ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)

Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Comparing Performance of Kalman Filtering and DWT based Speech Enhancement Techniques

K.Raghavendra Rao [1], Suneel Miriyala [2]

[1] PG Student, Department of ECE, Bapatla Engineering College, Bapatla, AP,India.

[2] Assistant Professor, Department of ECE, Bapatla Engineering College, Bapatla, AP,India.

raghukommana@gmail.com[1] , suneel.miriyala@gmail.com [2]

Abstract:

This paper gives an idea about the importance of speech enhancement and the performance analysis of DWT and Kalman filter based speech enhancement techniques. The objectives of Speech Enhancement vary widely reduction of noise level, increased intelligibility, reduction of auditory fatigue, etc. Speech processing systems are depending on the nature of the noise, and often on the signal-to-noise ratio (SNR) of the distorted speech. Enhancement of degraded speech is useful in aircraft, mobile, military and commercial communication.

Basically speech enhancement methods are available in time domain and transform domain approach. In this paper we are going to discuss about Kalman filter and Wavelet Transform based enhancement techniques which gives good results in high noise environment when compared to other techniques. The performance of these methods can be evaluated using PSNR and Mean Square Error (MSE).

Key words: Speech Enhancement, Filtering, DWT, Noise.

1) Introduction:

Speech is a general form of human communication. The speech signal is usually measured in terms of its quality and Intelligibility. The quality is a subjective measure that indicates the pleasantness of the delivered speech. Intelligibility is an objective measure which predicts the percentage of words that can be correctly identified by listeners. Enhancement means the improvement in the value or quality. When applied to speech, this simply means the improvement in intelligibility and/or quality of a degraded speech signal by using signal processing tools. By speech enhancement, it refers not only to noise reduction but also to separation of independent signals

Speech enhancement has three major objectives: (a) to improve the quality and intelligibility of the processed speech, i.e., to make it sound better or clear to the listener. (b) to improve the robustness of speech coders which tends to be affected by the presence of noise. (c) To increase the accuracy of speech recognition systems operating in less ideal locations. Significant advances in Speech Enhancement will results in better understanding of human speech. The basic Speech Enhancement methods (spectral subtraction, Wiener filtering) have existed for decades. They are largely based on simple ideas such as masking (e.g., suppressing noise at frequencies where the speech harmonics are weak) or on simple speech production models (cleaner speech, but of poor quality).

Contribution: In this paper, an overview of speech enhancement is provided. Different types of speech enhancement techniques are explained along with their merits and demerits. By using PSNR and Mean Square Error (MSE) of the speech enhancement using DWT and kalman filter, we can evaluate their performance. Enhancing speech in noisy environment is explained with latest techniques [1].

2) Need for speech enhancement:

Speech signals may contain degrading components like additive background noise along with required speech components. This make the listening task difficult and gives poor performance in automatic speech processing tasks like speech recognition, speaker identification etc. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Quality can be measured in terms of signal distortion but intelligibility and pleasantness are difficult to measure.

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

It is also a good idea to enhance speech for coding and recognition purposes. Speech codecs are limited for speech and they usually make the background noise sound weird. Moreover, enhanced speech can be compressed in fewer bits than non-enhanced. Speech recognition systems whose operation relies on the features extracted from speech will be disturbed by extra noise sounds [7].

Active noise suppression is a method in which the idea is to produce anti-noise to cancel the noise. The delay must be kept very small to avoid producing more noise instead of cancelling the existing noise. For this reason, most of the methods for active noise suppression are fully analog: A/D and D/A transforms inevitably produce some amount of delay. Speech enhancement is important in the telecommunications area for increasing the amount of information which can be transferred, stored, or heard, for a given set of time and space constraints.

3) Speech Enhancement Techniques:

The basic method for speech enhancement is Spectral Subtraction approach. Basically speech enhancement methods are available in time domain and transform domain approach. Other methods for Speech Enhancement are spectral subtraction along with Adaptive filter, Kalman filters, Wavelet Transform, Least Mean Square (LMS) algorithm, normalized LMS (NLMS) algorithm and Recursive Least Squares (RLS) algorithm.

