IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 11, Issue 1, Ver. III (Jan. - Feb .2016), PP 26-35
www.iosrjournals.org

Denoising Of Speech Signal By Classification Into Voiced, Unvoiced And Silence Region

Varshika Meshram[1], Preety D Swami[2]
[1](Department of Electronics & Communication, Samrat Ashok Technological Institute, Vidisha, India)
[2](Department of Electronics & Instrumentation, Samrat Ashok Technological Institute, Vidisha, India)

Abstract: *In this paper, a speech enhancement method based on the classification of voiced, unvoiced and silence regions and using stationary wavelet transform is presented. To prevent the quality of degradation of speech during the denoising process, speech is first classified into voiced, unvoiced and silence regions. An experimentally verified criterion based on the short time energy process has been applied to obtain the voiced, unvoiced and silence speech region. The stationary wavelet transform has been applied to each segment and compared with a pre-determined energy threshold. Different threshold methods have been applied for voiced, unvoiced and silence frames. The performance of the algorithm has been evaluated using a large speech database from the TIMIT database corrupted with white Gaussian noise for SNR levels ranging from -10 to 10 dB. Algorithm is compared with popular denoising methods and it is found that improved SNR is obtained through the use of semi-soft thresholding for unvoiced speech, modified hard thresholding for voiced speech and wavelet coefficients of silence region are set to zero.*

Keywords: *Speech Processing, White Gaussian noise, Short time energy, Voiced/unvoiced and silence speech region, Stationary Wavelet Transform, Semi-soft thresholding*

I. Introduction

In many speech processing applications such as teleconferencing system, voice communication, speech coding, automatic speech recognition applications, and hearing aid systems, speech has to be processed in the presence of unwanted background noise. These applications need removal of background noise and recover the original signal from noisy signal. Speech enhancement is an important technique in speech signal processing fields. It removes noise and improves the quality and intelligibility of speech communication by using several algorithms [1,2,3]. In the past few decades, various algorithms have been adopted for this objective [4,5,6,7]. Earliest speech denoising techniques were linear methods where the most common methods were Wiener filtering [6], spectral subtraction [5,6] and subspace filtering [7]. These methods have easy implementation and design. Therefore, many researchers were attracted to this approach.

Generally, the speech enhancement technique can be divided into two basic categories of single microphone and multi microphone method. Although multi microphone method has excellent performance in some applications, there are still many practical situations where one is limited to use a single microphone in presence of non-stationary noise. Several single microphone methods for speech enhancement are available. The spectral subtraction method is the oldest method for speech enhancement. This method is introduced by Boll and Berouti [4,5]. This linear method is capable to reduce the background noise and improve the SNR, but it usually introduces an annoying musical noise. When we look more into speech processing with the aid of an interesting technology, another method known as Wiener filtering comes into the picture. It was originally formulated by Lim and Oppenhiem [6]. This filtering method has the drawback that the estimation criteria are fixed. Another speech enhancement technique is signal subspace approach. This technique was introduced by Ephraim [7]. Signal subspace technique eliminates musical noise originating from fluctuating energy. It improves estimates by averaging over long windows. However, another disturbance source exists. These include rapid changes of model order and subspace swapping. The latter condition refers to noise basis vectors being incorrectly employed to describe the signal subspace. These methods are still gaining attention of many researchers to increase its performance [3].

Nowadays, a non-linear approach using wavelet transform has emerged as a powerful technique for removing noise from signal. The utilization of wavelets in signal processing has been found as a very useful tool for solving various problems; denoising is one of them. The wavelet denoising technique is also known as wavelet thresholding (shrinking). It is based on the assumption that in wavelet representation, magnitude of signal dominates the noise magnitude. To shrink the wavelet coefficient of noise a threshold is selected. Then an inverse wavelet transform is made on the residual coefficients to recover the original speech.

Donoho and Johnstone [8,9] employed a new algorithm based on wavelet thresholding (shrinking) for denoising the signal corrupted by white Gaussian noise. The application of wavelet thresholding for speech

enhancement has been reported in several works [10]. There are still many problems to be resolved for enhancement of speech signal corrupted by different types of noise.

To preserve useful information in a signal and eliminate as much noise as possible, this paper presents a speech enhancement system, which works in the wavelet domain and is based on classification of speech into voiced, unvoiced and silence regions. It uses some specific features of the speech signal to improve the performance. One of the methods of feature extraction is the short time energy that determines whether it is voiced, unvoiced or silence regions.

