A Wavelet Based Denoising of Speech Signal

* V.S.R Kumari[1] Dileep Kumar Devarakonda[2]
[1] Professor & HOD, Dept. of ECE, Sri Mittapalli College of Engineering, Guntur, A.P, India.
[2] PG Student (M. Tech), Dept. of ECE, Sri Mittapalli College of Engineering, Guntur, A.P, India.

Abstract: In this Paper we introduce an enhancement terminology in speech processing. Speech enhancement involves processing speech signals for human listening or as preparation for further processing before listening. The enhancement process aims to improve the speeches overall quality; to increase the speech intelligibility in order to reduce the listener fatigue, ambiguity etc depending on specific application. The wavelet transform plays an important role in signal analysis and widely used in many applications such as signal detection and Denoising. The basic idea behind the project is to estimate the uncorrupted speech from the distorted or noisy speech signal and sine signal, and is also referred to as speech "Denoising". There are various methods to help restore speech from noisy distortions. In this paper by using discrete wavelet transforms using different wavelet bases (Daubechies and Symlets) reduce the background noise in speech signals.

Keywords: Denoising, Wavelet transformation, Signal Processing, Noise, Speech Enhancement.

1. Introduction

During transmission and reception signals are often corrupted by noise which is unwanted signal. There are many forms of noise. One of the most common sources of noise is background noise which is always present at any location. Other types of noise include channel noise which affects both analog and digital transmission, quantization noise which results from over compression of speech signals, multi talker babble, reverberation noise or delayed version of noise are also present in some situations.

To overcome these problems we are introduce enhancement terminology in speech processing. Speech enhancement involves processing speech signals for human listening or as preparation for further processing before listening. The enhancement process aims to improve the speeches overall quality; to increase the speech intelligibility in order to reduce the listener fatigue, ambiguity etc depending on specific application. The enhancement system may be designed only to achieve one of these aims or several [3].

Speech enhancement is closely related to speech restoration. When speech is degraded, its restoration to the original speech signal often leads to speech enhancement. There are, however some important differences between enhancement and restoration. In speech restoration, an ideal speech signal is degraded and the objective is to make the processed speech signal as close as possible to the original. On the other the hand the objective of speech enhancement, is to make the processed signal sound better than the unprocessed signal. In substantiation of this it can be said that an originally un-degraded signal cannot be further restored, but can be enhanced by making it sound clearer.

Wavelets proven to be successful front end processors for speech recognition, by using the time resolution of the wavelet transform. For the speech recognition, the mother wavelet is based on the Hanning window. The recognition performance depends on the coverage of the frequency domain. The goal for good speech recognition is to increase the bandwidth of a wavelet without significantly affecting the time resolution. This can be done by compounding wavelets white noise is the most difficult to detect and to remove. The noise can easily be removed by discarding small coefficients. White noise can be handled either by hard and soft thresholding. The method is based on thresholding in the signal that each wavelet coefficient of the signal is compared to a given threshold. Using wavelets to remove noise from a signal requires identifying which components contain the noise, and then reconstructing the signal without those components.

The main objective is to minimize reconstructed error variance and maximize signal to noise ratio (SNR). Optimum wavelets can be selected based on the energy conservation properties in the approximation part of the coefficients. Wavelets with more vanishing moments should be selected as it provides better reconstruction quality and introduce less distortion into processed speech and concentrate more signal energy in few coefficients. Computational complexity of DWT increases with the number of vanishing moments and hence for real time applications it cannot be suggested with high number of vanishing moments.

The wavelet transform (WT) is a powerful tool of signal processing for its multi resolution possibilities. Unlike the Fourier transform, the WT is suitable for application to non-stationary signals with transitory phenomena, whose frequency response varies in time. The wavelet coefficients represent a measure of similarity in the frequency content between signal and a chosen wavelet function. These coefficients are computed as a convolution of the signal and the scaled wavelet function, which can be interpreted as a dilated band-pass filter because of its band-pass like spectrum.

The scale is inversely proportional to radian frequency. Consequently, low frequencies correspond to high scales and a dilated wavelet function. By wavelet analysis at high scales, we extract global information from a signal called approximations. Whereas at low scales, we extract fine information from a signal called details[5].

Signals are usually band-limited, which is equivalent to having finite energy, and therefore we need to use just a constrained interval of scales. However, the continuous wavelet transform provides us with lots of redundant information. The discrete wavelet transform (DWT) requires less space utilizing the space-saving coding based on the fact that wavelet families are orthogonal or bi-orthogonal bases, and thus do not produce redundant analysis.

