IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 2, Ver. VI (Mar - Apr. 2014), PP 48-53
www.iosrjournals.org

A Novel Hybrid System for Speech Enhancement In Wavelet Domain

SaurabhPandey[1], Arvind Kumar Jaiswal[2], NeeleshAgrawal[3]
$1_{PG\ Student}$, $2_{Professor}$, $3_{Assistant\ Professor}$
Department of Electronics and Communication Engineering
Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh, India -211007

Abstract: *The basic problem suffered by a communication system is in the removal of additive background noise. Spectral Subtraction method (SSM) has been successfully implemented to suppress the background noise. This paper proposed a novel hybrid system for enhancing noise-corrupted speech which involves a parallel combination of Spectral Subtraction and soft thresholding method in wavelet domain for speech enhancement. The performance of hybrid system is compared with both Spectral Subtraction method and Wavelet thresholding technique (WTT) and found to have better SNR improvement. Implementation and evaluation have been made using MATLAB.*

Keywords: *Inverse Discrete Wavelet Transform (IDWT), Spectral Subtraction method (SSM), voice activity detector (VAD), Wavelet thresholding technique (WTT).*

I. INTRODUCTION

There are many conditions when speech has processed in the presence of undesirable additive background noise that degrades speech quality and intelligibility. The reduction of these kinds of noise is essential in a noisy environment like Helicopter cockpit, in a car, for military purpose and many more. There are many ways to perform the noise reduction but our basic goal is to reduce the noise without deterioration in the speech quality (as minimum as possible). The spectral Subtraction method [1]is most commonly used method for suppression of background noise in frequency domain. In SSM an estimate of background noise is calculated during non-speech activity with the help of voice activity detector (VAD) [1]. This estimate of noise spectrum is subtracted from the noisy speech spectrum to obtain cleaned speech at the output. Wavelet thresholding is another method of denoising the speech signal but it is not successfully implemented for background noise[2]reduction. In this paper a new hybrid method of speech enhancement is proposed in which first we take the wavelet transform of noisy speech signal that yields Approximation and detail coefficients at level 1. Approximation coefficients are low frequency essential components of input signal while detail coefficients are high frequency less essential components of input signal. After getting these two coefficients, Approximation coefficients are given as input signal in SSM and detail coefficients are given as input signal in Wavelet thresholding technique (WTT)[3] . The output signals of both techniques are reconstructed using Inverse Discrete Wavelet Transform (IDWT).

II. THEORY

(a) Discrete Wavelet Transform

A novel noise reduction hybrid method is adopted which is based upon a combination of Spectral subtraction method and soft thresholding in wavelet domain [4]. This hybrid system needs to work in wavelet domain.Wavelet is a new technique for analysing and compressing a speech signal, it is more advantageous technique because it holds both time and frequency aspect of the signal and have localized analysis of a larger signal. The basic concept behind wavelet is to analyse a signal according to the scale. The first essential thing is to choose a mother wavelet then any signal can be represented by its translated and scaled version. Discrete wavelet transform [4] breaks the signal into high frequency and low frequency components. The output of high pass filter is known as detail coefficients and the output of low pass filter is approximation coefficients. Approximation coefficients [3] are high scaled low frequency components while detail coefficients are low scaled high frequency components. The output of low pass filter is further decomposed into high and low frequency components. This process is repeated to n levels; finally we have (n+1) outputs with analysed approximation coefficients at level 1 by using MATLAB command$\boldsymbol{sound}(\boldsymbol{cA1}, \boldsymbol{Fs}, \boldsymbol{bit\ depth})$. we can understand the speech with a little loss in speech quality. This shows that low frequency components contain essential information and that is why the output of LPF is called approximation coefficient. The output of HPF contains only high frequency non-essential information and is known as detail coefficient.

For applying wavelet technique, first we have to choose an appropriate mother wavelet and level of decomposition of the signal. Choosing a mother wavelet depends on the type of the signal we have to decompose. With speech de-noising , objective is to improve quality of the signal, so wavelet can be selected on the basis of energy conservation properties in approximation coefficients[4] .For selecting a decomposition level, if the frame based input is applied, then frame size must be a multiple of 2^n, where n represents the decomposition level.