The rest of the paper is organised as follows: Section 2 gives an overview of various wavelet transforms. Section 3 gives an overview of denoising by using thresholding. Section 4 explains the method of determination of voiced, unvoiced and silence regions. Section 5 contains an implementation of our speech enhancement algorithm. Section 6 and 7 include the criteria of evaluation and experimental result respectively. Finally, section 8 gives the overall conclusion of the work carried out.

II. Wavelet Transform

Wavelet transform decomposes a signal into a set of frequency bands by projecting the signal onto an element of a set of basic functions. The numerous types of wavelets are scaled versions of the mother wavelet. This requires one filter to be designed and others will follow the scaling rules in both the time and the frequency domain. Since the processing of signals in the frequency domain is easier to implement, most of the speech enhancement algorithm is implemented in the frequency domain. The Advantage of wavelet transform is the use of variable size time windows for different frequency bands. This gives a high frequency resolution in low bands and low frequency resolution in higher bands. Due to subband processing, wavelet transform can provide an appropriate model of speech signal, which gives better performance. The wavelet transform can be categorized into two classes; Continuous Wavelet Transform & the Discrete Wavelet Transform.

2.1 Continuous Wavelet Transform (Cwt)

The continuous wavelet transform is an implementation of wavelet transform using arbitrary scale and almost arbitrary wavelet. The wavelet coefficients obtained are a function of scale and position. The continuous wavelet transform of signal x(t) is defined as [11]:

$$CWT(a,\tau) = \int f(t)\,\psi_{a,\tau}(t)\,dt \qquad (1)$$

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\varphi\left(\frac{t-\tau}{a}\right)$$

Where a and τ are the scale and transition parameters respectively, and φ(t) is the mother wavelet. The wavelets are contracted (a < 1) or dilated (a > l), and moved with time shift over the signal. The inverse transform of continuous wavelet transform also exists.

2.2. Discrete Wavelet Transform (Dwt)

Discrete wavelet transform [18,19], is an implementation of the wavelet transform using a discrete set of wavelet scales and translations. In other words, this transform decomposes the signal into mutually orthogonal sets of wavelets. Scale and translation axis are based on powers of two, so called dyadic scale and translation. The advantage of DWT over CWT is that it is comparatively faster, easier to implement and avoids redundancy.

2.3. Stationary Wavelet Transform (Swt)

Stationary wavelet transform is designed to overcome the drawback of translation and invariance of the discrete wavelet transform. It is also known as un-decimated wavelet transform. Stationary wavelet transform does not decimate coefficients at each transformation level. This wavelet transform is a modification of discrete wavelet transform. The basic idea of stationary wavelet transform is to fill in the gap caused by the decimation step in the standard wavelet transform. This leads to over-decimation or redundant representation of the original signal. We simply apply high pass and low pass filters to the data at each level to produce two sequences at the next level. The two new sequences, each have the same length as the original sequence [12].

III. Denoising By Thresholding

Denoising by wavelets is performed by thresholding algorithm, in which wavelet coefficients that are smaller than a specific value, i.e. threshold will be scaled [8, 13]. In speech signal, energy is mostly concentrated in a small number of wavelet dimensions. The coefficients of these dimensions are large compared to other dimensions like noise. Noise has energy spread over a large number of coefficients. Setting smaller coefficients to zero eliminates noise while keeping important information of the original signal. In this section,

the most commonly used thresholding algorithms are reviewed. Thresholding algorithms that have a better performance of speech signals are introduced.

3.1 Threshold Algorithm

3.1.1 Hard Thresholding Algorithm

Hard thresholding function keeps the wavelet coefficients that are greater than the given threshold and sets components of the noise to zero. Donoho and Johnson used it in [9] for wavelet thresholding as:

$$T_{Hard}(x) = \begin{cases} x & |x| > \lambda \\ 0 & |x| \leq \lambda \end{cases} \quad (2)$$

In this thresholding algorithm, the wavelet coefficients that are less than the threshold (λ) will be replaced with zero, as shown in the Fig. 1

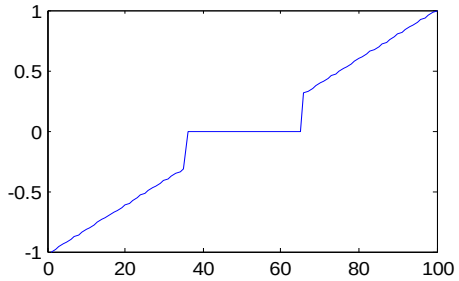

Fig. 1 Hard Thresholding

The hard thresholding algorithm in the wavelet domain is not continuous at threshold λ. This is the disadvantage of hard thresholding function.

