Morlet wavelets.

In CWT any time series can be decomposed into a series of dilations and compressions of a mother wavelet denoted as $w(t)$. The advantage of this view is that high frequencies can be localized to a smaller time interval than low frequencies. The Continuous Wavelet Transform (CWT) of $x(t)$ is given by (Rioul and Vetterli [1991]).

$$w(a,b)=\int_{-\infty}^{\infty} x(t)\psi^{*}_{a,b}(t)dt \qquad (13)$$

Where $x(t)$ is any square integrable function, “a” is the dilation parameter, “b” is the translation parameter and $\psi^{*}_{a,b}(t)$ is the dilation and translation (asterisk (*) denotes the complex conjugate) of the mother wavelet defined as

$$\psi^{*}_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \qquad (14)$$

The signal $x(t)$ can be reconstructed from the continuous wavelet transform provided the mother wavelet satisfies the admissibility condition,

$$C=\int_{-\infty}^{\infty}\frac{|\psi(\omega)|^{2}}{|\omega|}d\omega<\infty \qquad (15)$$

where $\psi(\omega)$ is the Fourier Transform of $\psi(t)$. The reconstructed signal $x(t)$ is given as

$$x(t)=\frac{1}{C}\int_{a=\infty}^{\infty}\int_{b=\infty}^{\infty}\frac{1}{|a|^{2}}w(a,b)\psi_{a,b}(t)dadb \qquad (16)$$

A wavelet is a continuous time signal that satisfies the following properties

$$\int_{-\infty}^{\infty}\psi(t)dt=0 \qquad (17)$$

$$\int_{-\infty}^{\infty}|\psi(t)|^{2}dt<\infty \qquad (18)$$

Where $\psi(t)$ is defined as the mother wavelet.

The Continuous Wavelet Transform (CWT) is two dimensional. It is obtained by the inner product of the signal and dilations and translations of the mother wavelet.

- CWT is represented as a time scale plot, where scale is the inverse of frequency. At a low scale (high frequency), CWT offers high time resolution and at higher scales (lower frequencies) CWT gives high frequency resolution.
- The interpretations of the time scale representations produced by the Wavelet Transform (WT) require the knowledge of the type of the mother wavelet.
- Thus the visual analysis of the wavelet transform is intricate. Direct reading of the frequency of the signal as well as its frequency components from the time scale plot is difficult.

Wavelet Transform uses a flexible movable window and is designed to have

- Poor frequency resolution and good time resolution at high frequencies.
- Poor time resolution and good frequency resolution at low frequencies.
- CWT can be practically computed by using analytical equations, integrals but fail in discrete case.

The Discrete Wavelet Transform (DWT), provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. Discrete Wavelet Transform [63] was discovered by Daubechies (Daubechies [1990]). The Discrete Wavelet Transform (DWT) is a linear transformation performed on a time series. Effectively, the DWT is nothing but a system of filters.

4. Results and conclusion

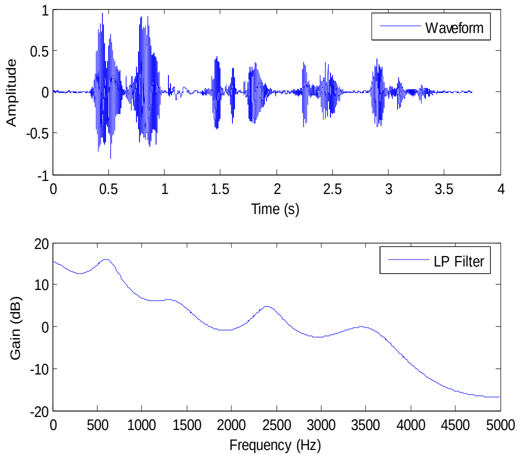

Fig-2:Noisy speech signal ; " PATTNAIK", Using LP Filter.