The wavelet transform (WT) is a powerful tool of signal processing for its multiresolutional possibilities. Unlike the Fourier transform, the WT is suitable for application to non-stationary signals with transitory phenomena, whose frequency response varies in time. The wavelet coefficients represent a measure of similarity in the frequency content between signal and a chosen wavelet function. These coefficients are computed as a convolution of the signal and the scaled wavelet function, which can be interpreted as a dilated band-pass filter because of its band-pass like spectrum.

The scale is inversely proportional to radian frequency. Consequently, low frequencies correspond to high scales and a dilated wavelet function. By wavelet analysis at high scales, we extract global information from a signal called approximations. Whereas at low scales, we extract fine information from a signal called details.

Signals are usually band-limited, which is equivalent to having finite energy, and therefore we need to use just a constrained interval of scales. However, the continuous wavelet transform provides us with lots of redundant information. The discrete wavelet transform (DWT) requires less space utilizing the space-saving coding based on the fact that wavelet families are orthogonal or biorthogonal bases, and thus do not produce redundant analysis.

VARIOUS WAVELET USED FOR SPEECH SIGNAL ANALYSIS

Several families of wavelets that have proven to be especially useful are included in the wavelet toolbox. This paper has used three wavelets: Symlets and Daubechies wavelets are used for speech signal denoising.

Daubechies wavelet transform

The Daubechies wavelet transforms are defined in the same way as the Haar wavelet transform by computing the running averages and differences via scalar products with scaling signals and wavelets the only difference between them consists in how these scaling signals and wavelets are defined. The Daubechies wavelet is more complicated than the Haar wavelet. Daubechies wavelets are continuous; thus, they are more computationally expensive to use than the Haar wavelet,

Audio de-noising and compression is more sonically pleasing with the Daubechies wavelet than with the Haar wavelet. The Daubechies 4 filter can be used to perform the Daubechies wavelet transforms.

Symlets Wavelets

The Symlets are nearly symmetrical wavelets proposed by Daubechies as modifications to the db family. The properties of the two wavelet families are similar. There are 7 different Symlets functions from sym2 to sym8.

2. ALGORITHM IMPLEMENTATION

In Automatic 1-D Denoising, Denoising is performed using one dimensional Denoising function[1]. It performs an automatic de-noising process of a one-dimensional signal using wavelets and returns a de-noised version of input signal obtained by thresholding the wavelet coefficients. The Denoising objective is to suppress the noise part of the signal and to recover the original one[5]. The de-noising procedure proceeds in three steps:

a. Decomposition: Choose a wavelet, and choose a level N. Compute the wavelet decomposition of the signal s at level N.

b. Detail coefficients thresholding: For each level from 1 to N, select a threshold and apply soft/hard thresholding to the detail coefficients.

c. Reconstruction: Compute wavelet reconstruction based on the original approximation coefficients of level N and the modified detail coefficients of levels from 1 to N.

a. Decomposition

In this algorithm, the Daubechies and Symlets wavelet with a decomposition tree of level 3 is used; because it can provide a well orthogonality to high frequency noise with a given number of vanishing moments. The decomposition tree with wavelet coefficients at different levels, in which the boxes of approximations cA1,cA2,cA3 represents the low frequency components obtained by low pass filter, and the boxes of details cD1, cD2,cD3 represents the high frequency components obtained by high pass filter and is represented in below figure;

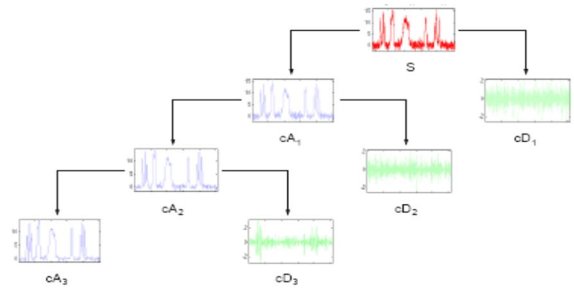

Figure 1: Three level decomposition of speech signal

b. Threshold Selection:

Wavelet thresholding is a signal estimation technique that exploits the capabilities of wavelet transform for signal Denoising. As one may observe, threshold selection is an important question when denoising. A small threshold may yield a result close to the input, but the result may still be noisy.

A large threshold on the other hand, produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness, however, destroys details and in image processing may cause blur and artifacts. Thresholding is a simple non-linear technique, which operates on one wavelet coefficient at a time.