3.1.2. Soft Thresholding Algorithm

The soft thresholding algorithm has different rule from the hard thresholding. The Soft thresholding function is defined as follow [8]:

$$T_{Soft}(x) = \begin{cases} sign(x)(|x| - \lambda) & |x| > \lambda \\ 0 & |x| \leq \lambda \end{cases} \quad (3)$$

The soft thresholding is shown in the Fig. 2

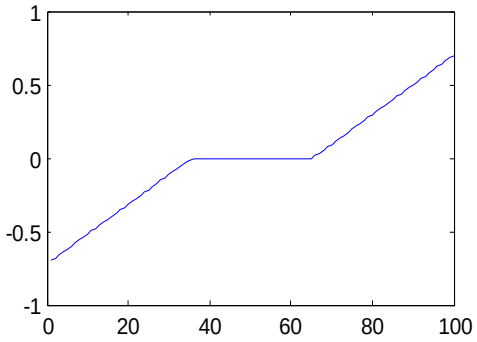

Fig. 2 Soft Thresholding

3.1.3 Semi-Soft/Firm Thresholding Algorithm

Semi-soft thresholding algorithm corrects the disadvantages of hard and soft thresholding algorithm with respect to the variance and bias of the estimated value. Geo [14] developed this algorithm. The equation of semi-soft thresholding can be given as:

$$\delta_{\lambda}^{F}(x) = \begin{cases} 0 & |x| \leq \lambda_1 \\ sign(x)\left(\frac{\lambda_2(|x| - \lambda_1)}{\lambda_2 - \lambda_1}\right) & \lambda_1 < |x| \leq \lambda_2 \\ x & |x| > \lambda_2 \end{cases} \quad (4)$$

Where λ_1 and λ_2 denote lower and upper threshold values, respectively. Fig. 3 shows the semi-soft thresholding algorithm.

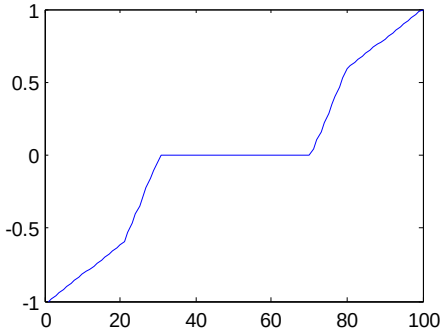

Fig. 3 Semi-Soft Thresholding

3.1.4 Modified Hard Thresholding Algorithm

We have used a modified version of hard thresholding algorithm instead of standard hard thresholding algorithm. In modified hard thresholding, we apply a nonlinear function to the threshold value. The proposed algorithm is defined as:

$$\delta = \begin{cases} x & |x| \geq \lambda \\ sign(x) * \left(\frac{|x|^{\gamma}}{\lambda^{\gamma-1}}\right) & |x| < \lambda \end{cases} \qquad (5)$$

Fig. 4 shows the input-output characteristics of the modified hard thresholding algorithm. The γ parameter can be determined by optimization.

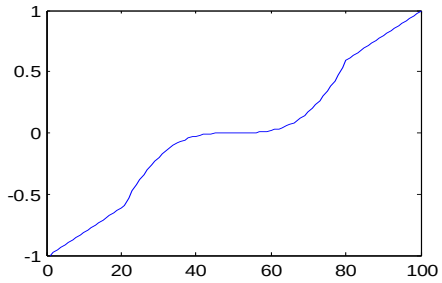

Fig. 4 Modified Hard Thresholding

4. Voiced, Unvoiced And Silence Determination

Speech signal can be divided into voiced, unvoiced and silence regions. The silence region has the lowest energy as compared to voiced and unvoiced regions. By detection of silence region, we can remove noise completely by setting the wavelet coefficients of the silence region to zero. The voiced speech is produced because of excitation of the vocal tract by the periodic flow of air at the glottis [15]. It has the highest energy as compared to unvoiced and silence regions because of its periodicity.

Unvoiced speech is non-periodic and is caused when air passes through a narrow constriction of the vocal tract when consonants are spoken [15] and it shows lower energy than voiced speech. The waveform of the unvoiced speech is similar to noise and it will suffer from being eliminated in wavelet thresholding.

The classification of speech into voiced, unvoiced and silence provides a preliminary acoustic segmentation for the speech processing application. In speech enhancement system, voiced, unvoiced and silence regions are enhanced separately. To separate these regions, short time energy parameter is used. The short time energy can be mathematically expressed as:

$$E_n = \sum_{n=-\infty}^{\infty} [x(n)w(n-m)]^2 \qquad (5)$$

Where x(n) represents speech signal and w(n) is the window frame.

IV. Methodology

The proposed speech enhancement system has been illustrated using block diagram of Fig. 5.

