A Comparative Performance Analysis of Wavelets in Denoising of Speech Signals

A. K. Verma
Electrical Engineering Department
Dayalbagh Educational Institute
Dayalbagh, Agra, UP, India

Neema Verma
Electronics and InstrumentationDeptt.
Hindustan Institute of Technology& Management
Keetham, Agra, UP, India

ABSTRACT
It is known that data or signal obtained from the real world environment is corrupted by the noise. In most of the cases this noise is strong causing poor SNR and therefore, need to be removed from the desired signal before further processing of signal. Research in the area of wavelets showed that wavelet shrinkage method performs well and efficiently as compared to other methods of denoising. Here, we present a comparative analysis of performance of various types of wavelets i.e. Haar, Db10, Coif5, Bior3.3 and Sym5 in denoising of speech signals in the presence of White Gaussian noise. In the process of denoising, scale dependent Visu Shrink employing universal threshold selection criteria (*square-root-log)* method for deciding the threshold levels for truncating the wavelet coefficients with soft thresholding is used. Along with the performance evaluation of different types of wavelets, the effect of wavelet decomposition levels is also investigated. The quality of denoised speech signal is expressed in terms of Peak Signal to Noise Ratio (PSNR) as compared to original noiseless speech signal.

General Terms
Signal Enhancement, Denoising.

Keywords
Wavelets, PSNR, Denoising, Universal Threshold, Soft Thresholding, Level of Decomposition.

1. INTRODUCTION
Signal degradation by noise is an important phenomenon. Therefore, for various signal processing fields denoising is a key problem. All audio signals are one dimensional; their range of denoising techniques is constrained to linear. All audio samples contain well defined features which should be retained during denoising. Signals are corrupted by various kinds of noises present in environment like White Gaussian noise, Colored noise, Burst noise etc. In order to develop an efficient and effective technique of denoising audio signals, it is necessary to separate useful contents of audio from noise by suitable method. Due to various advantages wavelets have been the first choice in the area of signal denoising. An efficient method using wavelet shrinkage, which performs well over conventional frequency selective filter approach, is given in [1, 2]. An improved method using adaptive thresholding is suggested in [3]. To improve SNR, a block thresholding estimation method in presence of transients & harmonics, is suggested in [4, 5]. It adjusted parameters adaptively & minimized stein estimation of risk. The performance comparison of two mean–square-error estimators in reconstruction of signal was suggested in [6]. It shows that heuristic argument estimator removes noise effectively than Steins Unbiased Risk Estimator (SURE) by discarding purely noise coefficients in thresholding. An effective improvement in SNR can be achieved by selective smoothing at each scale of time-frequency plot thus avoiding manual selection of coefficients [7]. Further to obtain maximum signal to noise ratio & minimum error, wavelet with more vanishing moments is suggested in [8]. It gives better reconstruction quality & more signals are concentrated in only few coefficients. Noise estimation is a difficult problem, to have better estimation, a method employing Empirical Mode Decomposition (EMD) scheme is suggested in [9]. It decomposes noisy signal adaptively into Intrinsic Mode Functions (IMFs) and reconstructs signal by wavelet shrinkage on IMFs. This scheme suggests thresholding on different components of signal is more effective than on signal itself. This paper is organized in various sections, section II gives brief introduction to wavelet transform analysis, section III describes denoising scheme, section IV displays the experimental results, section V represents the acknowledgements & section VI represents conclusion followed by references. Several types of wavelets have been developed for various applications depending upon their specific properties. Thus it gives immense motivation to observe the effectiveness of these wavelets in denoising of speech signals. This paper presents a comparative analysis of performance with Haar, coif5, Db10, bior3.3 and sym5 wavelets in denoising.

2. DISCRETE WAVELET TRANSFORM (DWT)
The wavelet transform (WT) is a mathematical tool useful in the analysis of signals. Its representation involves the decomposition of the signals in wavelet basis functions $\psi(t)$given by,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) a, b \in \mathbb{R}$$

Here a, b are called *scale* and *position* parameters respectively. If scales and positions are chosen based on powers of two, so-called *Dyadic* scales and positions, then analysis becomes much more efficient and just as accurate. It was developed in 1988 by S. Mallat. In this case, wavelet function becomes,

$$\psi_{m,n}(k) = 2^{-\frac{m}{2}}\psi(2^{-m}k - n) m, n \in \mathbb{Z},$$

in orthonormal basis for $\mathbb{L}^2(\mathbb{R})$.For a given function $f(k)$, the inner product $\langle f, \psi_{m,n}\rangle$ then gives the discrete wavelet transform as, [10]

$$DWT(m,n) = \langle f, \psi_{m,n}\rangle = 2^{-\frac{m}{2}}\sum_{k=-\infty}^{\infty} f(k).\psi^*(2^{-m}k - n)$$

The multi resolution theory given by S. Mallat and Meyer proves that any conjugate mirror filter characterizes a wavelet ψthat generates an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$, and that a fast discrete wavelet transform is implemented by cascading these conjugate mirror filters. The wavelet decomposition of a signal $f(k)$ based on the multi resolution theory can done using digital FIR filters as shown in figure1 [11].