In its most basic form, each coefficient is threshold by comparing against threshold, if the coefficient is smaller than threshold, set to zero; otherwise it is kept or modified. Replacing the small noisy coefficients by zero and inverse wavelet transform on the result may lead to reconstruction with the essential signal characteristics and with less noise. Since the work of Donoho & Johnstone there has been much research on finding thresholds. There are two types of thresholding available. They are as:

1. Hard Thresholding
2. Soft Thresholding

1. Hard thresholding

In this method input is kept only if it is larger than the threshold T, otherwise it is set to zero.

$$Y = T(X, Y) = X; \text{ for } |X| > \lambda \quad (1)$$
$$0; \text{ for } |X| \leq \lambda$$

In the hard thresholding scheme given in equation 1, the input is kept, if it is greater than the threshold λ, otherwise it is set to zero. The hard thresholding procedure removes the noise by thresholding only the wavelet coefficients of the detailed sub bands, while keeping the low-resolution coefficients unaltered.

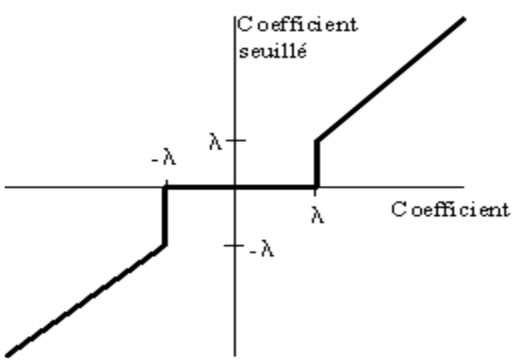

Figure 2: Characteristics of Hard Thresholding

2. Soft thresholding

It is also called shrinkage function. It takes the argument and shrinks it towards zero by the threshold T.

$$Y = T(X, Y) = \text{sign}\{X\} (|X|-1); \text{ for } |X| > \lambda$$
$$= 0 \quad ; \text{ for } |X| \leq \lambda \quad (2)$$

The soft thresholding scheme shown in equation 2 is an extension of the hard thresholding. If the absolute value of the input X is less than or equal to λ then the output is forced to zero. If the absolute value of X is greater than λ then the output is |Y| = |X - 1|. When comparing both hard and soft shrinking schemes graphically from the Figures 2 and 3, it can be seen that hard thresholding exhibits some discontinuities at λ and can be unstable or more sensitive to small changes in the data, while soft thresholding avoid discontinuities and is therefore more stable than hard thresholding.

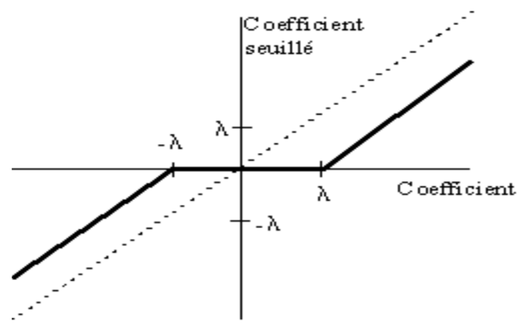

Figure 3: Characteristics of Soft Thresholding

Threshold Rules-Rigrsure

Steins unbiased risk estimator (SURE) or Rigrsure is an adaptive thresholding method which is proposed by Donoho and Jonstone and it is based on Stein's unbiased likelihood estimation principle. This method computes is likelihood estimation first using the given threshold t, and then minimize the non-like hood t, so the threshold has been obtained.

Heursure

Heursure threshold is a combination of SURE and global thresholding method. If the signal-to-noise ratio of the signal is very small, then the SURE method estimation will have more amounts of noises. In this kind of situation, the fixed form threshold is selected by means of global thresholding method.

Minimax

Minimax threshold is also used fixed threshold and it yields minmax performance for Mean Square Error (MSE) against an ideal procedures. Because the signal required the denoising can be seen similar to the estimation of unknown regression function, this extreme value estimator can realize minimized of maximum mean square error for a given function.

$$W_{tm} = \begin{Bmatrix} 0.3936 + 0.1829 * (log(n)/log(2)), & |n|>32 \\ 0 & |n|\leq 0 \end{Bmatrix} \quad (3)$$

In this method, the threshold value will be selected by obtaining a minimum error between wavelet coefficient of noise signal and original signal.

Sqtwolog

The threshold selection for Sqtwolog is calculated by following function

Wsq = sqrt (2*log (length (signal))) (4)

c. Reconstruction:

After decomposing the signal in approximate and detail coefficients, threshold value is selected by using threshold selection rule. Then by using this value, thresholding is done on detailed coefficients. After detail coefficients thresholding, signal is reconstructed by using original approximation coefficients and modified detail coefficients. Both hard and soft thresholding techniques are used to compare the efficiency of the method.