3.1 Spectral Subtraction:

The basic method for speech enhancement is Spectral Subtraction method. In this section, we provide a brief overview on conventional spectral subtraction. Spectral subtraction assumes that a signal is composed of two additive components. The noisy speech can be expressed as

$$y(t) = s(t) + d(t) \qquad \text{….(3.1.1)}$$

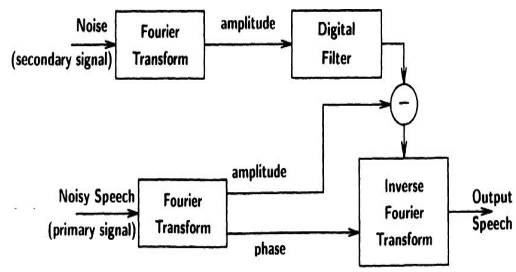

Figure 3.1.1 Block diagram of Spectral Subtraction Technique

Where t is time, s (t) represents the uncorrupted speech signal; d (t) represents the additive noise signal and y (t) is the corrupted speech signal available for processing. The observed signal y(t) is divided into overlapping frames using the application of a window function and implemented in the short-time Fourier transform (STFT) magnitude domain. In the frequency domain this can be represented as

$$Y(\omega)=S(\omega)+D(\omega) \qquad \text{….(3.1.2)}$$

The power spectrum of noisy speech can be estimated as:

$$Y(\omega)^2=S(\omega)^2+\delta n(\omega) \qquad \text{…..(3.1.3)}$$

Where $\delta n(\omega)$ is the statistical average values of D(ω)2 during non-speech stage. So the enhanced speech amplitude is

$$\hat{S}(\omega)=[Y(\omega)^2-E(D(\omega)^2)]1/2=[Y(\omega)^2-\delta(\omega)^2]1/2..(3.1.4)$$

Combined with the phase of the noise corrupted signal to re-synthesize the signal

$$S(\omega)=|\hat{S}(\omega)|ejargY(\omega) \qquad \text{… (3.1.5)}$$

The inverse short-time Fourier transform is performed to transform the signals into time domain. Conventional spectral subtraction algorithm estimating noisy energy during no speech stage, however, it can't update noise during speech stage. Also the method requires a VAD that might not work very well under low SNR [2].

3.2) Adaptive filter:

The speech enhancement methods aimed at suppressing the background noise are based on its estimation. If the noise is slower than the speech, it is

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

easy to estimate the noise during the pauses in the speech. If the noise is varying rapidly, then estimation is more difficult. The reduction of additive background noise in speech is done by Weiner filter and Kalman filter respectively [3].

Wiener filtering estimates noise free speech signal from the noisy signal corrupted by additive noise. Estimation is performed by minimizing the Mean Square Error (MSE) between the noise free signal (n) and its estimation *x*(n). The problem with this method is that it has fixed frequency response at all frequencies and it also requires estimation of power spectral densities of noise free and noise signal before filtering. To solve this problem, M.A. Abd E-Fattah presented adaptive wiener filtering approach in 2008. According to this approach enhanced speech signal of small segment stationary noisy signal can be represented as

$x\hat{}(n)= m_x + (x (n) - m_x)*V_x /(V_x + V_d)$ ….(3.2.1)

Where m_x is mean of noise free speech signal, V_x and V_d are variance of noise free speech and noise respectively. If V_x is smaller than V_d input signal x (n) is attenuated due to filtering effect. Different steps involved in implementation of speech enhancement using Adaptive Wiener Filtering are shown in Fig

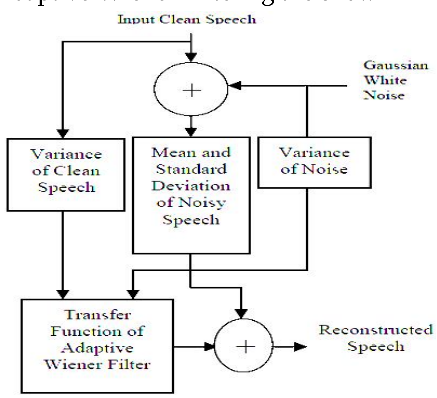

Figure 3.2.1 Adaptive Wiener Filtering based speech enhancement

3.3) LMS Algorithm:

The simplicity of the Least Mean Square (LMS) algorithm and ease of implementation makes it the best choice for many real-time systems. The implementation steps for this algorithm can be stated as;

i. Define the desired response and set each coefficient weight to zero.