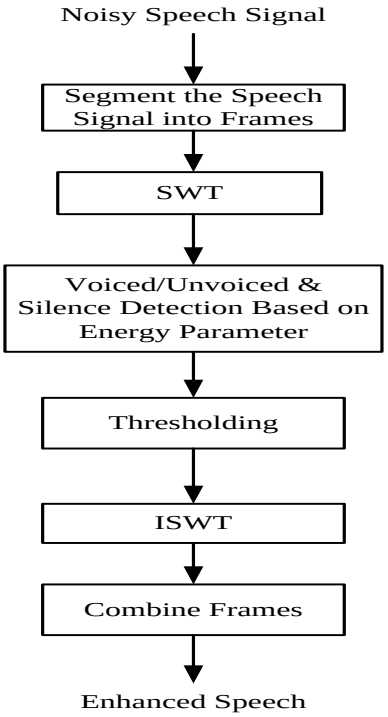

Fig. 5 Block Diagram of the Proposed Speech Enhancement System

First, the speech signal is divided into several short-time segments which is also known as framing. Each frame has 250 samples at 16 kHz sampling rate. Then stationary wavelet transform has been applied to each frame. Accordingly, for each frame the energy of each frame is calculated. The frame with the largest energy will be considered as voiced and the frame with the energy less than a certain threshold will be considered as unvoiced signal. Silence part has lower energy than voiced and unvoiced speech. The above mentioned thresholding function is then applied to the wavelet coefficients of each voiced and unvoiced frames and the wavelet coefficient of the silence frame are set to zero. Fig. 6 shows the graph of threshold value with respect to input SNR. After thresholding, inverse stationary wavelet transform of the thresholded coefficients is computed. Finally the frames are combined to generate the complete denoised signal.

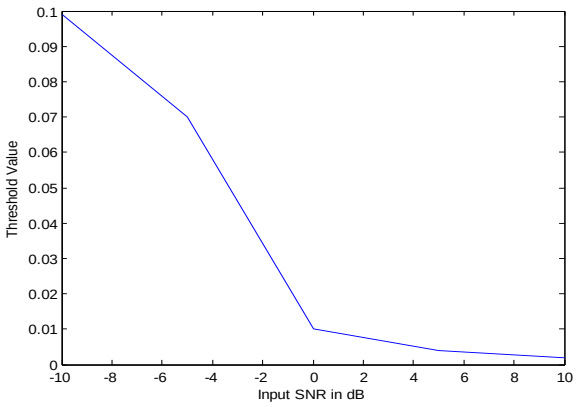

Fig.6 Graph of Input SNR versus Threshold Value

6. Performance Evaluation Parameter

This section presents the performance evaluation parameter used to measure the efficiency of the speech enhancement system. For quantitative evaluation of the denoising performance, the Signal to Noise Ratio (SNR) is a robust measure. The SNR is the difference between the denoised signal and the noisy signal, which is given by the following formula:

$$SNR(db) = 10log_{10}\left[\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1}(x(n)-\hat{x}(n))^2}\right] \quad (5)$$

Where x(n) and x̂(n) are the original and enhanced speech signals respectively, and N is the number of samples in the speech signal.

7. Experimental Results

This section presents the experimental results of the speech enhancement system at various SNR levels from -10 to 10 dB. The speech signal taken from the TIMIT Acoustic-Phonetic Continuous Speech Corpus [16], were used to evaluate the proposed algorithm. For this purpose, clean speech signal sampled at 16 kHz is used. The duration of the speech signal is 4 sec. Then white Gaussian noise is added to speech signal to obtain noisy speech signal. Additive white Gaussian noise is used because, in speech communication from a noisy acoustic environment, the signal is observed as an additive random noise. The reason for choosing white Gaussian noise is that it is one of the most difficult types of noise to remove, because it does not have a localized structure either in the time domain or in the frequency domain [17]. The chosen frame size is N = 269 with no overlapping and the signal is decomposed into four levels by using sym1 wavelet. The whole experiment is performed on Matlab Wavelet toolbox (The MathWorks Inc., 2013).

In order to evaluate the denoising performance of the proposed adaptive thresholding method, the results are compared with the four thresholding methods described in this paper (hard, soft, garrote and semi-soft). These tests are done under different SNR inputs (-10dB, -5dB, 0dB, 5dB and 10dB) of speech signal and the experimental results are shown in TABLE I.