3. Results and Discussions

Objective Quality measures provide a measure based on a mathematical comparison of the original and processed speech signals that can be easily implemented and reliably reproduced using MATLAB.

A. Signal to noise ratio (SNR)

The global SNR values are determined by the ratio of square of clean speech to the square of the difference between the clean speech and the enhanced speech. If the summation is performed over the whole signal length, the operation is called as global SNR.

SNR of denoised signal can be calculated as

$$\mathrm{SNR}_{out} = 10\log_{10} \frac{\sum s^2(n)}{\sum(s(n) - s^{\wedge}(n))^2} \quad (5)$$

Where s (n) is the clean speech and sˆ (n) the enhanced signal

B. Mean Square Error (MSE)

Minimizing mean square error (MSE) between the processed speech and the clean speech is a commonly used technique in the filtering algorithms. MSE is a valid distance measure between two speeches and it is computed directly as,

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=0}^{N-1} (\hat{s}(n) - s(n))^2 \quad (6)$$

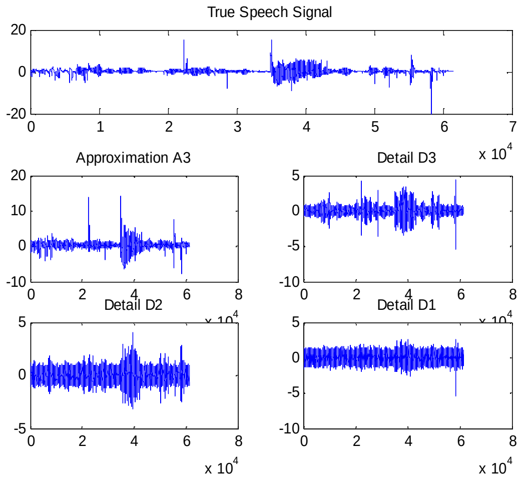

Figure 4: True speech signal and approximations using Db13 wavelet by the selection of soft thresholding (Heursure)

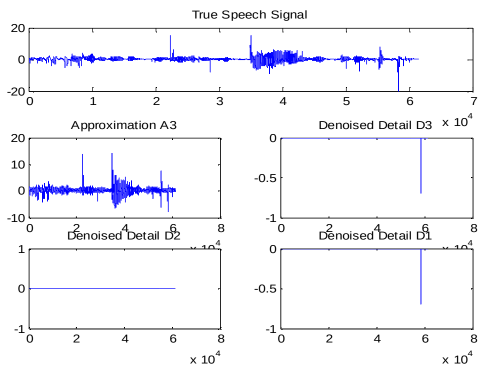

Figure 5: True speech signal and approximations of denoised speech signal using Db13 wavelet by the selection of soft thresholding (Heursure)

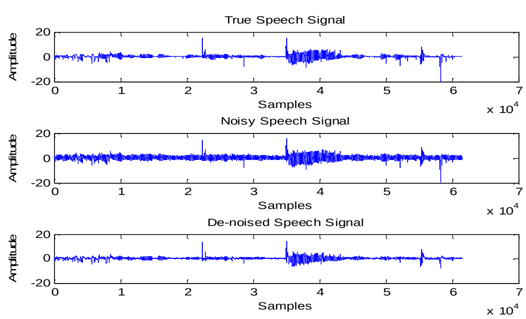

Figure 6: True, Noisy and Denoised speech signals using Db13 wavelet by the selection of soft thresholding (Heursure)

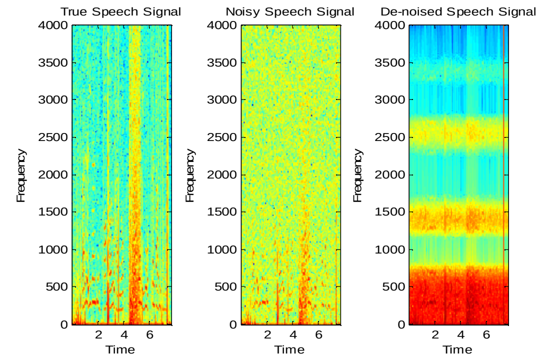

Figure 7: Spectrograms of True, Noisy and Denoised speech signals using Db13 wavelet by the selection of soft thresholding (Heursure)

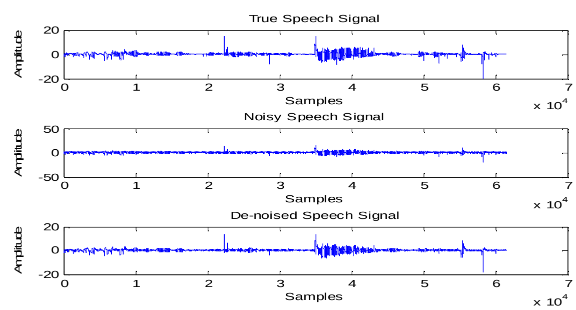

Figure 8: True, Noisy and Denoised speech signals using Db13 wavelet by the selection of hard thresholding (Heursure)