The discrete wavelet transform DWT can be simply thought of in terms of filter banks. A filter bank [4],[5] is defined as a set of filters which are applied to a signal together with changes in sampling rates. The simplest case is the two-channel filter bank which consists of a low-pass and a high-pass filter, represented by the coefficients **h** and **g** respectively.

(b) Spectral Subtraction Method

Spectral subtraction method is one of the most popular methods for reducing the background noise [2] . In this method only an estimate of noise is needed for speech enhancement. In case of single channel, that noise estimation is taken when speaker is silent referred as 'non-speech activity'. The noise estimation is determined with the help of 'voice activity detector' (VAD) [1]. This plays most essential role in SSM. The role of VAD is to determine the estimate of noise during non-speech activity. Spectral subtraction method is suitable for stationary or slowly varying noise.

Spectral subtraction assumes that a signal is composed of two additive components i.e. the input signal can be expressed as the sum of speech spectrum and noise spectrum. The noisy speech signal $y(m)$ can be represented as

$$y(m) = x(m) + n(m)$$

$x(m)$s the uncorrupted speech signal and $n(m)$ is the noise signal. The additive noise introduced in speech signal impairs the speech quality. Take the Fourier transform of noisy speech signal

$$Y(\omega) = X(\omega) + N(\omega)$$

If $\hat{N}$ is an estimate of the noise spectrum, then an approximation of speech $\hat{X}$ can be obtained from Y

$$\hat{X}(\omega) = Y(\omega) - \hat{N}(\omega)$$

The noise estimate in spectral subtraction uses the VAD to decide when to update the noise reference during non-speech activity.

After subtracting, the values having negative magnitude are to zero by half-wave rectification process. This produces some artifacts, when converting the speech signal from frequency domain to time domain.

(c) Wavelet Thresholding Technique

Unlike conventional techniques, in wavelet transform input signal decomposes into high and low frequency components and the output is known as detail and approximation coefficients respectively. When the signal get decomposed into approximation and detail coefficients, then we truncate the small valued coefficients considering it to non- essential part of the signal. For this purpose there are two types of thresholding- hard and soft thresholding [6], [7].

Hard Thresholding- Hard thresholding set any coefficient whose absolute value is less than or equal to the threshold to zero. Suppose, Y refer the coefficient of a wavelet dimension of the noisy speech signal and T be the threshold value for the denoising the speech signal.

$$\hat{Y} = \begin{cases} Y, & if\ Y > T \\ 0, & if\ Y \leq T \end{cases}$$

Where $\hat{Y}$ denotes the thresholded coefficients.

Soft Thresholding- Soft thresholding sets any coefficient whose absolute value is less than or equal to the threshold to zero and subtracts the threshold value from the other coefficient.

$$\hat{Y} = \begin{cases} sgnY(|Y| - T), if\ Y > T \\ 0, \qquad\qquad if\ Y \leq T \end{cases}$$

It can be easily observed that soft thresholding perform de-noising operation better than hard thresholding as it removes more noise components. However the speech quality degradation is also higher in it.

A MATLAB function *wdencmp* enables us to choose whether it is global thresholding or level-dependent thresholding. Level dependent thresholds are calculated using Brige-Massart strategy. According to this strategy all the approximation coefficients are kept at the level of decomposition j. the number of detail coefficients that are to be kept at level i from 1 to j are given by the formula

$$n_j = \frac{M}{(J + 2 - j)^a}$$

Typically, a = 1.5 for compression and a = 3 for de-noising.
If wavelet coefficients are high scarcely distributed then value of M is equal to L, where L is the length of coarsest approximation coefficients.

III. EXPERIMENTAL SETUP

HYBRID SYSTEM (SSM+WT)

Spectral Subtraction Method is very popular for background noise reduction while Wavelet thresholding technique is used for de-noising a speech signal with variety of noises. For background noise reduction SSM has been more successful than WTT. This paper proposed a new hybrid technique for background noise reduction which contains a parallel connection of SSM and WTT in Wavelet domain. The effectiveness of this hybrid system is greater than the other two methods SSM and WTT in terms of signal de-noising.

