

Studies on Implementation of Wavelet for Denoising Speech Signal

Dr. Mahesh S. Chavan

Professor, Department of Electronics & Telecommunication,
Rajarambapu Institute of Technology,
Rajaramnagar Sakharale (MS) India.

Mrs Manjusha N.Chavan

Lecturer, Department of Electronics & Telecommunication,
Rajarambapu Institute of Technology,
Rajaramnagar Sakharale (MS) India.

Dr. M.S.Gaikwad

Professor, Department of Electronics & Telecommunication,
Sinhgad Academy of Engineering,
Kondwa, Pune(MS) India.

ABSTRACT

The wavelet transform has become a powerful tool of signal analysis and is widely used in many applications which include signal detection and denoising. In hands free speech communication environments situation occurs that speech is superposed by background noise. Over the past few decades there is tremendous increase in the level of ambient environmental noise. This has been due to growth of technology. Noise is added by various factors like noisy engines, heavy machines, pumps, vehicles, over noisy telephone channel or using radio communication device in an aircraft cockpit. The wavelet denoising technique is called thresholding; it is a non linear algorithm. It can be decomposed in three steps. This paper is based on wavelet as denoising algorithm. Haar and Daubechies wavelets are implemented on speech signals and performance is evaluated.

Categories and Subject Descriptors

Speech signal Processing, Wavelets, Denoising.

General Terms

Types of noise, Harr, Daubechies wavelets, Thresholding.

Keywords

Speech Signal, Harr wavelet, Daubechies Wavelet, Spectrogram.

1. INTRODUCTION

Generally enhancement of speech can be used to increase the quality of speech processing device like digital hearing aids, mobiles and other man machine interface in our daily life to make them more robust in noisy environmental conditions. Speech is a very basic way for humans to convey information with the emotion of a human voice to one another. People use speech to communicate messages. [1] Human speech can be modeled as filter acting on excitation waveform. Speech can be divided into voiced and unvoiced. Voiced speech has a spectrum with energy concentrated at discrete frequencies i.e. at the fundamental frequency of the vocal folds and its multiples (harmonics). About one third of speech is completely a periodic (unvoiced) resulting from a random excitation that resembles white noise, caused by air rapidly passing through a narrow constriction in vocal tract. [2] There often occur conditions under which we measure and then transform the speech signal to another form in order to enhance our ability to communicate with a band width of only 4 kHz. With the advent of the wonders of digital technology, the analog to digital converter samples the electrical speech e.g. 8000 samples per second for telephone speech, so that speech signal can be digitally transmitted and processed. 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.[5] Denoising has long been a focus of research and yet there always remains room for improvement. Speech enhancement in general has three main objectives 1) to improve perceptual aspects such as quality and intelligibility of the processed speech i.e. to make it sound better so that to reduce listener's fatigue.2) To improve robustness of speech coder which tend to be severely affected by the presence of noise.3) to increase accuracy of speech recognition systems operating in noisy environments. [5]

The problem of denoising consists of removing noise from corrupted signal without altering it. Fourier domain was long been the method of choice to suppress noise [3].The classical methods based on spectral subtraction are effective for this purpose; however they introduce artificial noise and alter the original signal. The utilization of wavelets in signal and image processing has been found to be a very useful tool for solving various engineering problems, denoising is one of them. Early transform domain approach leave behind annoying musical noise. Later methods solve problem to some extent.

1.1 Choice of wavelet

Choosing a wavelet that has compact support in both time and frequency in addition to significant number of vanishing moments is essential for an algorithm. Several criteria can be used in selecting an optimal wavelet function. The 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.

2 RESULTS AND IMPLEMENTATION OF WAVELET:

The speech signals used for the work are pronounced by male and female speakers, recorded using sound recorder facility with external microphone using mono channel. Samples used are 11000. Signal is sampled at sampling frequency $F_s=8000$ Hz,

encoded using 16 bits, and degraded by additive Gaussian white noise. Signal is corrupted by 5db, 10 db and 15 db additive noises. Thus we have the noisy signal in required SNRs. In case of additive background noise the assumptions made for developing enhancement methods are (i) speech and noise signals are uncorrelated at least over a short-time basis, (ii) noise is either stationary or slowly varying as compared to speech, and (iii) noise can be represented as zero mean random process. The degradation level of additive background noise is normally specified by the measure called Signal to Noise Ratio (SNR) and is defined as the ratio of signal energy to noise energy. For evaluating performance of the method both objective and subjective tests are conducted. In objective test, SNR of signal after denoising is computed. Other two parameters used for comparing results are time required for reconstruction of signal and mean square error between clean signal and denoised signal. Haar and Daubechies wavelets are implemented on noisy signal and effort is made to remove additive white Gaussian noise from noisy signal. Signal is decomposed to level 4 and level 5.