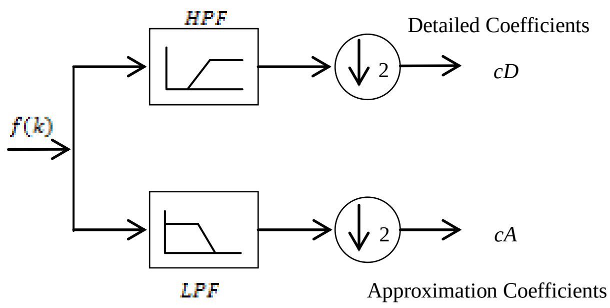


Fig. 1: One level wavelet decomposition (Analysis)

The arrangement shown above has used two wavelet decomposition (Analysis) filters which are High Pass and Low Pass respectively followed by down sampling by 2 producing half of input data point of High and Low frequency. The High frequency coefficients are called *Detailed Coefficients* (cD) and Low frequency coefficients are called *Approximation Coefficients* (cA). After decomposition, the signal can be reconstructed back by Inverse Wavelet Transform. The corresponding Filter Bank structure for reconstruction is shown in figure 2.

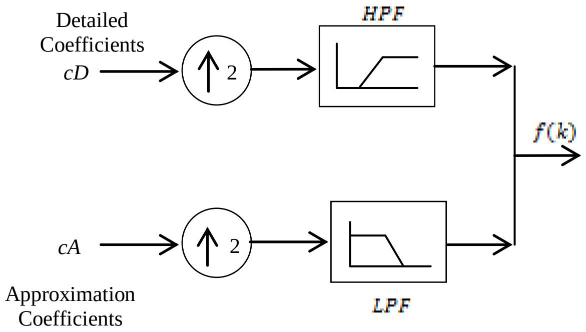


Fig. 2: One level wavelet reconstruction (Synthesis)

The signal $f(k)$can be decomposed in several levels. A three level wavelet decomposition tree is shown in figure 3.

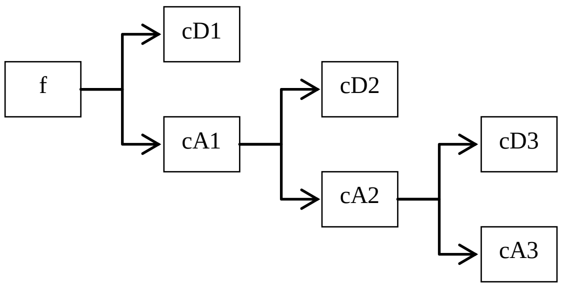


Figure 3: A 3 level wavelet decomposition tree

3. DENOISING SCHEME

The performance of various wavelets is investigated in denoising of speech signals in the environment of white Gaussian noise. Therefore, a speech signal $f(k)$ is corrupted by additive White Gaussian noise$w(k)$ as,

$$x(k) = f(k) + w(k)$$

For the purpose of denoising the noisy signal $x(k)$ is first decomposed into N levels and then level dependent soft thresholding is performed. The threshold values (λ) are calculated by universal threshold (square root log) method proposed by Donoho and Silverman [12] is given by,

$$\lambda_j = \sigma_j \sqrt{2log(N_j)}$$

Where, N_j is the length of the noisy signal at j^{th}scaleand σ_j is Median Absolute Deviation (MAD) at j^{th}scale given by,

$$\sigma_j = \frac{MAD_j}{0.6745} = \frac{median(|c|)}{0.6745}$$

Where, c represent wavelet coefficients at scale j. This is the optimal threshold in the asymptotic sense and minimizes the cost function of the difference between the function. In the proposed scheme, Soft thresholding is used over hard thresholding as hard thresholding is a “keep or kill” procedure and sometimes, pure noise coefficients may pass the hard threshold and appear as annoying ’blips’ in the output. While the soft thresholding shrinks coefficients above the threshold in absolute value. The continuity of soft thresholding and some advantages such as it makes algorithms mathematically more tractable. Moreover, Soft thresholding avoids these annoying ‘blips’ in output. The transfer functions of soft and hard thresholding is shown in figure 4.