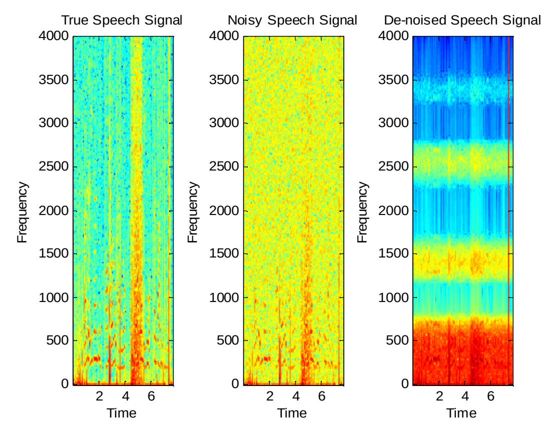

Figure 9: Spectrograms of True, Noisy and Denoised speech signals using Db13 wavelet by the selection of hard thresholding (Heursure)

For AWGN (Additive White Gaussian Noise) hard thresholding condition two wavelets namely Db13 and Sym21 have maximum Denoised SNR. In soft thresholding also Db13 and Sym13 have higher SNR than others. From above table of contents Db13 of hard thresholding is best technique for denoising of speech signal compared other type of wavelets because less error and higher denoised SNR.

Table 1 Objective Measures evaluation of different wavelets on the speech signal using soft & hard thresholding (Heursure)

Objective Measures	Db13		Db40		Sym13		Sym21	
	Soft	Hard	Soft	Hard	Soft	Hard	Soft	Hard
Error	11.6307	5.8969	12.5350	6.5713	13.5753	8.6986	13.3313	8.1828
Nosy SNR	5.0263	5.0047	4.9875	4.9857	5.0105	4.9981	5.0111	5.0259
Denoised SNR	6.2431	6.7701	6.2324	6.2511	6.2462	6.2558	6.2283	6.2568

4. Conclusion

Denoising of speech signals has been achieved successfully using wavelets. This paper provides a practical approach on how noisy audio (in wavelet form) incorporated with white Gaussian noise can be denoised. Db13 wavelet is comparatively good performance other type of wavelets.

As Daubechies is require maximum time for hard thresholding as compared to other wavelets. The biorthogonal wavelets can be used and are found to be suitable for future work done.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments which were very helpful in improving the quality and presentation of this paper.

References:

[1] S.Manikandan, "Speech enhancement based on wavelet denoising" Academic Open Internet Journal www.acadjournal.com Volume 17, 2006.

[2] G. Oppenheim J. M. Poggi M. Misiti, Y. Misiti. Wavelet Toolbox. The MathWorks, Inc., Natick, Massachusetts 01760, April 2001.

[3] S. Mallat. A Wavelet Tour of Signal Processing. Academic Press, San Diego, USA, 1998.

[4] A. K. Verma, Neema Verma. A Comparative Performance Analysis of Wavelets in Denoising of Speech Signals, National Conference on Advancement of Technologies – Information Systems & Computer Networks (ISCON – 2012) Proceedings published in International Journal of Computer Applications® (IJCA).

[5] Pradnya B. Patil et.al "A Wavelet Based Method for Denoising of Biomedical Signal", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering , March 21-23, 2012

[6] Hamid Sheikhzadeh, Hamid Reza Abutalbi,"An improved wavelet based speech enhancement system", Dept.of Electrical Eng.,Amirkabir University of technology,Tehran

[7] Soon Ing Yann "Transform based Speech Enhancement Techniques",PhD Thesis 2003, Nanyang Technological University.

[8] Pradnya B. Patil , Dr. Mahesh S. Chavan. A Wavelet Based Method for Denoising of Biomedical Signal, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 21-23, 2012

Authors Profile:

V.S.R KUMARI is working as a Professor & Head of the department of Electronics & Communication Engineering in Sri Mittapalli College of Engineering, Guntur, A.P, India. She has over 18 years of teaching experience and she is carrying out her research under Andhra University, Vishakhapatnam, AP, India.

Dileep Kumar Devarakonda is Pursuing his M. Tech from Sri Mittapalli College of Engineering, Guntur, A.P, India in the department of Electronics & Communications Engineering (ECE).