Formant 1 Frequency 629.0,
Formant 2 Frequency 1495.8
Formant 3 Frequency 2441.0
Formant 4 Frequency 3358.0

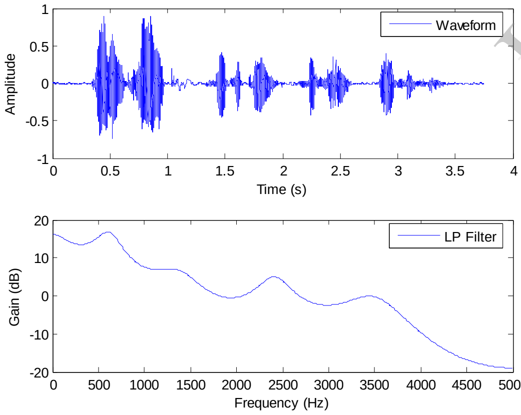

Fig-3: Denoising of signal; " PATTNAIK". Using LP Filter

Formant 1 Frequency 615.5
Formant 2 Frequency 1376.1
Formant 3 Frequency 2412.8
Formant 4 Frequency 3494.8

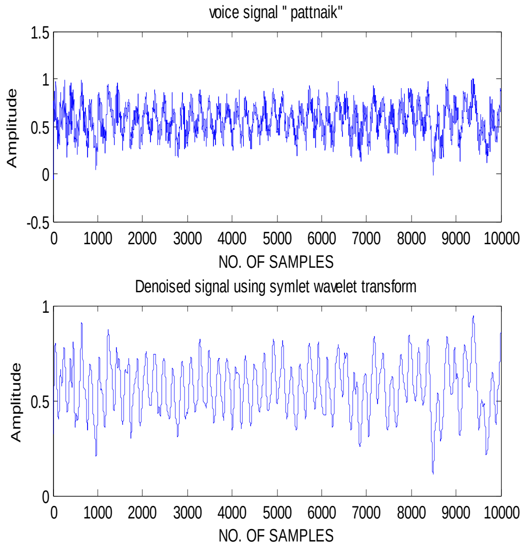

Fig-4: Denoising of signal; " PATTNAIK" by using "symlet" wavelet

Formant 1 Frequency 152.7
Formant 2 Frequency 2041.6
Formant 3 Frequency 3080.5
Formant 4 Frequency 4050.4

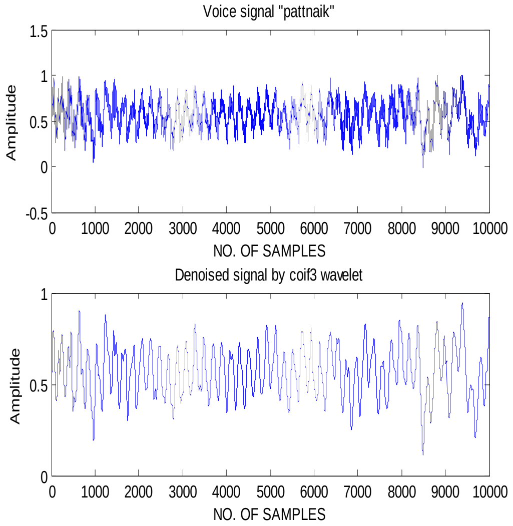

Fig-5: Denoising of signal; " PATTNAIK" by using "coif3" wavelet

Formant 1 Frequency 147.7
Formant 2 Frequency 1999.8
Formant 3 Frequency 3088.1
Formant 4 Frequency 4058.7