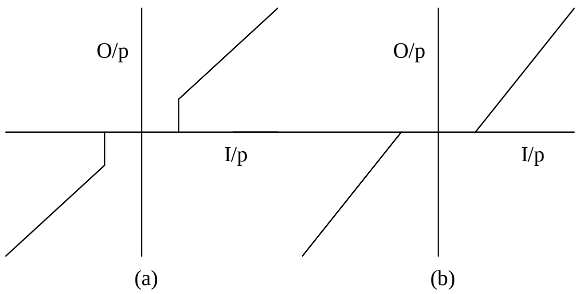


Fig. 4: (a) Hard Thresholding, (b) Soft Thresholding

Threshold determination is an important problem. When denoising, a small threshold may yield a result closer to the input, but the result may still be noisy. A large threshold on other hand produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness destroys details (Sharp changes) of signal.

4. EXPERIMENTAL RESULTS

As samples, five speech signals each of size 10 seconds duration sampled at 8000 samples per second are analyzed for experiment. These five speech samples contain three of English language (f_1, f_2, f_3) and two of Hindi language (f_4, f_5). For comparing the performance of various wavelets in speech signals following five wavelets, Haar, Db10, Coif5, bior3.3 and sym5 are used. Besides observing the performance of various wavelets, the effect of decomposition level is also investigated.
For the performance comparison and measurement of quality of denoising, the Peak Signal to Noise Ratio (PSNR) is calculated between original speech signal $f(k)$and denoised speech signal $f_d(k)$ given by,

$$PSNR = 10log_{10}\left(\frac{f_{max}{}^2}{MSE}\right)$$

Where, $f_{max}$is maximum value of signal and is given by,

$$f_{max} = max\left(max\left(f(k)\right), max\left(f_d(k)\right)\right)$$

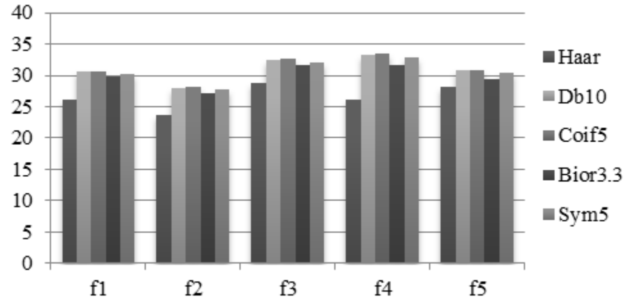
And MSE is mean Square Error given by,

$$MSE = \frac{1}{N}\sum_{k=1}^{N}[f_d(k) - f(k)]^2$$

PSNR values for various speech signals are shown comparatively in figure 5 to figure 8.

Signals	Haar	Db10	Coif5	Bior3.3	Sym5
f_1	26.18	30.56	30.67	29.80	30.27
f_2	23.67	27.86	28.19	27.23	27.82
f_3	28.77	32.53	32.61	31.69	32.10
f_4	26.09	33.36	33.45	31.66	32.93
f_5	28.21	30.82	30.84	29.32	30.49

(a)

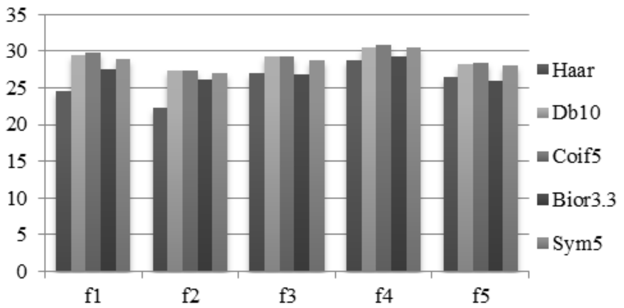


(b)

Figure 5 (a, b): Comparison of PSNR at level 2 decomposition

Signals	Haar	Db10	Coif5	Bior3.3	Sym5
f_1	24.58	29.48	29.72	27.51	28.93
f_2	22.32	27.32	27.29	26.13	27.07
f_3	26.97	29.23	29.30	26.75	28.82
f_4	28.78	30.55	30.89	29.30	30.48
f_5	26.51	28.13	28.35	26.03	28.05

(a)

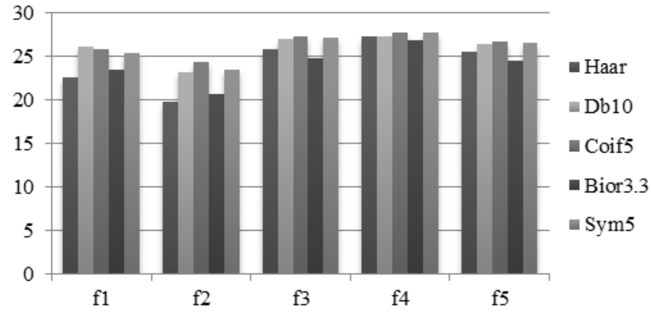


(b)

