Speech Enhancement Using Super Soft Thresholding in Wavelet Domain

R. Santhoshkumar, Dr. B. Kirubagari
Department of CSE & Annamalai University
Tamil Nadu, India

Abstract— Speech is being a fundamental way of communication among human beings. In many unavoidable situations, unwanted background noises are added to the speech signal. The proposed speech enhancement technique is to remove the background noise and to improve the quality of the speech signal. Noisy signal are decomposed by wavelet decomposition technique. Super soft thresholding technique is applied to the decomposed signal to remove the background noise. The thresholded signal can be reconstructed by wavelet reconstruction technique. The performance of the noisy signal and denoised signal can be measured using SNR (Signal to Noise Ratio). The proposed super soft thresholding algorithm can achieve better performance, when compared to hard or soft thresholding algorithm.

Keywords— Decomposition; Reconstruction; SNR; Speech signal; Super soft Thresholding.

I. INTRODUCTION

Speech is a fundamental and common medium, hence important for us, to communicate. Advancement in technology have made way for many more speech oriented applications like cellular voice calls, VoIP, teleconferencing systems, speech recognition, and hearing aids, etc. In many cases, these systems work well in nearly noise-free conditions, but their performance deteriorates rapidly in noise conditions. In general, there exists a need to increase the reliability of these systems in noisy environments. Therefore, improvement in existing pre-processing algorithms or introducing entire new class of algorithm for speech enhancement is the basic objective of research community. Noise can be classified into many types depending on the nature and properties of the noise sources. Additive noises like background noise, impulse noise, speaker interfering noise and non additive noises like speaker stress, non-linearity's of microphones etc. affect the quality of the speech produced. In speech enhancement, the goal is used to improve the quality of degraded speech.

Wavelet Transforms are used in various research areas including signal and image denoising, data compression and classification problems. The wavelet coefficients are denoised using wavelet denoising techniques which use soft thresholding. The wavelet coefficients of the noise and the target signal are separated using a boundary called the threshold, which is estimated depending on standard rules. But simple threshold can suppress the noise only up to an extent. Soft thresholding can be applied for further noise reduction. The principle underlying the wavelet-based methods is similar to the subspace concept. The wavelet based methods achieve noise reduction through thresholding, which relies on the fact that only a few significant wavelet coefficients contribute to the signal synthesis. In 1995, wavelet thresholding (shrinking) was introduced by Donoho as a powerful tool in denoising signals degraded by additive white noise. Although the application of wavelet shrinking for speech enhancement has been reported in several works, there are many problems yet to be resolved for a successful application of the method to speech signals degraded by various noise types.

In this project, we present a new system for speech enhancement. The core of our system is an improved Wavelet thresholding that uses the speech signal features for improved performance, The proposed system also uses a time adaptive threshold selection method that selects the current time interval threshold depending on the estimate of the energy of the clean speech signal for current frame. The proposed system exploits the advantage of a new thresholding algorithm that yields fewer artefacts and better subjective result in comparison with hard or soft thresholding algorithms. A further advantage of the proposed system is that, unlike most other wavelet-based algorithms in which the detection of unvoiced segments affects their performances, it does not require any voiced/unvoiced detection method.

This paper is organized as follows: Section II focuses on the wavelet thresholding and block diagram of the wavelet denoising method. In Section III, we discuss about thresholding algorithm. Section IV describes the detailed description of the proposed super soft thresholding method and Section V shows the experiments and results analysis of the proposed super soft thresholding. Finally, conclusions and acknowledgement are given.