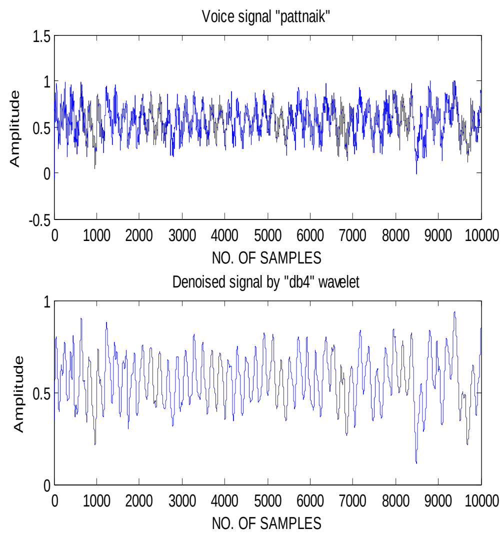

Fig-6: Denoising of signal; " PATTNAIK" by using "db4" wavelet

Formant 1 Frequency 448.8
Formant 2 Frequency 2085.2
Formant 3 Frequency 3096.2
Formant 4 Frequency 4055.8

Conclusion

Although the WT (Wavelet Transform) is also restricted by Heisenberg uncertainty principle, the window in WT can be adjusted. In the WT (Wavelet Transform), the mother wavelet can be stretched according to frequency to provide reasonable window, a long time window is used in low frequency and a short time window is used in high frequency. This time-frequency analysis which fully reflects the thought of multiresolution analysis is in accordance with the features of time varying non stationary signals.Estimation of formant frequencies is generally more difficult than estimation of fundamental frequency. The problem is that formant frequencies are properties of the vocal tract system and need to be inferred from the speech signal rather than just measured. The spectral shape of the vocal tract excitation strongly influences the observed spectral envelope, such that we cannot guarantee that all vocal tract resonances will cause peaks in the observed spectral envelope, nor that all peaks in the spectral envelope are caused by vocal tract resonances. To find the formant frequencies from the filter, we need to find the locations of the resonances that make up the filter. This involves treating the filter coefficients as a polynomial and solving for the roots of the polynomial. Practically, it is indeed very difficult to obtain a speech analysis. We, however, expect from the discussion that the proposed technique can be applied to analyze speech data when the conventional model of linear prediction is only an approximation to speech signal uttered by female and children speakers. Though the method is intended for analyzing high-pitched speech signal, the results demonstrate that it can also be used for analyzing typical male speech with better accuracy. Formant frequency estimation shows the frequency of vocal tract of male and female voice signals. The above estimation can be done through wavelet transform. The above result shows different formant frequency and its comparison.

References

[1] F. S. Chen, "Wavelet Transform In Signal Processing Theory And Applications", National Defense Publication of China, 1998.

[2] I. Daubachies, "Ten Lectures On Wavelets", Philadelphia, PA: SIAM, 1992.

[3] S. Mallat, " A Wavelet Tour Of Signal Processing", London,U.K.:Academic,1998.

[4] Ingrid Daubechies, "The Wavelet Transform, Time–Frequency Localization and Signal Analysis", *IEEE Trans. On Information Theory*, Vol.36, No.5, pp.961–1005, 1990.

[5] P. Rakovi, E. Sejdic, L.J. Stankovi and J. Jiang, "Time–Frequency Signal Processing Approaches with Applications to Heart Sound Analysis", *Computers in Cardiology*, Vol.33, pp.197–200, 2006.

[6] B. S. Atal and S. L. Hanauer, ''Speech analysis and synthesis by linear prediction of the speech wave,'' J. Acoust. Soc. Am.,50, 637–655 (1971).

[7] J. Makhoul, ''Linear prediction: A tutorial review,'' Proc.IEEE, 63, 561–580 (1975).

[8] A. K. Krishnamurthy and D. G. Childers, ''Two-channel speech analysis,'' IEEE Trans. Acoust. Speech Signal Process.,34, 730–743 (1986).

[9] M. Yanagida and O. Kakusho, ''A weighted linear prediction analysis of speech signals by using the Given's reduction,''Digital Signal Processing, M. H. Hamza, Ed., IASTED Int. Symp. Appl. Signal Processing and Digital Filtering, Paris,pp. 129–132 (1985).

[10] C. H. Lee, ''On robust linear prediction of speech,'' IEEE Trans. Acoust. Speech Signal Process., 36, 642–650 (1988).