$$w(n) = 0, n=1,2,3..N \quad ....(3.3.1)$$

For each sampling instant (n) carry out the following steps;

ii. Move all samples in the input array one position to the right, now load the current data sample (n) into the first position in the array. Calculate the output of the adaptive filter by multiplying each element in the array of filter coefficients by the corresponding element in the input array and all the results are summed to give the output corresponding to that data that was earlier loaded into the input array, such that the output y(n) is;

$$y(n) = \Sigma\, w(n)\, x(n) \quad ....(3.3.2)$$

iii. Before the filter coefficients can be updated, the error must be calculated, simply find the difference between the desired response, d(n), and the output of the adaptive filter, y(n).

$$e(n) = y(n) - d(n) \quad ....(3.3.3)$$

iv. To update the filter coefficients multiply the error by the learning rate parameter, μ, and then multiply the results by the filter input and add this result to the values of the previous filter coefficients.

$$\ddot{w}(n+1) = \ddot{w}(n) + \mu.e(n).\ddot{x}(n) \quad ....(3.3.4)$$

Where μ is the step size of the adaptive filter, $\ddot{w}(n)$ is the filter coefficients vector, $\ddot{x}$ (n) is the filter input vector.

Then LMS algorithm calculates the cost function J(*n*) by using the following equation:

$J(n) = e^2(n)$ Where $e^2(n)$ is the square of the error signal at time(n). The resources required to implement the LMS algorithm for a transversal adaptive FIR filter of L coefficients in real time is given in Table I. The computations given are those required to process one sample.

3.4) Kalman Filtering:

The filter is a mathematical procedure which operates through a prediction and correction mechanism. A kalman filter is simply an optimal recursive data processing algorithm.

The complete estimation procedure is as follows:

The model is formulated on state-space and for an initial set of parameters given, the model prediction errors are generated by the filter. these are used

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)

Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

recursively to evaluate the probability function until its maximization.

As a summary, we can say that the kalman filter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized [4].

Algorithm:

The kalman filter estimates the previous process using a feedback control, that is, it estimates the process to a moment over the time and then it gets the feedback through the observed data.From the equation point of view that is used to derivate the kalman filter,it is possible to separate them into two groups:

1.) Those which update the time or prediction equations

2.) Those which update the observed data or update equations

The first group of equations has to throw the state to the n moment taking as reference the state on n-1 moment and the intermediate update of the covariance matrix of the state. the second group of equations has to take care of the feedback; they add new information inside the previous estimation to achieve an improved estimation of the state.

The specified equations for the state prediction are detailed as follows:

$$x_{n|n-1} = A_{n,n-1} \cdot x_{n-1} \quad \ldots (3.4.1)$$

$$R_{e,n|n-1} = A_{n,n-1} \cdot R_{e,n-1} \cdot A^{T}_{n,n-1} + u \cdot R_{w} \cdot u^{T} \ldots (3.4.2)$$

The specified equations for the state correction are detailed as follows:

$$R_{e,n} = R_{e,n|n-1} - R_{e,n|n-1} \cdot C^{T} \cdot F_{n}^{-1} \cdot C \cdot R_{e,n|n-1} \ldots (3.4.3)$$

$$x_{n} = x_{n|n-1} + R_{e,n|n-1} \cdot C^{T} \cdot F_{n}^{-1}(y_{n} - C \cdot x_{n|n-1}) \ldots (3.4.4)$$

$$F_{n} = C \cdot R_{e,n|n-1} \cdot C^{T} + R_{v} \ldots (3.4.5)$$

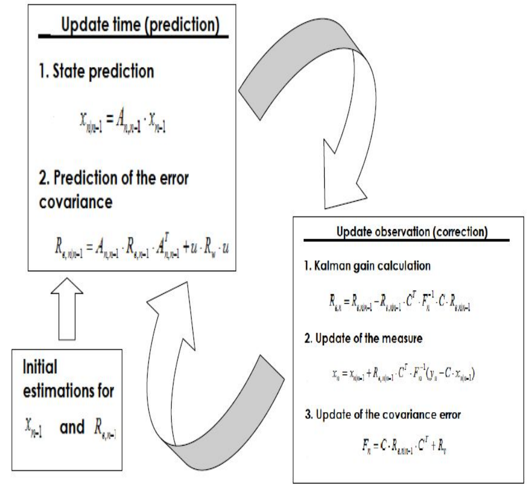

Figure 3.4.1. Block Diagram complete operation of kalman filtering

3.5) Wavelet Thresholding (using DWT):

Wavelet transform is a powerful tool for modeling non-stationary signals like speech that exhibit slow temporal variations in low frequency and abrupt temporal changes in high frequency. Moreover, when one is restricted to use only one (noisy) signal (as in single-microphone speech enhancement), generally the use of the subband processing can result in a better performance. Therefore, wavelet transform can provide an appropriate model of speech signal for denoising applications [5].