Let $s(n)$ is the clean speech, $y(n)$ the noisy, $\hat{s}(n)$ the enhanced signal and $w(n)$ the noise then we have:

$$y(n) = s(n) + w(n) \quad 1$$

SNR of denoised signal can be calculated as

$$SNR_{out} = 10 \log_{10} \frac{\sum s^2(n)}{\sum (s(n) - \hat{s}(n))^2} \quad 2$$

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,

$$MSE = \frac{1}{N} \sum_{n=0}^{N-1} (\hat{s}(n) - s(n))^2 \quad 3$$

Speech signal pronounced by male speaker is as shown in figure and respective spectrogram of '.wav' file is as shown. Results are shown for level 4 decomposition of noisy signal with Haar wavelet implementation.

3 RESULTS AND DISCUSSION

For AWGN (Additive White Gaussian Noise) hard thresholding condition two wavelets namely Db14 and Db16 have maximum SNR. Both of these need more time as compared with others. In soft thresholding also Db8 and Db18 have same and higher SNR than others. Db8 with very less reconstructing time gives highest SNR for hard and Db16 with 0 MSE and more time has higher SNR for soft in random noise. Db14 has maximum SNR and all other have similar SNR response with more or less time for reconstruction. Db16, Db18 need maximum reconstruction time and Haar needs least time for hard thresholding of AWGN. Db12 to Db20 require same time to reconstruct signal out of which Db18 is better for soft thresholding. If hard thresholding is applied to signal corrupted by random noise Db16 performs better with MSE of 0.002. Db20 takes maximum and haar takes minimum time to reconstruct. For soft thresholding Db20 gives maximum SNR. Db16 also gives approximately close SNR with same reconstructing time and less MSE than DB20. For AWGN hard thresholding Db16 has highest SNR but needs long time to reconstruct signal. DB20 and haar have same SNR values where DB20 needs much more time than haar. Here haar is found to be performing well. For soft thresholding of AWGN Db16 gives maximum SNR. Db14, Db16 have same SNR which is approximately close to SNR of Db12. In case of hard thresholding for random noise Db14 has highest SNR value. Haar and Db12 have 0 MSE and all other wavelets have

some error in reconstructed signal as compared to original signal. In soft threshold method, signal Db20 has better results than Db10. Reconstructed signals which are denoised are as shown in following figures. Male and female speaker signals corrupted by white noise decomposed to level 4 and level 5 and reconstructed after thresholding the coefficients. Following figures indicate results for speech signal of male speaker with level 4 decomposition.