II. WAVELET THRESHOLDING

The principle under which the wavelet thresholding operates is similar to the subspace concept, which relies on the Speech is a fundamental and common medium, hence important for us, to communicate. Advancement in technology

have made way for many more speech oriented applications like cellular voice calls, VoIP, teleconferencing systems, speech recognition, and hearing aids, etc. In many cases, these systems work well in nearly noise-free conditions, but their performance deteriorates rapidly in noise conditions. In general, there exists a need to increase the reliability of these systems in noisy environments. Therefore, improvement in existing pre-processing algorithms or introducing entire new class of algorithm for speech enhancement is the basic objective of research community. Noise can be classified into many types depending on the nature and properties of the noise sources. Additive noises like background noise, impulse noise, speaker interfering noise and non additive noises like speaker stress, non-linearity's of microphones etc. affect the quality of the speech produced. In speech enhancement, the goal is used to improve the quality of degraded speech.
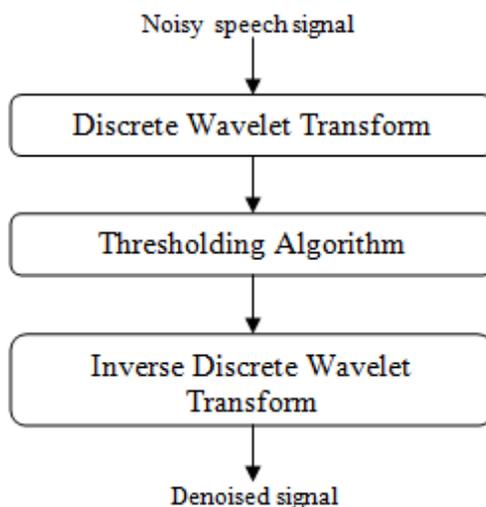
A. Block Diagram

Fig. 1: Denoising by wavelet thresholding block diagram

Wavelet Transforms are used in various research areas including signal and image denoising, data compression and classification problems. The wavelet coefficients are denoised using wavelet denoising techniques which use soft thresholding. The wavelet coefficients of the noise and the target signal are separated using a boundary called the threshold, which is estimated depending on standard rules. But simple threshold can suppress the noise only up to an extent. Soft thresholding can be applied for further noise reduction.   The principle underlying the wavelet-based methods is similar to the subspace concept. The wavelet based methods achieve noise reduction through thresholding, which relies on the fact that only a few significant wavelet coefficients contribute to the signal synthesis. In 1995, wavelet thresholding (shrinking) was introduced by Donoho as a powerful tool in denoising signals degraded by additive white noise. Although the application of wavelet shrinking for speech enhancement has been reported in several works, there are many problems yet to be resolved for a successful application of the method to speech signals degraded by various noise types.

The proposed denoising algorithm is summarized as follow:
  i) Compute the discrete wavelet transform for noisy signal.
  ii) Based on an algorithm, called thresholding algorithm and a threshold value, shrink some detail wavelet coefficients.
  iii) Compute the inverse discrete wavelet transform.

Wave shrink, which is the basic method for denoising by wavelet thresholding, shrinks the detail coefficients because these coefficients represent the high frequency components of the signal and it supposes that the most important parts of signal information reside at low frequencies. This assumption is true for a large group of signals; this idea is also based on the sparsity characteristic of wavelet transform. Therefore, the assumption is that in high frequencies the noise can have a bigger effect than the signal. In other words, wave shrink supposes that at high frequencies, noise forms a bigger part of coefficients in comparison with low frequencies.

Wave shrink has drawbacks for signals like speech, because in some parts of speech like consonants or unvoiced regions the high frequencies that are in detail sections of wavelet transform contain important information that influence the quality and intelligibility of speech signal, for example [8] mentioned that in many languages, consonant phonemes carry more semantic information rather than the vowels. The other disadvantage is that the basic Wave shrink method just enhances the objective tests and hence for applications such as speech enhancement where both objective and subjective tests are important, it is not particularly useful, for example sometimes listeners prefer the noisy speech to the enhanced speech with the basic Wave shrink method. In this project, we develop an improved wavelet thresholding as the core of our system that is customized for speech enhancement.

B. Discrete Wavelet Transform
In Discrete Wavelet Transform, a signal can be expressed in both time and frequency representation. A signal can be analysed and reconstructed with DWT using its multi-resolution filter banks and special wavelet filters [6]. The main

characteristics of the wavelet transforms are that they can use windows of varying size, which is broad at low frequencies and narrow at high frequencies. This gives an optimal time–frequency resolution in all frequency ranges. In DWT, the original signal passes through 2 filters namely a low-pass filter and a high-pass filter and produces 2 signals called approximation (low frequency) coefficients and detail (high frequency) coefficients. In speech signals, approximation coefficients are of more importance than detail coefficients because they represent the characteristics of a signal more. Selection of the wavelet family and hence wavelets plays an important role in signal denoising. In this work, we have used the most popular wavelets called the Daubechies wavelets that are found to be efficient in speech processing applications. The main criterion for selecting an optimal wavelet function is to reduce reconstructed error variance and to increase SNR. If the number of vanishing moments is more, it causes complexity. But they provide better performance in reconstruction and cause less distortion into the processed speech signals. So here we have used wavelets with more vanishing points.