Removing noise components by thresholding the wavelet coefficients is based on the observation that in many signals (like speech), energy is mostly concentrated in a small number of wavelet dimensions. The coefficients of these dimensions are relatively large compared to other dimensions or to any other signal (specially noise) that has its energy spread over a large number of coefficients. Hence, by setting smaller coefficients to zero, one can nearly optimally eliminate noise while preserving the important information of the original signal.

Algorithm:

Let **y** be a finite length observation sequence of the signal **x** that is corrupted by zero-mean, white Gaussian noise **n** with variance σ2.

$$y = x + n \quad \ldots (3.5.1)$$

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

The goal is to recover the signal **x** from the noisy observation **y**. If W denotes a discrete wavelet transform (DWT) matrix, equation (3.8.1) (which is in time domain) can be written in the wavelet domain as

$$\mathbf{Y}=\mathbf{X}+\mathbf{N} \qquad \dots (3.5.2)$$

Where $\mathbf{Y}=W\mathbf{y}, \mathbf{X}=W\mathbf{x}, \mathbf{N}=W\mathbf{n}$ … (3.5.3)

Let $\mathbf{X}_{est}$ be an estimate of the clean signal **X** based on the noisy observation **Y** in the wavelet domain. The clean signal **x** can be estimated by

$$\mathbf{x}=W^{-1}\mathbf{X}_{est}=W^{-1}\mathbf{Y}_{thr} \qquad \dots (3.5.4)$$

Where $\mathbf{Y}_{thr}$ denotes the wavelet coefficients after thresholding.

The proper value of the threshold can be determined in many ways. D.L Donoho has suggested the following formula for this purpose

$$T=\sigma[2\log(N)]^{1/2} \qquad \dots (3.5.5)$$

Where T is the threshold value and N is the length of the noisy signal (**y**). Thresholding can be performed as *Hard* or *Soft* thresholding as given below,

$$THR_H(Y,T)=\begin{cases} Y & ,|Y|>T \\ 0 & ,|Y|<T \end{cases} \qquad \dots (3.5.6)$$

$$THR_S(Y,T)=\begin{cases} Sgn(Y)(|Y|-T) & ,|Y|>T \\ 0 & ,|Y|<T \end{cases} \qquad \dots (3.5.7)$$

Some major problems arise when the basic wavelet thresholding method is applied to a complex signal such as speech degraded by real-life noises. This problem can be avoided by using improved wavelet based speech enhancement technique as mentioned in below figure

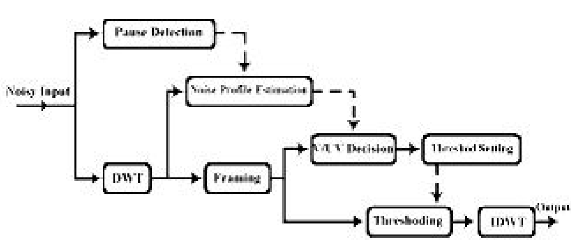

Figure 3.4.2 Improved wavelet based speech enhancement technique

4) Speech Objective Quality Measures:

The objective comparison of three single channel speech enhancements is carried by evaluating performance of parameters such as, Mean Square error (MSE), peak Signal to Noise Ratio (SNR). It is based on mathematical comparison of the original and processed speech signal [6].

i. Peak Signal to Noise Ratio (PSNR):

$$\text{PSNR (dB)} = 10\log10(\,NX^2/||x-r||^2) \qquad \dots (4.1)$$

Where N is the length of the reconstructed signal, X is the maximum absolute square value of signal 'x' and $||x-r||^2$ is the energy of the difference between the original and reconstructed signal.

ii. Mean Square Error (MSE):

$$\text{MSE}=1/N[(r(n)-x(n))^2] \qquad \dots (4.2)$$

Where N is length of input speech signal, x(n) is input speech signal and r(n) is reconstructed speech signal.