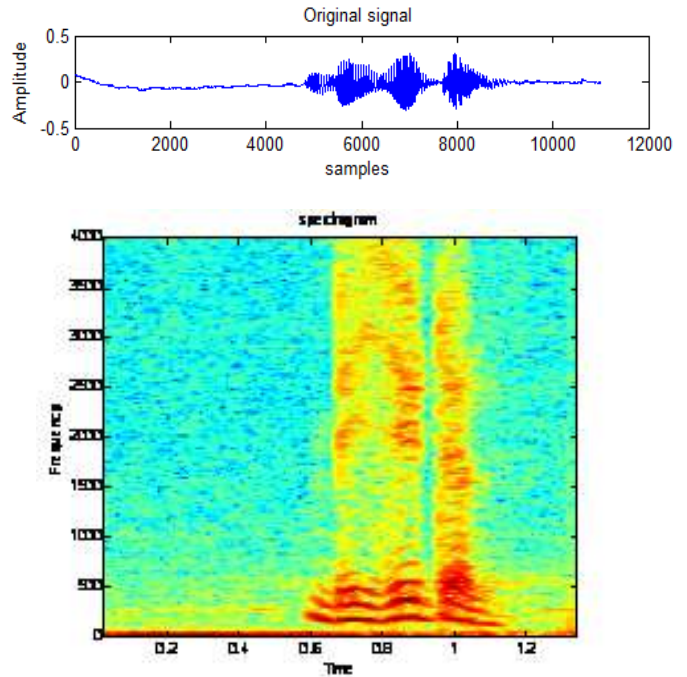


Figure 1. Original speech signal and its spectrogram

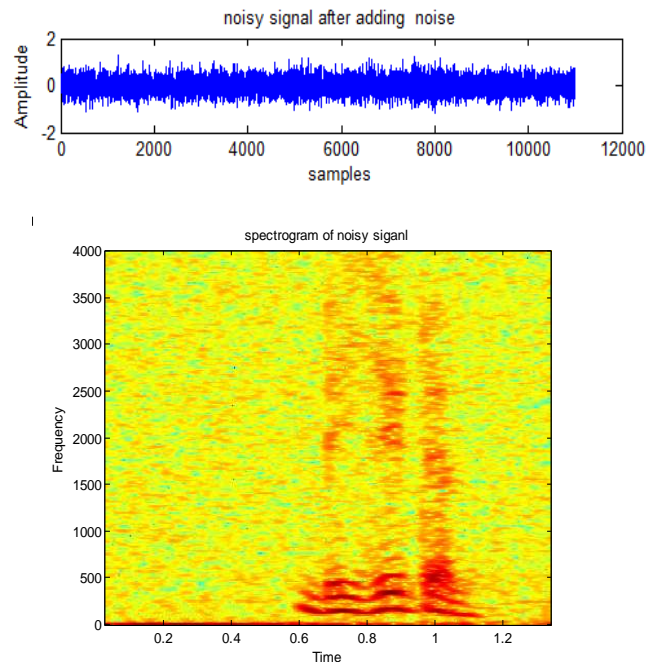


Figure 2 Noisy signal corrupted by 10 db AWGN and its Spectrogram

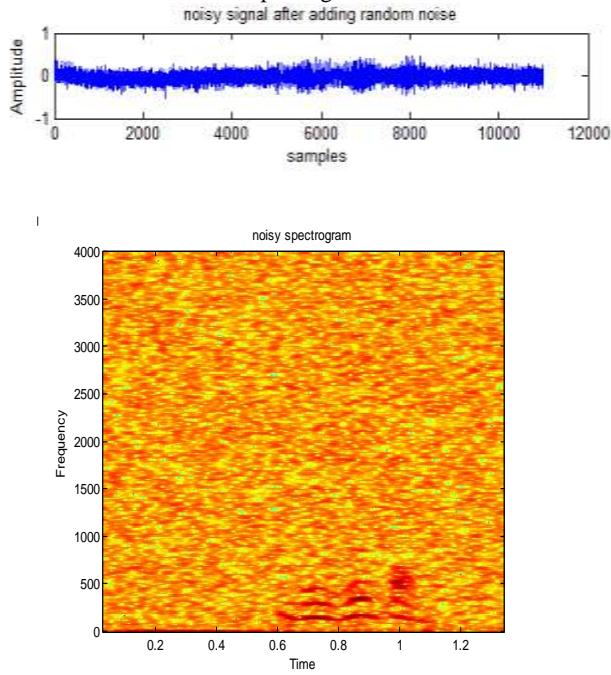


Figure 3 noisy signal corrupted by of 10 db random noise and Spectrogram of noisy signal.