We take Wavelet transform of a background noise corrupted, noisy speech signal $y(m)$with mother wavelet db2.Wavelet decomposition at level 1 yields high frequency components and low frequency components with the help of High Pass Filter and Low Pass Filter respectively. The output of HPF is known as detail coefficients and the output of LPF is known as Approximation coefficients.

Calculation of the concentrated energy in Approximation and detail coefficients results in energy in Approximation coefficients (Ea) = 97.0545% and Energy in Detail coefficients (Ed) = 2.9455%. The energy concentrated in both types of coefficients displays that Approximation coefficients are essential components of speech signal while detail coefficients are high frequency, very less essential components of speech signal. Figure (2) is noisy speech signal while figure (3) and (4) are approximation and detail coefficients of noisy speech signal respectively.

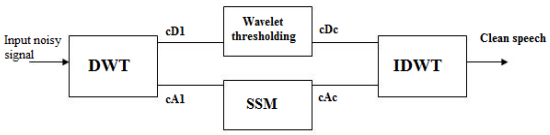

Figure (1) Hybrid System

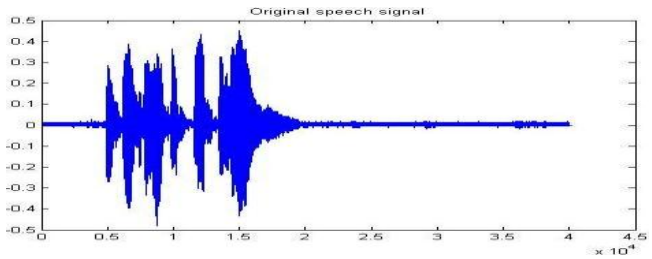

Figure (2) Original speech signal

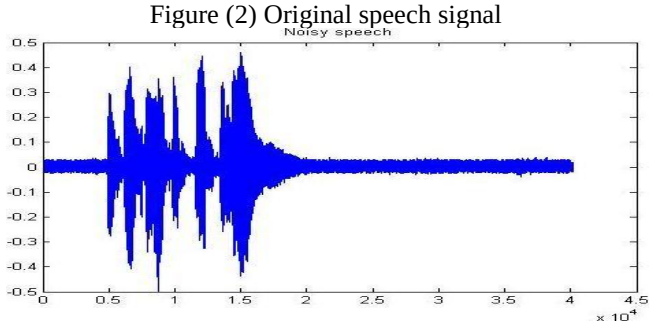

Figure (3) Noisy speech signal

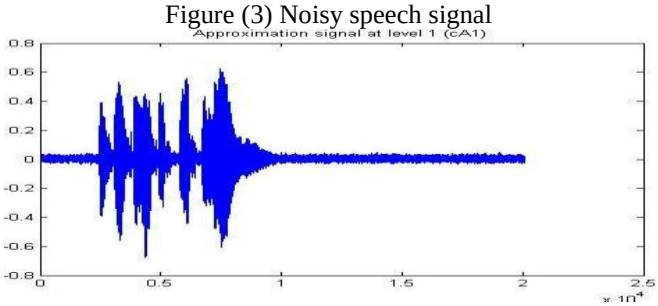

Figure (4) Approximation coefficients at level 1

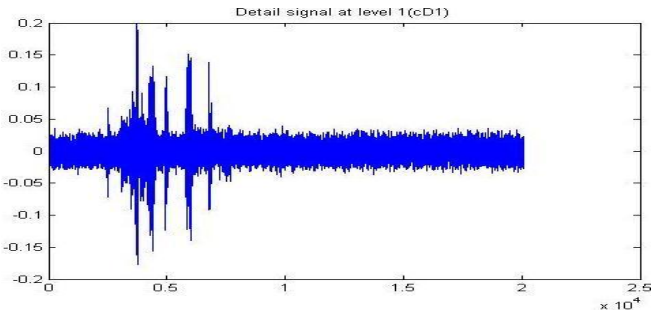

Figure (5) Detail coefficients at level 1

The work has exploited the fact that approximation coefficients represent the original signal (in this case, noisy signal) and detail coefficients are less essential noisy components. Approximation coefficients at level 1 (cA1) is treated as input signal for Spectral Subtraction method (SSM) and detailed coefficients at level 1 (cD1) is treated as input signal for Wavelet thresholding technique (WTT). The outputs of SSM and WTT are used for Inverse discrete Wavelet Transform (IDWT) whose purpose is to reconstruct the clean speech signal in time domain.