C. Inverse Discrete Wavelet Transform
A process by which components can be assembled back into the original signal without loss of information. This process is called reconstruction, or synthesis and the mathematical manipulation that effects synthesis is called the IDWT. IDWT reconstructs a signal from the approximation and detail coefficients derived from decomposition. The IDWT differs from the Discrete Wavelet Transform (DWT) in that it requires upsampling and filtering, in that order. Upsampling, also known as interpolating, means the insertion of zeros between samples in a signal.

$$idwt(cA, cD, wavelet[, mode='sym'[, correct\_size=0]])$$

The idwt() function reconstructs data from the given coefficients by performing single level Inverse Discrete Wavelet Transform.

- cA – Approximation coefficients.
- cD – Detail coefficients.
- Wavelet – Wavelet to use in the transform.
- Mode – Signal extension mode to deal with the border distortion problem.
- Correct_size – Typically, cA and cD coefficients lists must have equal lengths in order to perform IDWT.

III. THRESHOLDING ALGORITHM
A. Wavelet Denoising using Hard Thresholding
There are two popular thresholding functions used for denoising signals using wavelets namely hard and soft thresholding functions. In hard thresholding, elements whose absolute values are less than the threshold is set to 0.Hard thresholding can be expressed as

$$X_{Hard} = \begin{cases} x & \text{if} & |x| > \tau \\ 0 & \text{if} & |x| \leq \tau \end{cases} \tag{1}$$

B. Wavelet Denoising using Soft Thresholding
In soft thresholding, the elements whose absolute values are lower than the threshold are first set to zero. Then the nonzero coefficients are shrinked towards 0. Soft thresholding can be expressed as

$$X_{Soft} = \begin{cases} sign(x)(|x| - |\tau|) & \text{if} & |x| > \tau \\ 0 & \text{if} & |x| \leq \tau \end{cases} \tag{2}$$

C. Proposed Super-Soft thresholding algorithm
The proposed Super-Soft thresholding algorithm avoids forcing the wavelet coefficients smaller than the threshold to zero but instead replaces them by a fraction of their original values.

$$X_{Super-Soft} = \begin{cases} sign(x)(a|x|) & \text{if} & |x| \leq \tau \\ sign(x)(|x| - \tau) & \text{if} & |x| > \tau \end{cases} \tag{3}$$

Where X represents the wavelet coefficients and is the threshold value. Here we have used soft thresholding technique. The value of is taken as the universal threshold developed by Donoho and Jonstone. which is defined as

$$\tau = \sigma\sqrt{2\log(N)} \tag{4}$$

where σ is the standard deviation and N is the length of the signal.

Suppose x(t) is the original signal and the noise added is n(t). Then a signal y(t) can be represented as the summation of the original signal and the noise as y(t) = x(t) + n(t) [11][12].

IV. PROPOSED SUPER-SOFT THRESHOLDING ALGORITHM
In the proposed Super-Soft thresholding algorithm instead of setting some wavelet coefficients to zero, the algorithm attenuates the coefficients depending on their distance from the threshold. This idea is based on the fact that forcing some wavelet coefficients to zero causes observable sharp time-frequency discontinuities in the speech spectrogram [17] that can decrease the quality of the enhanced speech signal. The proposed Super-Soft thresholding algorithm avoids forcing the wavelet coefficients smaller than the threshold to zero but instead replaces them by a fraction of their original values. From mathematical point of view we use a slope as (3):

$$sign(x)(a|x|) \qquad if \quad |x| \leq \tau \tag{5}$$

where a is the line slope for the values smaller than threshold, so it should be a small value. To avoid discontinuity for the values bigger that the threshold, we continue this slope to cross the soft Thresholding algorithm for the values greater than threshold , so we have:

$$y = sign(x)(a|x|) \quad if \quad |x| \le \tau \tag{6}$$
$$y = sign(x)(|x| - \tau) \quad if \quad |x| > \tau \tag{7}$$