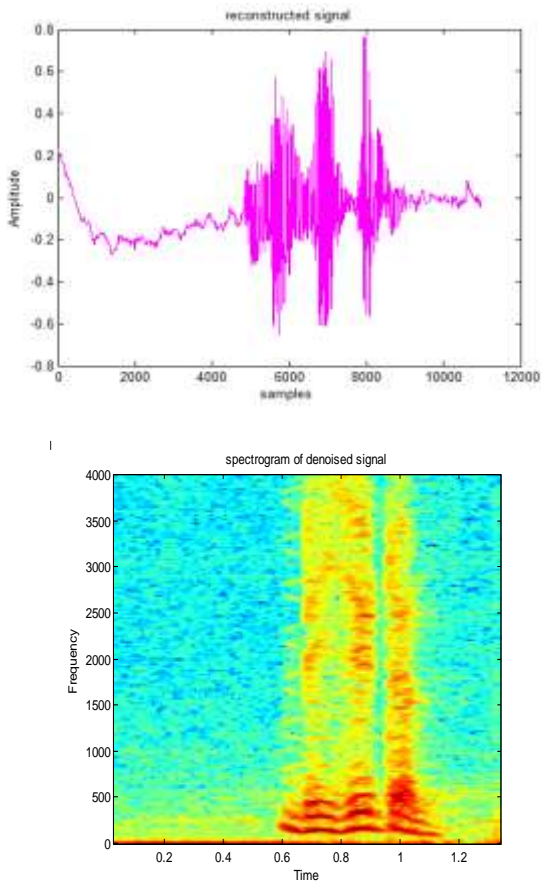


Figure 4 Reconstructed signal using Haar wavelet for 10 db AWGN

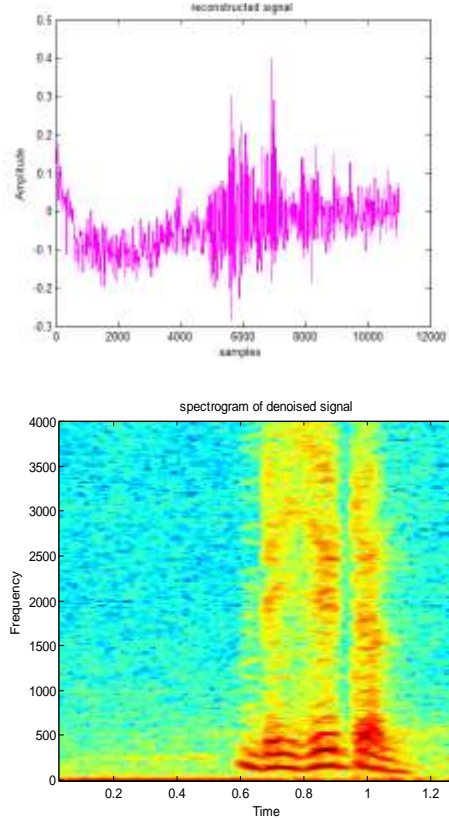
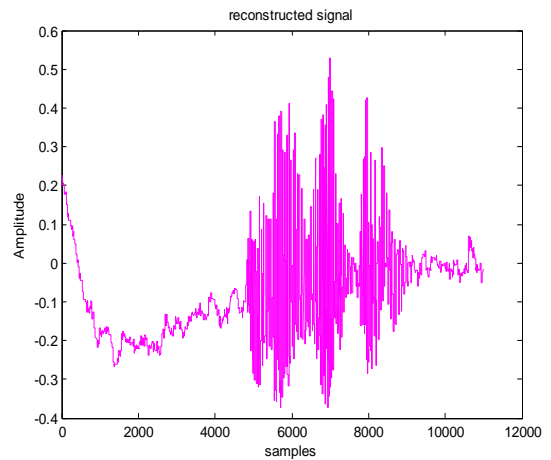


Figure 5 Reconstructed signal using Haar wavelet for random noise 10 db

From figure 4 it is observed that denoised signal is approximately same as of original signal with Haar wavelet for hard thresholding in additive white Gaussian noise. For random noise of 10 db for the same speech signal reconstructed signal has some noise which is tolerable as shown in figure 5. When same speech signal is denoised using soft thresholding, SNR of reconstructed signal improves. Reconstructed signal using soft thresholding is as shown in figure 6 and 7.