VI. Experimental Results

A male spoken sentence "Compression of speech signal" of 5 second with sampling frequency of $8000Hz$ and bit depth of 16 has been taken. For making this original speech signal noisy, noise has been added digitally in original speech. Three methods have been used for background noise reduction and it has been observed that Hybrid system is better for background noise reduction. The performance evaluation of proposed method has been conducted using the quality measure SNR which is calculated as detailed in following paras.

Unprocessed noisy Speech-to-Noise Ratio

The SNR of the unprocessed noisy speech is defined as the ratio of the clean signal power to the noise power.

$$SNR_n = 10log_{10}\frac{\sum_{m-1}^{N} x(m)^2}{\sum_{m-1}^{N} n(m)^2}$$

Where N is the length of the sentence expressed in number of samples.

Processed speech Signal to Noise Ratio

As the enhancement can amplify or attenuate the signal, and for a homogeneous evaluation and comparison methods, the enhanced signal is scaled to the same dynamic range as of the clean speech. It is accomplished by normalizing the enhanced speech $\hat{x}(m)$ to the clean speech $x(m)$. The resulting scaled signal $\tilde{x}(m)$ is defined as:

$$\tilde{x}(m) = \hat{x}(m)\frac{\max(|x(m)|)}{\max(|\hat{x}(m)|)}$$

The efficiency of the enhancement method is defined by the SNR of the enhanced speech i.e.

$$SNR_p = 10log_{10}\frac{\sum_{m=1}^{N} x^2(m)}{\sum_{m-1}^{N}(\tilde{x}(m) - x(m))^2}$$

The denominator is the difference between the clean original signal and the enhanced scaled signal. A small difference characterizes a good match between the two signals.