After solving this equation the cross point will be the Point $\left(x = \frac{\tau}{1-a}, y = \frac{a\tau}{1-a}\right)$, that we used it as the threshold point. For the values greater than the cross point, this method is similar to soft thresholding algorithm, that itself has a better performance for speech than hard thresholding, because it tries to improve the wavelet coefficients greater than threshold which will not be changed by hard thresholding. Therefore, the Super-Soft thresholding algorithm is defined as follow:

$$X_{\text{Super-Soft}} = \begin{cases} sign(x)(a|x|) & if \quad |x| \le \tau \\ sign(x)(|x| - \tau) & if \quad |x| > \tau \end{cases} \tag{8}$$

In our experimental results, we will see that this thresholding algorithm has much better SNR than hard or soft thresholding for the enhanced speech.

A noisy speech corpus (NOIZEUS) was developed to facilitate comparison of speech enhancement algorithms among research groups. The noisy database corrupted by six different real-world noises at different SNRs. The noise are taken from the NOIZEUS database and includes airport, babble, car, exhibition, restaurant, street noise with four different DB (0db, 5db, 10db, 15db). This corpus is available to researchers free of charge.

V. EXPERIMENTS AND RESULTS

A. Evaluation using SNR

Computational, this is the simplest test, but the most un- reliable one . let, s(t), z(t), $\hat{s}$(t) be the clean, corrupted and enhanced speech signal, respectively, and T the sample size. Define by:

$$SNR_{in} = 10 log 10 \frac{\sum_{t=1}^{T} s^2(t)}{\sum_{t=1}^{T} [s(t) - z(t)]^2} \tag{9}$$

$$SNR_{out} = 10 log 10 \frac{\sum_{t=1}^{T} s^2(t)}{\sum_{t=1}^{T} [s(t) - \hat{s}(t)]^2} \tag{10}$$

The SNR levels in the input and in the output of the evaluated enhancer. Define by the difference:

$$G = SNR_{out} - SNR_{in} \tag{11}$$

These noisy speech examples are at different input SNRs equal to 0dB, 5dB, 10dB and 15dB.

TABLE 1: Speech signal corrupted by Airport

DECIBLES	VALUE FOR ENHANCED SIGNAL
0dB	1.1794
5dB	6.0775
10dB	10.7947
15dB	15.2741

TABLE 2: Speech signal corrupted by Car noise

DECIBLES	VALUE FOR ENHANCED SIGNAL
0dB	0.7002
5dB	5.5894
10dB	10.9582
15dB	16.2438

TABLE 3: Speech signal corrupted by Babble noise

DECIBLES	VALUE FOR ENHANCED SIGNAL
0dB	0.5536
5dB	5.3017
10dB	10.831
15dB	16.3785

TABLE 4: Speech signal corrupted by Exhibition noise

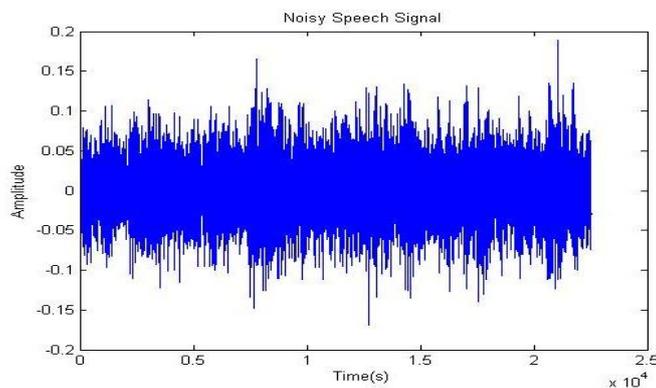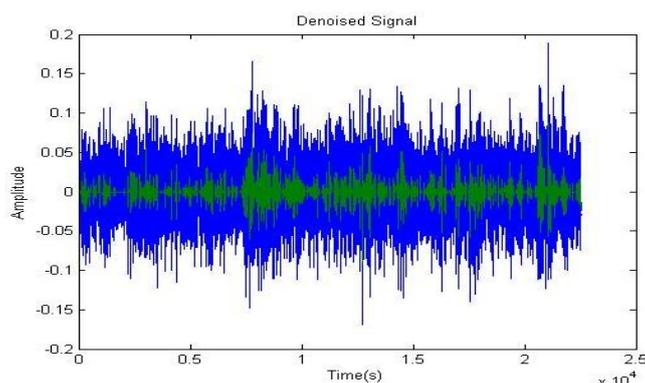| DECIBLES	VALUE FOR ENHANCED SIGNAL
0dB	0.9563
5dB	6.1772
10dB	11.9573
15dB	16.1129