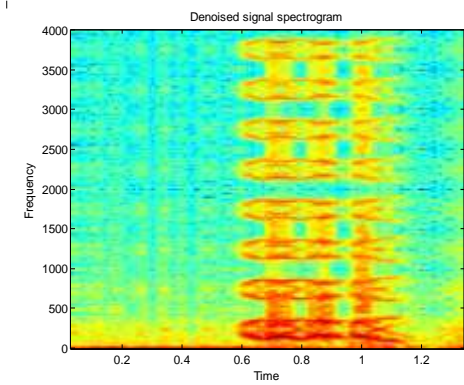


Figure 6 Reconstructed signal for AWGN of 10 db for soft thresholding

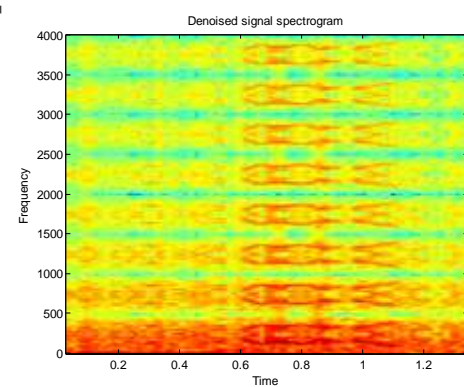
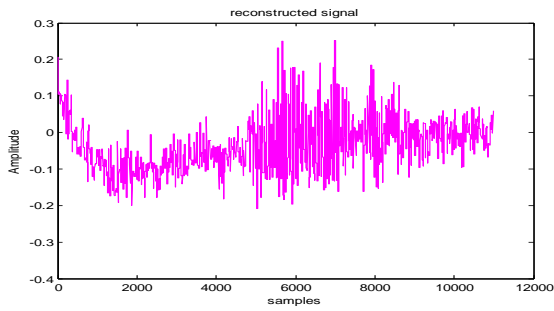


Figure 7 Reconstructed signal for Random noise of 10 db for soft thresholding

Following are the results of signal decomposed up to level5 with Haar wavelet implementation for hard as well as soft thresholding

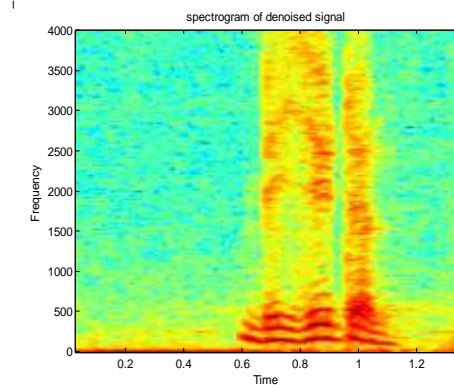
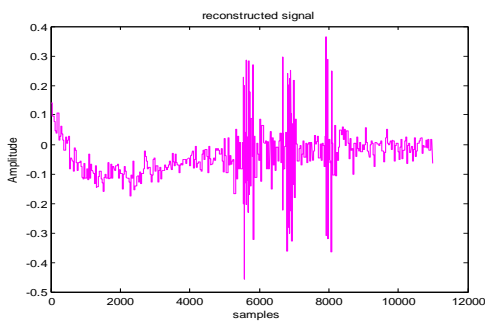


Figure 8 Reconstructed signal for random noise of 10 db using hard threshold

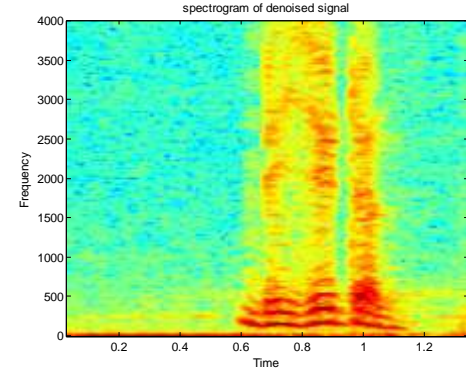
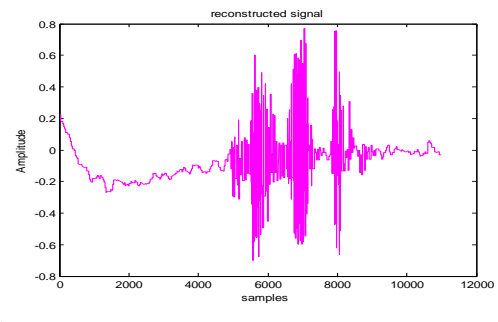
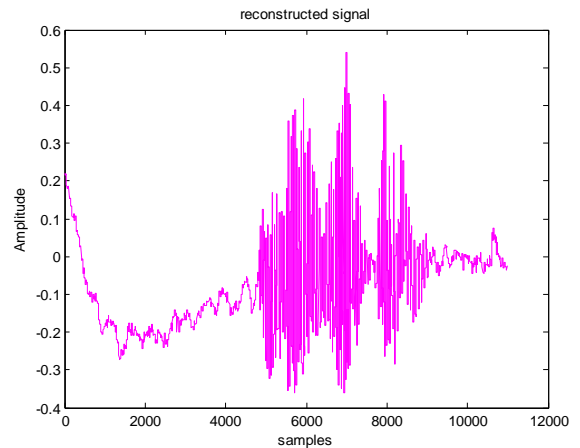


Figure 9 Reconstructed signal for AWGN 10 db
Following are the results of implementation of soft thresholding



IMPLEMENTATION OF DAUBECHIES WAVELET

Following figures show signals reconstructed by hard thresholding and decomposed up to level 4 using db18 wavelet. Reconstructed signal seems to be close to the original signal. SNR is also around 24 db. Signal was corrupted using random noise of 10 db. It is observed that Daubechies wavelet needs more time to reconstruct signal.