Figure 6 (a, b): Comparison of PSNR at level 3 decomposition

Signals	Haar	Db10	Coif5	Bior3.3	Sym5
f_1	22.49	26.06	25.77	23.43	25.38
f_2	19.78	23.10	24.26	20.68	23.41
f_3	25.84	27.03	27.33	24.77	27.08
f_4	27.28	27.29	27.76	26.84	27.75
f_5	25.50	26.44	26.69	24.47	26.51

(a)

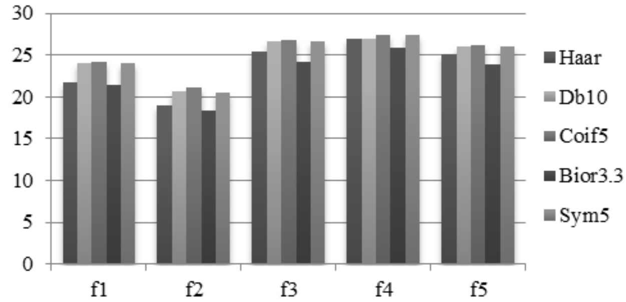


(b)

Figure 7 (a, b): Comparison of PSNR at level 4 decomposition.

Signals	Haar	Db10	Coif5	Bior3.3	Sym5
f_1	21.78	24.05	24.17	21.45	23.97
f_2	18.96	20.61	21.19	18.42	20.57
f_3	25.50	26.61	26.76	24.18	26.69
f_4	26.88	26.91	27.36	25.86	27.43
f_5	25.14	26.02	26.23	23.81	26.07

(a)



(b)

Figure 8 (a, b): Comparison of PSNR at level 5 decomposition

5. CONCLUSIONS

In this paper, performance of various wavelets in denoising process is investigated along with effect of level of decomposition in the environment of white Gaussian noise. The method for threshold estimation used is level dependent Visu Shrink employing universal threshold.

The results show that as level of decomposition increases the value of PSNR decreases. Thus lower levels of decomposition can be preferred for such purposes. As seen from the result, performance of Coif5 and Db10 is better than as compared to other wavelets in case of all decomposition levels. The Haar wavelet is performing poorer as compared to others also it gives unwanted distortion in reconstructed voice when heard as it is not a smooth wavelet. Overall from the results, it is clear that Coif5 at level 2 is best suitable giving minimum distortion and maximum PSNR as compared with other wavelets.

6. REFERENCES

[1] Donoho D. L. and Johnston I.M., "'De-noising by soft-thresholding', *IEEE* Transaction on Information Theory, 1995, 41, (3). pp. 613-627.

[2] D. L. Donoho and I. M. Johnstone "Minimax Estimation via wavelet shrinkage" Annals of Statistics, 26(3):879 921, 1998.

[3] Tie Caiand Xing Wu, "Wavelet Based De-Noising of Speech using Adaptive Decomposition".

[4] Guoshen Yu, Stephane Mallat, Emmanuel Bacry, "Audio Denoising by Time-Frequency Block Thresholding", IEEE transactions on Signal Processing, Vol. 56, No. 5, May 2008.

[5] S. Sreekanth, P. Dinesh Khanna, P. Uma Mallikarjuna, "An Efficient Noise Reduction by using Diagonal and Non diagonal Estimation Techniques", Proceedings of the International Conference on Communication and Computational Intelligence, 2010, pp.393-398.

[6] Hamid Krim, Dewey Tucker, Stephane Mallat and David Donoho,"On Denoising and Best Signal Representation" IEEE Transactions on Information Theory, Vol. 45, No. 7,

[7] A. Sony John, Uday B. Desai, "Signal De-Noising using the Wavelet Transform and Regularization", 1997, IEEE.

[8] Mahesh S. Chavan, Manjusha N. Chavan, M. S. Gaikwad, "Studies on Implementation of Wavelet for Denoising Speech Signal", International Journal of Computer Applications (0975 – 8887), Volume 3, No.2, June 2010.

[9] Kais Khaldi, Abdel-Ouahab Boudraa, Abdelkhalek Bouchikhi, Monia Turki-Hadj Alouane and El-Hadji Samba Diop, "Speech Signal Noise Reduction by EMD", ISCCSP 2008, Malta, 12-14 March 2008.

[10] Daubechies Ingrid, "Ten lectures on wavelets", 9e, SIAM, ISBN: 780-89871-274-2, 2006.

[11] Mallat Stephane, "A Wavelet Tour of Signal Processing", 3e, 2008, Academic Press, ISBN 978-01-237-4370-1.

[12] Johnston I. M., Silverman B. W., "Wavelet Threshold Estimators for data with correlated Noise", J. Roy. Statist. Soc. Ser. B 59, 319–351.