Table. 1 SNR Output Of Four Thresholding Methods

METHOD↓	INPUT SNR (dB)				
	-10	-5	0	5	10
Hard Thresholding	0.8357	4.0649	7.6543	10.3534	12.7951
Soft Thresholding	1.1368	4.3987	7.7550	10.7289	12.7264
Garrotte Thresholding	1.1510	2.5173	6.7744	10.4586	14.1030
Semi-soft Thresholding	1.7425	3.5937	7.7375	10.7656	14.4009
Proposed Adaptive Thresholding	2.2582	4.5254	7.9353	10.8902	14.4913

From the Table I it can be seen that when SNR input is low, soft thresholding algorithm has better performance as compared to hard thresholding. In the garrotte thresholding, SNR improves only at 10dB input SNR level, then hard and soft thresholding. Semi-soft thresholding algorithm performs better than hard and soft thresholding algorithm to some extent, except SNR level of -5 dB. Performance of the proposed threshold algorithm is the best under different input conditions.

Table.2 Comparison Of Snr Of Proposed System And Speech Enhancement System In [17]

SNR input(dB)	SNR Output(dB) [17]	Proposed Adaptive Thresholding Method SNR Output(dB)
-10	-0.86	2.43
-5	2.38	4.63
0	5.13	7.93
5	8.52	11.01
10	11.51	14.49

The proposed adaptive speech enhancement method is also compared with a speech enhancement system that threshold the wavelet coefficients on the basis of speech signal features [17]. SNR is calculated under different SNR inputs and the results are shown in TABLE II. From the table it can be seen that the proposed adaptive thresholding method can greatly improve SNR with less distortion of the original speech signal.

The qualitative performance of the algorithm can be seen from the Fig. 7, Fig. 8, Fig. 9, Fig. 10 and Fig. 11. Part (a) of each figure represents the original speech signal. Part (b) of each figure shows noisy speech signals at -10, -5, 0, 5 and 10 dB input SNR levels. The noisy speech signal was enhanced by using the proposed speech adaptive thresholding method. Part (c) of each figure shows enhanced speech signals at -10, -5, 0, 5 and 10 dB input SNR levels.

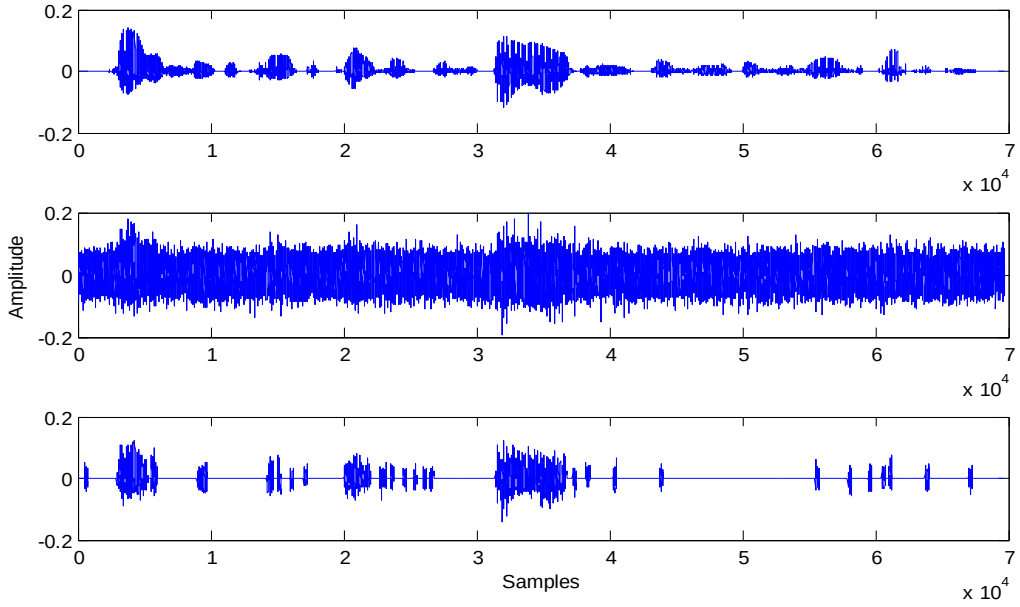

Fig. 7 Speech enhancement result: (a) Original Speech Signal, (b) Noisy Speech Signal at -10 dB input SNR level (c) Denoised Speech Signal at -10 dB input SNR level.