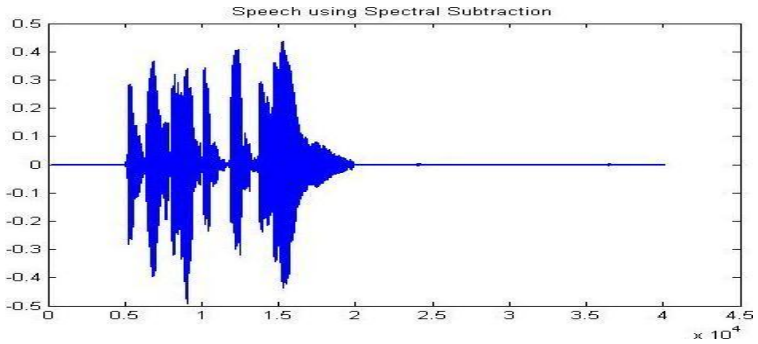

Figure (6) clean speech using SSM

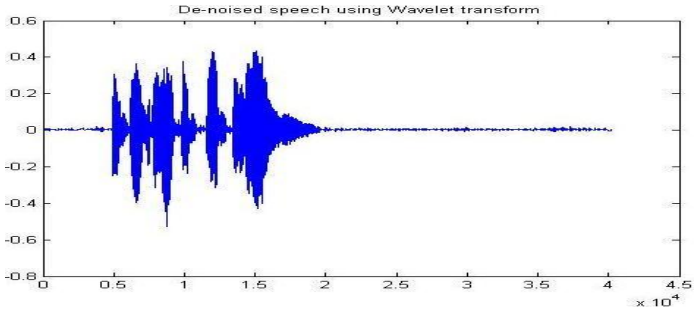

Figure (7) clean speech using WTT

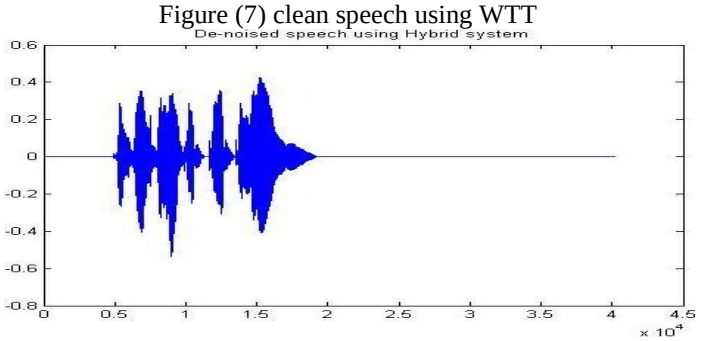

Figure (8) clean speech using hybrid system

Table (a) indicates the value of SNR with different techniques adopted while conducting the performance evaluation. It has also been presented graphically in fig (9)

Table (a) Values of SNR with different techniques

Signals	**SNR(in dB)**
Noisy	-0.1085
WTT	17.3344
SSM	23.7310
Hybrid system	26.3062

Fig 10 to Fig 13 shows the spectrograms of Noisy Speech signal and clean speech signals using,WTT, SSM and Proposed Hybrid system respectively, supporting the observed analysis.