Comparing the performance of discussed speech enhancement techniques are given below(with AWGN and Random noise)

Speech Enhancement Technique	Mean square error (MSE)	Peak signal to noise ratio (PSNR)
Speech enhancement Using Kalman filter with AWGN	0.0035	72.59
Speech enhancement Using DWT with AWGN	0.0014	88.73

Table 1: Comparison of MSE and PSNR when speech signal effected by AWGN

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 288 - 293 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Speech Enhancement Technique	Mean square error (MSE)	Peak signal to noise ratio (PSNR)
Speech enhancement Using Kalman filter with random noise	0.0036	72.52
Speech enhancement Using DWT with random noise	0.0021	88.53

Table2: Comparison of MSE and PSNR when speech signal is effected by Random noise

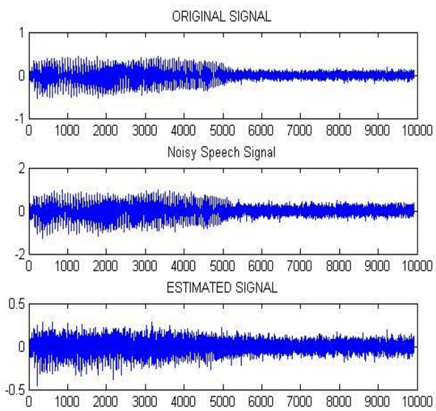

Figure4.1:Experimental results speech enhancement using Kalman filtering

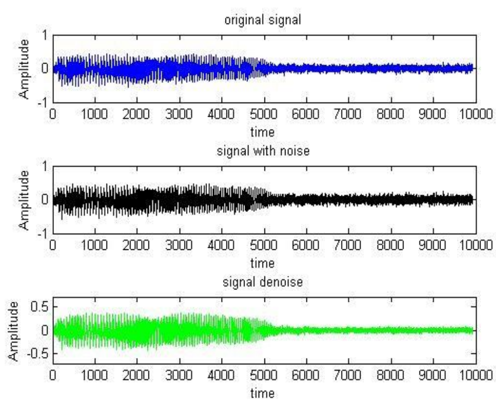

Figure4.1:Experimental results speech enhancement using Discrete Wavelet Transform

Conclusion:

In this paper, Implementation of employing kalman filtering and DWT for speech enhancement is analyzed. The main aim of this paper is to show the performance of the speech enhancement techniques with respect to PSNR and MSE. In general speech enhancement the test results of Kalman filtering and DWT shows somewhat similar results. But, by considering noisy speech enhancement case, the DWT based techniques shows good results when compared with the Kalman filtering.

References:

[1] Milind U. Nemade, Prof. Satish K. Shah "Performance Comparison of Single Channel Speech Enhancement Techniques for Personal Communication", International Journal of Innovative Research in Computer and Communication Engineering, ISSN (Print) : 2320 – 9798, Vol. 1, Issue 1, March 2013.

[2] Paurav Goel ,Anil Garg "Developments In Spectral Subtraction For Speech Enhancement" , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 ,Vol. 2, Issue 1, Jan-Feb 2012, pp.055-063.

[3] T. Lalith Kumar and K. Soudarya Rajan "Speech Enhancement Using Adaptive Filters", VSRD-IJEECE, Vol. 2 (2), 2012, 92-99.

[4] T. Lalith Kumar and K. Soudara Rajan "Performance Comparison Of Adaptive Filters With Kalman Filter For Speech Enhancement", Bookman International Journal of Electrical & Electronics Engineering, Vol. 1 No. 1 Sep. 2012, ISSN No. 2319-4294.

[5] Hamid Sheikhzadeh and Hamid Reza Abutalebi "An Improved Wavelet-Based Speech Enhancement System", Amirkabir University of Technology (Tehran Polytechnic) Hafez Ave, Tehran, Iran.

[6] Allam Mousa, Marwa Qados, Sherin Bader " Speech Signal Enhancement Using Adaptive Noise Cancellation Techniques", Canadian Journal on Electrical and Electronics Engineering Vol. 3, No. 7, September 2012

[7] A text book on „Speech Communications" by Douglas O'Shaughnessy, IEEE Press, 2000.