Fig. 2: Noisy Speech Signal

Fig. 3: Enhanced Speech Signal

VI.   CONCLUSION

In this paper, an improved adaptive wavelet thresholding speech enhancement system, which uses the proposed Super-Soft thresholding algorithm, improves the noisy speech wavelet coefficients in a way that avoids sharp time-frequency discontinuities in the speech spectrogram that can decrease the quality of the enhanced speech signal. This system also uses the estimation of the clean speech signal energy for each frame to select the threshold for the thresholding algorithm of the current frame.

A further advantage of this algorithm is that unlike most of the other wavelet-based algorithms in which the detection of unvoiced segments highly affects their performances, the proposed method does not require any voiced/unvoiced detection method. By applying many tests, we evaluated our technique by giving different types of noisy signal (airport, babble, car, exhibition, restaurant, street).The results confirmed the improvements in performance and achievements of our work.

REFERENCES

[1]   D. O'Shaughnessy. Speech Communication: Human and Machine,IEEE press: Addison-Wesley Publishing Co; 1999.

[2]   Byung-Jun Yoon, P. P. Vaidyanathan. Wavelet-based denoising by customized thresholding, IEEE International Conference on Acoustics, Speech and Signal Processing; 2004; 925-928.

[3]   Mohammed Bahoura, Jean Rouat. Wavelet speech enhancement based on time-scale adaptation, Speech Communication; Vol. 48: Issue 12: 2006; 1620–1637.

[4]   Hadhami Issaoui, Aïcha Bouzid, Noureddine Ellouze. Comparison between Soft and Hard Thresholding on Selected Intrinsic Mode Selection, IEEE conference on Sciences of Electronics, Technologies of Information and telecommunications; 2012; 1-5.

[5]   Slavy G. Mihov, Ratcho M. Ivanov, Angel N. Popov. Denoising Speech Signals by Wavelet Transform, Annual Journal Of Electronics; 2009; ISSN 1313-1842.

[6]   Mahesh S. Chavan, Manjusha N.Chavan, M.S.Gaikwad. Studies on Implementation of Wavelet for Denoising Speech Signal, International Journal of Computer Applications; Vol. 3: No.2: 2010; 1-7

[7]   Matko Saric, Luki Bilicic, Hrvoje Dujmic. White Noise Reduction of Audio Signal Using Wavelets Transform With Modified Universal Threshold, R. Boskovica B. B Hr 21000 Split, Croatia. 2005

[8]   Elif Derya Ubeyil. Combined Neural Network model employing wavelet coefficients for ECG signals classification, Digital Signal Processing; Vol 19: 2009; 297-308.

[9]   S. Kadambe, P. Srinivasan. Application of adaptive wavelets for speech, Optical Engineering; Vol 33(7): 1994; 2204-2211.

[10]   D.L. Donoho. De-noising by soft thresholding, IEEE transactions on information theory; Vol. 41: no. 3:1995; 613-627.

[11]   Yasser Ghanbari, Mohammad Reza Karami. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets, Speech Communication; Vol. 48 (8): 2006; 927–940.

[12]   Tie Cai, Xing Wu. Wavelet-Based De-Noising of Speech Using Adaptive Decomposition, Proc. of IEEE International Conference On Industrial Technology; 2008; 1-5.