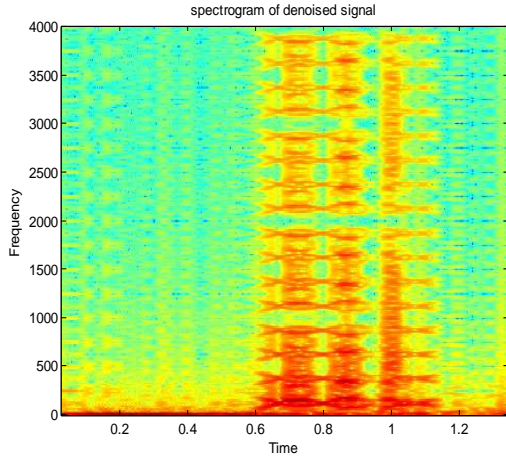


Figure 10 Reconstructed b1.wav for AWGN 10 db

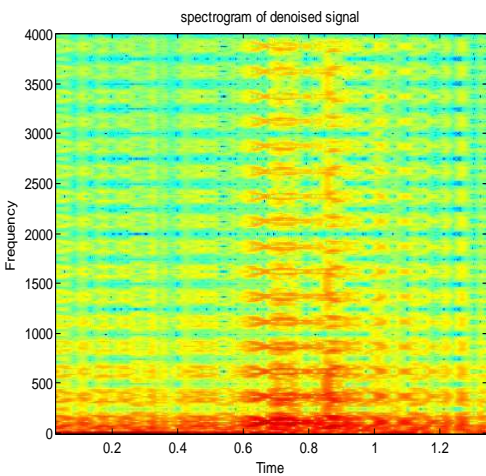
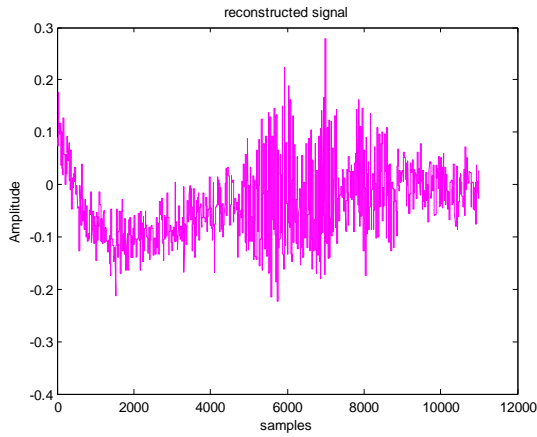


Figure 11 Reconstructed signal for Random 10 db

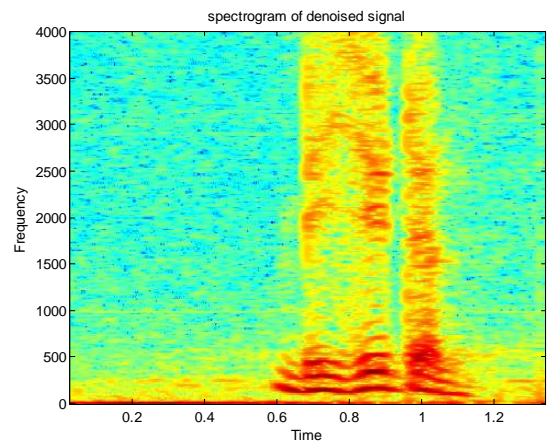
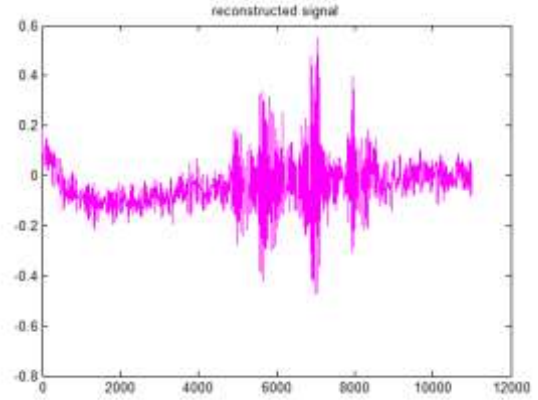
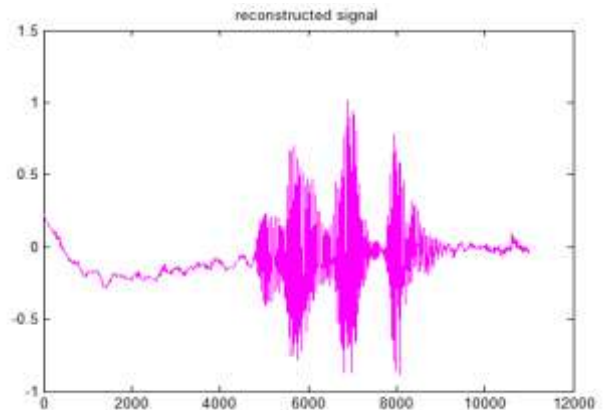