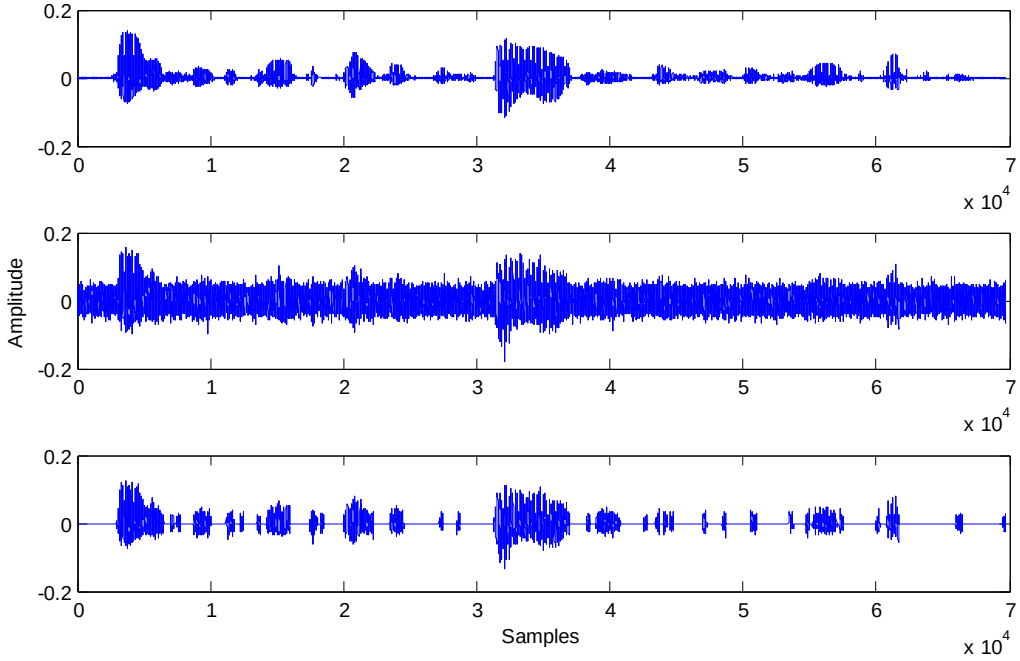

Fig. 8 Speech enhancement result: (a) Original Speech Signal, (b) Noisy Speech Signal at -5 dB input SNR level (c) Denoised Speech Signal at -5 dB input SNR level.

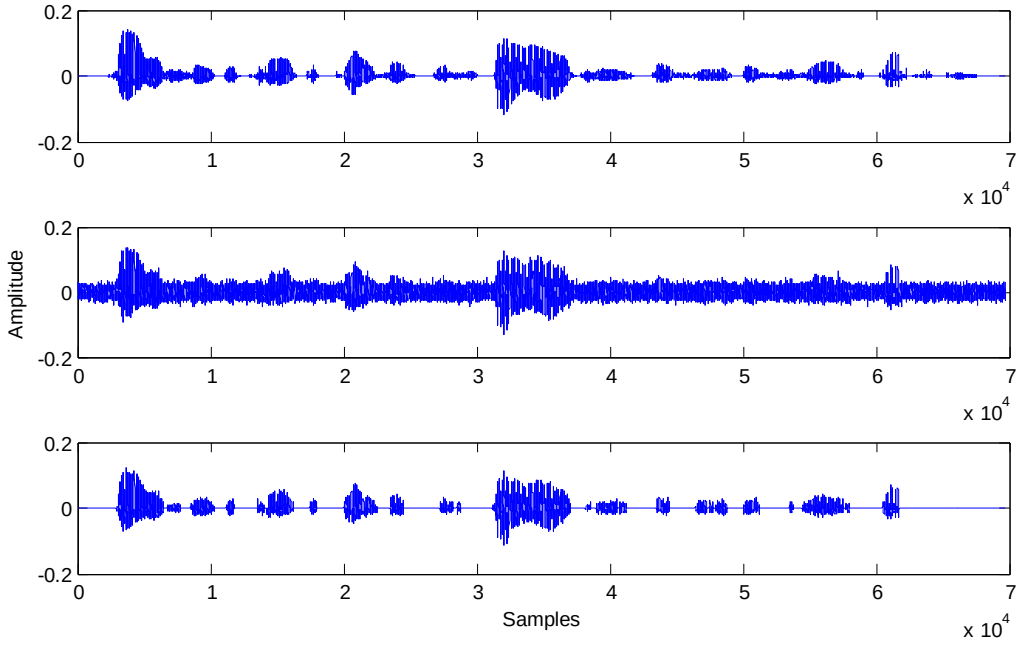

Fig. 9 Speech enhancement result: (a) Original Speech Signal, (b) Noisy Speech Signal at 0 dB input SNR level (c) Denoised Speech Signal at 0 dB input SNR level.