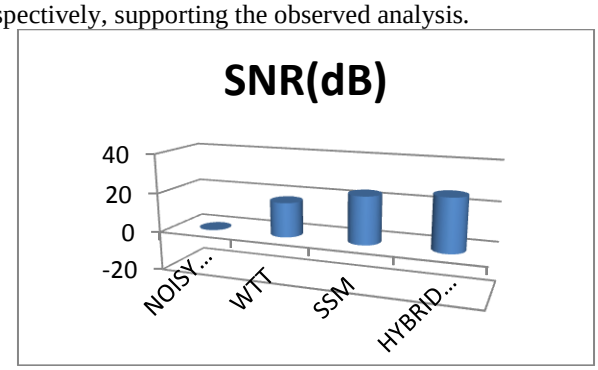

Figure (9) Plot of SNR with different techniques

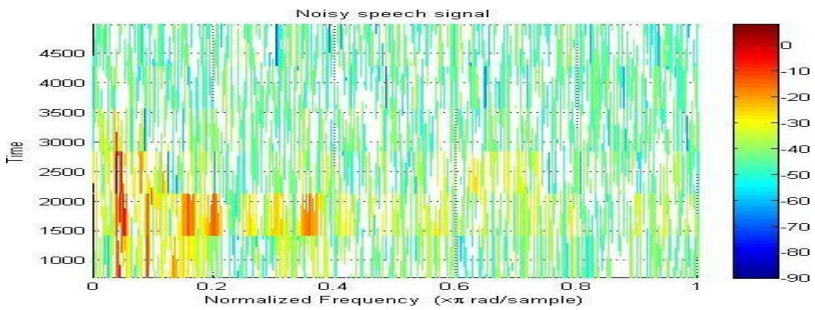

Figure (10) Spectrogram of Noisy speech signal

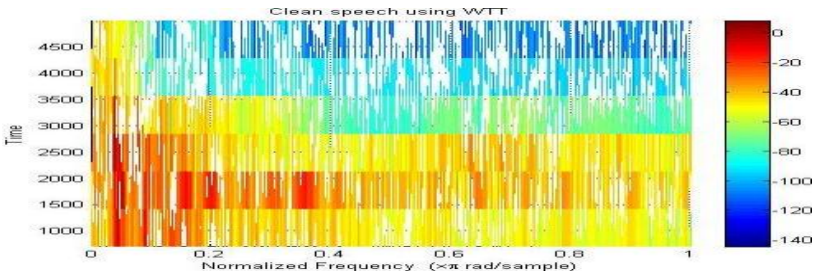

Figure (11) Spectrogram of clean signal using WTT

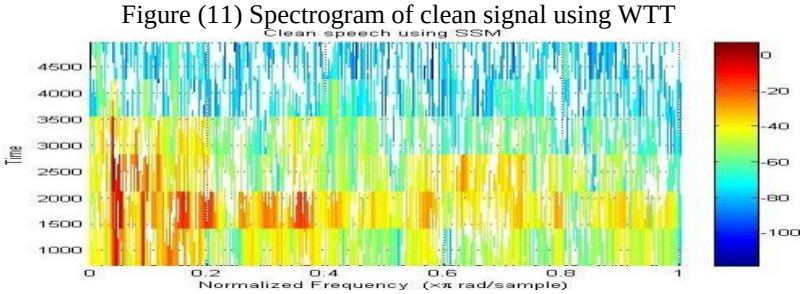

Figure (12) Spectrogram of clean signal using SSM

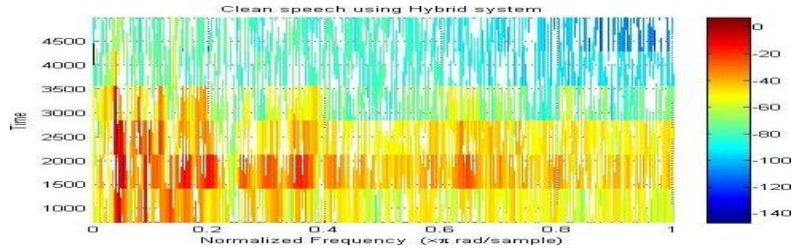

Figure (13) Spectrogram of clean signal using Proposed Hybrid system

VII. Conclusion

The proposed hybrid system is implemented for reducing additive background noise. The performance of this hybrid system is compared with SSM and WTT separately on the basis of SNR and it has been observed that Hybrid system works better for background noise reduction. Performance analysis brought in the values of signal to noise ratio (SNR) are given in table (a).

References

[1] Farid Ykhlef, A. Guessoum and D. Berkani, "Speech Enhancement Based on a Combination of Spectral Subtraction and a Minimum Mean- Square Error Short-Time Log-Spectral Amplitude Estimator in Wavelet Domain"2012.

[2] Malihehassani, M. R. Karamimollaei "Speech Enhancement Based on Spectral Subtraction in Wavelet Domain"*IEEE 7th International Colloquium on Signal Processing and its Applications*. 2011

[3] S.F.Boll, "Suppression of acoustic noise in speech, using spectral subtraction" .*IEEE. Acoustic. Speech, Signal Processing, vol. ASSP-27*, pp. 113-120, Apr. 1979.

[4] Birgé, L., P. Massart "From model selection to adaptive estimation," in D. Pollard (ed.), *Festschrift for L. Le Cam*, Springer, (1997), pp. 55-88.

[5] Anuradha R. Fukane, Shashikant L. Sahare, "Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments"

[6] Ekaterina Verteletskaya, Boris Simak, "Noise Reduction Based on Modified Spectral Subtraction Method" *IAENG International Journal of Computer Science*, 38:1, IJCS_38_1_10

[7] G. R. Mishra, Saurabh Kumar Mishra, Akanksha Trivedi, O.P. Singh, Satish Kumar, "Improving the Efficiency of Spectral Subtraction Method by combining it with Wavelet Thresholding Technique". International Journal of Research in Computer Science, 3 (3): pp. 29-33, May 2013. doi: 10.7815/ijorcs. 33.2013.065

[8] Saeed V. Vaseghi "Advanced Digital Signal Processing and Noise Reduction", Second Edition. Copyright © 2000 John Wiley & Sons Ltd ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic).

[9] IngYann Soon Soo Ngee Koh Cii Kiat Yeo, "Wavelet for Speech De-noising", *IEEE Tencon* - Speech and Image Technologies for Computing and Telecommunications, 1997

[10] Donoho, D.L. (1995), "De-noising by soft-thresholding," *IEEE Trans. on Inf. Theory*, Vol. 5, 2004.

[11] Ben Gold and Nelson Morgan. 'Speech and Audio Signal Processing'. John Wiley and Sons, 2000.

[12] K.P. Soman and K.I. Ramachandran, "Insight into Wavelets from Theory to Practice"2005.