Figure 12 Reconstructed signal b1.wav with hard thresholding using db18



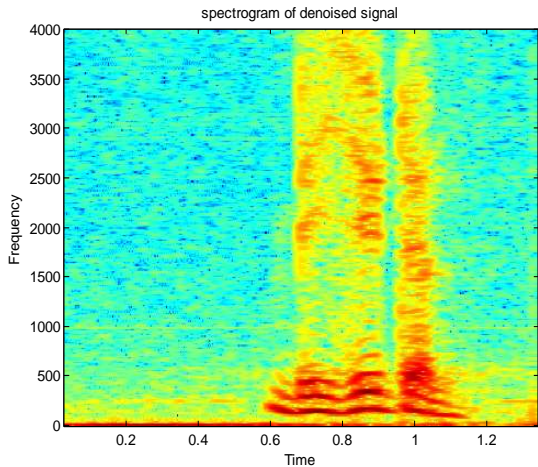


Figure 13 Reconstructed signal b1.wav with hard thresholding using db18 for 10 db AWGN

When denoised using soft thresholding, reconstructed signal is found to be similar to the original signal. Soft thresholding performs better than hard thresholding. Reconstructed signals are as shown in figure 14.

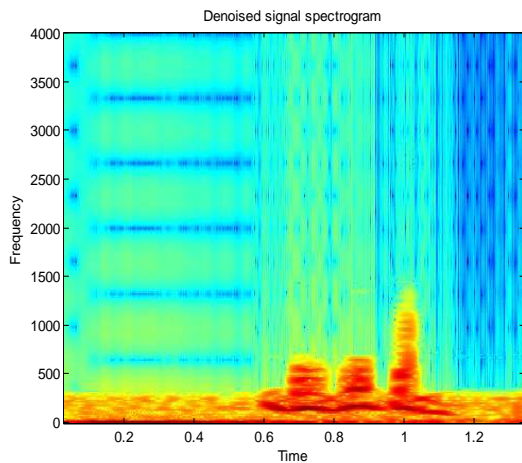
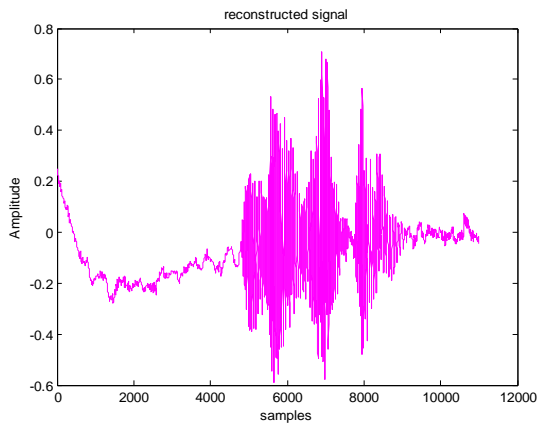


Figure 14 Reconstructed signal with soft thresholding using db18 for AWGN

Figure 15 shows reconstructed signal which was corrupted with 10 db random noise. From figure it is clear that still noise is present in reconstructed signal and not totally removed, spectrogram also do not match to that of original and is quite blurred & indicates presences of noise