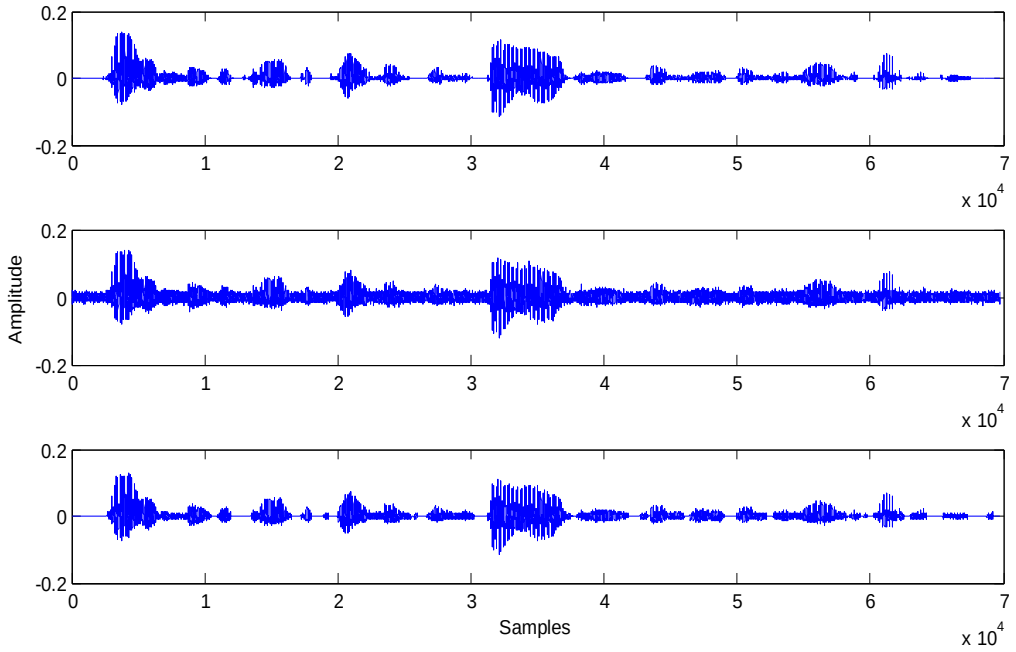

Fig. 10 Speech enhancement result: (a) Original Speech Signal, (b) Noisy Speech Signal at 5 dB input SNR level (c) Denoised Speech Signal at 5 dB input SNR level.

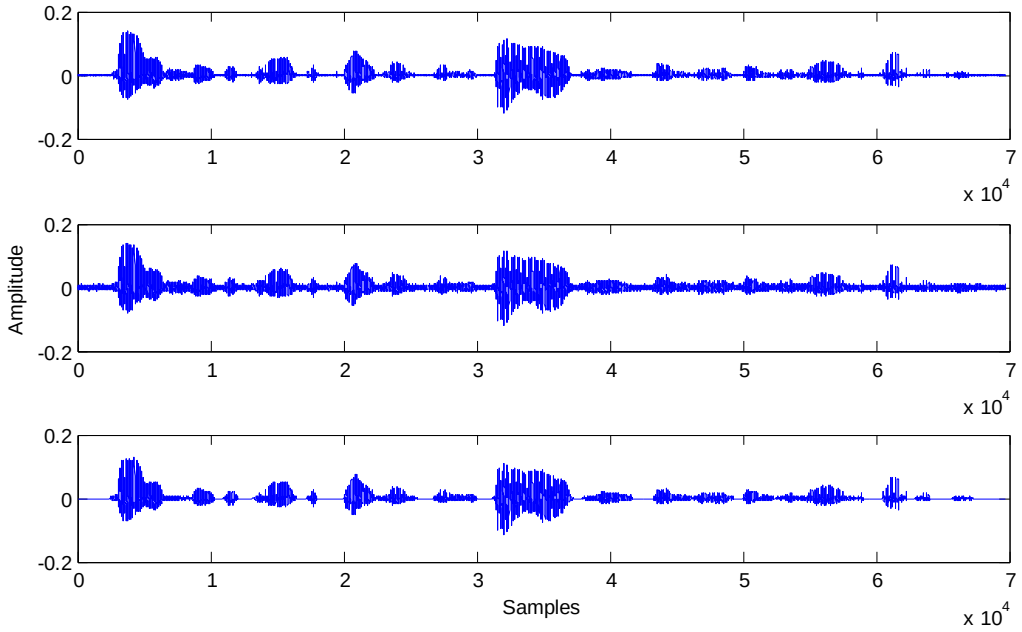

Fig. 11 Speech enhancement result: (a) Original Speech Signal, (b) Noisy Speech Signal at 10 dB input SNR level (c) Denoised Speech Signal at 10 dB input SNR level.

V. Conclusion

In this paper, a denoising approach based on the classification of speech into voiced, unvoiced and silence regions is implemented for speech enhancement using stationary wavelet transform. The speech enhancement system uses the energy threshold (energy threshold is manually calculated) of the wavelet coefficient for the separation of voiced, unvoiced and silence regions. Thresholding used for unvoiced speech is semi-soft thresholding, for voiced speech modified hard thresholding has been used and for silence region the coefficients are set to zero. For various noise levels the thresholding parameters that provide the best SNR were obtained manually. Then inverse stationary wavelet transform has been applied to recover the enhanced speech signal.

After performing experimental evaluations on speech signals from the TIMIT database and corrupting then by Gaussian noise at various input SNR levels, the result obtained shows that our method constitutes a successful application of the adaptive wavelet thresholding. The performance was evaluated in terms of the Signal to Noise Ratio (SNR). The proposed algorithm has achieved a much better performance in reducing noise with great intelligibility of the original speech.

Future work includes speech enhancement corrupted by various other disturbances like pink noise, babble noise, street noise, railway noise etc. Different methods for classification of speech signal into voiced, unvoiced and silence region may further lead to improvement in results.

References

Journal Papers:

[1] Wenhao Yuan and Bin Xia, "A Speech Enhancement Approach Based on Noise classification", Applied Acoustics, 96, September 2015, 11-19.
[2] Bingyin Xia and Changchun Bao, "Wiener Filtering Based Speech Enhancement with Weighted Denoising Auto-encoder and Noise Classification", Speech Communication, 60, May 2014, 13-29.
[3] Adam Borowicz and Alexandr Petrovsky, "Signal Subspace Approach for Psychoacoustically Motivated Speech Enhancement", Speech Communication, 53(2), February 2011, 210-219.
[4] S. F. Boll, "Suppression of Acoustic Noise in Speech using Spectral Subtraction," IEEE Transactions on Acoustics, Speech and Signal Processing, 27(2), April 1979.
[5] M. Berouti, R. Schwartz, and J. Makhoul, Enhancement of Speech Corrupted by Acoustic Noise, IEEE International Conference on ICASSP , Washington DC, 4, April 1979, 208-211.
[6] J. S. Lim and I. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech", Proceedings of the IEEE, 67(12), December 1979, 1586-1604.
[7] Y. Ephraim and H. L. V. Trees, "A Signal Subspace Approach for Speech Enhancement", IEEE Transactions on Speech and Audio Processing, 3(4), July 1995, 251-266.
[8] D.L. Donoho, "Denoising by Soft Thresholding", IEEE Transactions on Information Theory, 41(3), May 1995, 613-627.
[9] D.L. Donoho, and J. M. Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage", Biometrika, 81(3), August 1992, 425-455.

[10] Sheikhzadeh H and Abutalebi HR, "An Improved Wavelet Based Speech Enhancement System", In EUROSPEECH, 2001, 1855-1858.
[11] Seok, J. and Bae, K., "Speech Enhancement with Reduction of Noise Components in the Wavelet Domain", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2, 1997, 1323-1326.
[12] G. P. Nason and B. W. Silverman, "The Stationary Wavelet Transform and Some Statistical Application", Wavelets & Statistics: Lecture Notes in Statistics, 103, 1995, 281-299.
[13] M. Jansen, "Noise Reduction by Wavelet Thresholding", Lecture Notes in Statistics, 161, 2001.
[14] H. Y. GAO and A. G. Bruce, "WaveShrink with Firm Shrinkage", Statistica Sinica, 7, 1997, 855-874.
[15] Jong Kwan Lee and Chang D. Yoo, "Wavelet Speech Enhancement Based on Voiced/Unvoiced Decision", Korea Advanced Institute of Science and Technology, The 32nd International Congress and Exposition on Noise Control Engineering, Jeju International Convention Center, Seogwipo, Korea, August 25-28, 2003
[16] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, et al., TIMIT Acoustic–Phonetic Continuous Speech Corpus: Linguistic Data Consortium, 1993.
[17] Saeed Ayat, M.T. Manjuri-Shalmani and Roohollah Dianat, "An Improved Wavelet Based Speech Enhancement by using Speech Signal Features", Computer and Electrical Engineering, 23 May 2006, 411-425.

Books:

[18] R. Polikar, "The Wavelet Tutorial" Available: (http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html, 19960).
[19] D Lee Fugal, Conceptual Wavelet in Digital Signal Processing (Space and Signals Technical Publishing, San Diego, California, First Edition, 2009)
[20] S. V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction (Wiley, Third Edition, 2005).