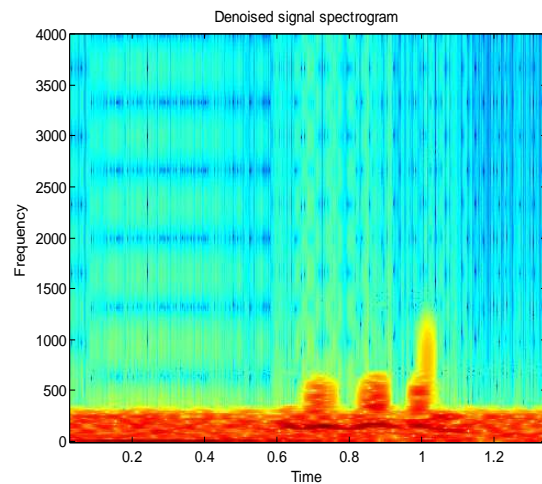
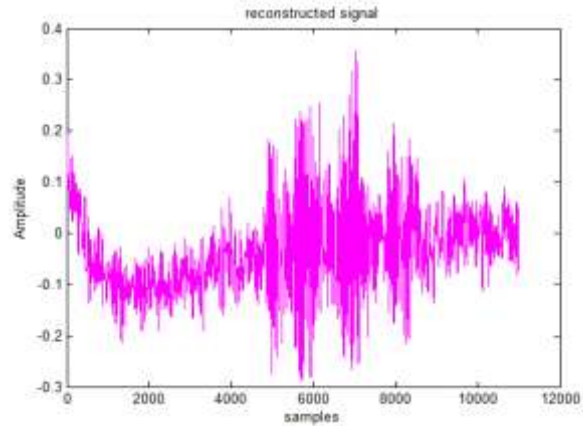


Figure 15 Reconstructed b1.wav with soft thresholding for random noise using db18

SUMMARY AND CONCLUSION

Wavelet denoising is a non-parametric estimation method that has been proposed in recent years for speech enhancement applications. In Present work both objective and subjective methods used for evaluation of wavelets performance in speech denoising confirms its superiority. Haar wavelet has comparatively good performance than some other wavelets from Db family. SNR for level4 is better than for level5. If SNR obtained for speech signals in any noise with all type of wavelets is below 19 db then it is not properly denoised and still some noise components are remained.

Subjective test was conducted with a group of 10 people with normal hearing capability. They were asked to listen carefully noisy signal and reconstructed signal which is our signal of interest (denoised signal). Throughout the test SNR of noise was set to 10 dB. Listeners preferred denoised signal with soft thresholding against hard thresholding. In some cases denoised signal has less strength than original but still it is audible.

From results obtained it can be concluded that Haar wavelet is not suitable for speech denoising application. As Haar is not smooth as compared to other wavelets it has limitations when applied to non stationary signal such as speech. Higher order Daubechies can be used and are found to be suitable for the work done. Also soft thresholding is better than hard thresholding.

REFERENCES

- [1] S.Manikandan, "Speech enhancement based on wavelet denoising" Academic Open Internet Journal www.acadjournal.com Volume 17, 2006.
- [2] Hamid Sheikhzadeh, Hamid Reza Abutalbi,"An improved wavelet based speech enhancement system", Dept.of Electrical Eng.,Amirkabir University of technology,Tehran
- [3] Lallouani, M. Gabrea and C.S. Gargour, "Wavelet based Speech enhancement using different threshold based denoising algorithms", Canadian conference, Electrical and Computer Engg., May 2004.
- [4] Amara Graps, "An introduction to Wavelets",IEEE Computation Science and Engineering.
- [5] Soon Ing Yann "Transform based Speech Enhancement Techniques",PhD Thesis 2003, Nanyang Technological University.
- [6] Chris Perkins, Tobin Fricke, "Wavelets", University of California at Berkely.
- [7] "Speech Compression using Wavelets", By Nikhil Rao, B.E Thesis.
- [8] Yan Long, Lin Gang and Guo Jun," Selection Of The best Wavelet Base For Speech Signal", Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, October 20-22, 2004.
- [9] Qin Linmei, Hu Guangrui and Li Chongni," A new speech enhancement method",Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and speech processing, May 2-4 2001, Hongkong.
- [10] V. Balakrishnan, Nash Borges, Luke Parchment," Wavelet Denoising and Speech Enhancement", Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore.