GU J Sci, Part A, 5(4): 159-171 (2018)

Gazi University

Journal of Science

PART A: ENGINEERING AND INNOVATION

http://dergipark.gov.tr/gujsa

Speech Enhancement using Maximal Overlap Discrete Wavelet Transform

Selma OZAYDIN[1,*], Iman Khalil ALAK[2]

[1] *Çankaya University, Department of Electronics and Communication Engineering, Ankara, Turkey, email:selmaozaydin@cankaya.edu.tr*

[2] *Çankaya University, Department of Electronics and Communication Engineering, Ankara, Turkey, email:mony9918@yahoo.com*

Article Info

Received: 07/08/2018
Accepted: 17/12/2018

Keywords

Wavelet thresholding, Discrete wavelet transform, Maximal overlap, Speech enhancement, Signal denoising

Abstract

Signal denoising for non-stationary digital signals can be effectively succeeded by using discrete wavelet transform. Selecting of a suitable thresholding method is important to minimize the loss of useful signal information. This paper demonstrates the application of the maximal overlap wavelet transform (Modwt) technique in speech signal denoising. The analysis algorithm was performed on Matlab platform. In this algorithm, different kinds of input noisy speech signals having environmental background noises such as restaurant, car, street or station were tested. The noisy signals were filtered from the speech signal by thresholding of wavelet coefficients with threshold estimation methods known as sgtwolog, modwtsqtwolog, heursure, rigrsure and minimax. The performance of the Modwt in denoising process was evaluated by comparing signal-to noise ratio (SNR) and mean square error (MSE) results to those of well-known threshold estimation methods. First, denoising effectiveness of a Modwt based threshold method was tested in different scenarios and very important improvements in denoising process were achieved by Modwt based scenarios. Next, the influence of the different wavelets families on Modwt based threshold estimation method was evaluated by experimental results. The results revealed that Modwt based method outperforms conventional threshold methods while providing nearly up to a %24 increase in SNR value.

1. INTRODUCTION

Recovery of an original speech signal from background noise is a challenging task in a speech denoising process. The goal of a speech denoising algorithm is to recover original speech signal by removing noise with a minimum distortion. Wavelet analysis is a mathematical model that transforms the original signal (especially in time domain) into time-frequency domain for analysis and processing. It is very appropriate method for non-stationary data analysis. The one-dimensional wavelet transform technique is used for time/frequency analysis of speech signals by achieving very good noise reduction performance. In wavelet transform, low-frequency information is analyzed in a longer area and high-frequency information is in a shorter one [1]. The idea of wavelets for the analysis of non-stationary signals by using translation and dilation of a single function was first introduced in 1982 by Jean Morlet [2,3]. In a wavelet transform, a function is analyzed as a sum of time-shifted and scaled representations of a mother wavelet. Signal de-noising and perfect reconstruction can be realized easily for orthonormal wavelets. Yves Meyer has proposed a very popular orthonormal wavelet method in 1992 [5].

It is possible to classify Wavelet transform into two groups as discrete wavelet transform (DWT) for discrete-time signals and continuous wavelet transform (CWT) for continuous-time signals [3,6]. DWT method transforms the digital input signal from time domain to time scale domain for analysis and the input signal is decomposed into several other signals with different levels of resolution to allow recovering the original input signal without losing any information. DWT is performed by first sampling of the signal and then filtering with low-pass and high-pass finite impulse response (FIR) filters. Among the typical reasons of using wavelet transforms is that the Fourier transform analysis is not in a position to

*Corresponding author, e-mail: selmaozaydin@cankaya.edu.tr

concern to the local information of signals. Therefore, the Fourier transform is not used for analyzing signals in a joint time-frequency domain of localized signals. Unlike Short Time Fourier transform, wavelet transform supplies variable window sizes which are used to analyze different frequency components with different resolutions and to find discontinuities or bursts in an audio signal. The wavelets are compressed at low scale to catch fast-changing details in a high frequency and they are stretched to follow slow changing properties in a low frequency. In wavelet transform, the usage possibility of variable size windows results in a high frequency-resolution in low frequency bands and low frequency-resolution in high frequency bands. Consequently, non-stationary signals can be modelled to analyze slow temporal variations in low frequencies and rapid temporal changes in high frequencies. Since Wavelets are characterized by scale and position, they are very useful in analyzing variations in signals in terms of scale and position [3,4,7].

A noise in a speech can be in white noise, pink noise, realistic noise or any other kind of noises. There are special advantages of the wavelet transform when denoising a speech signal. For example, its time frequency localization and multi resolution analysis are among its advantages. The objective quality of denoised signal is usually measured by SNR value. In wavelet analysis low frequency coefficients particularly stands for signal and high frequency coefficients represent noise. It is possible to get denoising by selecting a threshold and removing the noise from these high frequency coefficients according to the selected threshold. Denoising is done by soft or hard threshold of wavelet coefficients [8]. The Modwt technique [9,10] has many advantages over DWT [7] due to its specific properties as mentioned in Section 4. It is a time shift-invariant, redundant and non-orthogonal transform technique. In this paper, the effects of these properties of the Modwt on denoising performance were tested in different scenarios and compared with the well-known DWT methods. A remarkable performance improvement was obtained for each test scenarios.

2. RELATED WORKS AND MOTIVATION

Mihov, S. G., Ivanov, R. M. & Popov, A. N. [22] investigated the performance of wavelet transform for denoising speech signals contaminated with various noises in several SNR levels. The experiment was done for a large database of reference noisy speech signals. They applied the wavelet denoising method by using soft and hard thresholding. They calculated the output SNR for four different methods of threshold estimation (heursure, rigsure, minimax, sqtwolog). It is seen from the results that minimax and sqtwog criterion did not improve the perception and SNR quality of the signals. An improvement in signal quality was achieved by using heursure or rigsure methods and by selecting the soft thresholding. They increased the levels of decomposition to increase the performance but it was found that there was no sensible improvement in signal quality, besides computational complexity was increased. It can be concluded from the test results that the output SNR values corresponding to several common wavelet functions (such as haar, db3, db5).are not so different from each other. Besides, increasing of decomposition level has no sensible effect on signal quality. Furthermore, it is seen that there is no particular wavelet denoising procedure that would be suitable for all kinds of input noises.

Sumithra, A., Thanushkodi, B. [23] proposed a speech improvement called as trimmed thresholding. The performance of different thresholding methods are evaluated against various noises. It was concluded that the soft thresholding was best in denoising but worst in preserving edges, and hard thresholding is best in preserving edges but worst in denoising. The objective and subjective tests demonstrated that the proposed scheme with trimmed thresholding was better in denoising of experimented environmental noises when compared to hard and soft thresholding strategies. It was also indicated that the proposed technique gave better MSE performance than other wavelet thresholding methods. The limitation of this proposed scheme is the proper tuning of a parameter (α) for each noise conditions.

Sanam, T. F. & Shahnaz, C. [24] proposed a thresholding-method based on the Teager energy-operation. Different noisy input data were evaluated and objective and subjective measures were performed. The objective and subjective evaluations represented that the proposed method was capable of enhancing the speech with better quality and intelligibility compared to some existing methods. However, the histogram model of the Teager energy-operated wavelet packet coefficients was based on Gaussian distribution. For

a future work, the usage of other types of distributions was proposed to approximate the histograms of the WP coefficients. Besides, the use of a perceptually weighted filter would be an option to mask the residual noise.

Du, L, Xu, R, et al., [25] investigated various wavelet threshold functions used in wavelet packet denoising. The experimental results showed that some key parameters should be considered including wavelet basis, layers of wavelet packet decomposition, values of threshold, and threshold functions to obtain an ideal effect for speech denoising by the nonlinear filtering based on threshold in wavelet domain. But in this study, the only noise type used in the tests were the additive White Gaussian noise. Any types of noise was recommended as a future work.

From the studies on DWT, it can be concluded that wavelet transform based methods provide a powerful tool for non-linear filtering of speech signals contaminated by noise. Here, selecting of a suitable thresholding is a key issue. Therefore, the motivation in this study was to design an effective wavelet based speech denoising algorithm. The goal was to perform a speech enhancement method to remove different kinds of daily life environmental background noises from an input signal such as restaurant, car, street or station noises. Thus, the performance of the Modwt based thresholding method in denoising process was evaluated in this paper due to the its superior properties over DWT as explained in Section 3.3 and very important improvements were achieved as can be found in Section 4.

3. THRESHOLDING IN DISCRETE WAVELET TRANSFORM

There are many useful properties in DWT for the time arrangement and advanced information examination. The objective is generally to lessen the noisy motion. In DWT, the time arrangement signal is first applied on wavelets to separate the signal into various constituents. At that point, DWT isolates the constituents into an alternate recurrence at different scales [7]. The DWT gives adequate data both to examination and union of the original signal, with a noteworthy decrease in the calculation time. The DWT picks a subset of scales and positions to ascertain. DWT is an intense instrument to use in an extensive variety of uses [11,12]. Wavelet performs multi resolution analysis of a signal with localization in both time and frequency [8]. The DWT is represented by equation as shown below.

$$DWT_{\psi}f(m,n) = \int_{-\infty}^{\infty} f(t)\psi_{m,n}^{*}(t)\,dt \qquad (1)$$

where,

$$\psi_{m,n}(t) = 2^{-m}\psi(2^{m}\,t - n) \qquad (2)$$

DWT utilizes multi determination channel banks and wavelet channels to break down and reproduce the original speech signal. It gives adequate data and diminishes calculation time for analysis and synthesis. The DWT first uses a down sampling and afterward decays the signal using an arrangement of low and high pass filters. The yields are estimation and detail coefficients. DWT is basically a sampled version of CWT. In order to remove a noise with wavelet transform, the following steps can be followed. Firstly, wavelet transform of noisy signal is performed. Then, thresholding is done with the resulting wavelet coefficients. Finally, inverse transform of the signal is performed to obtain the denoised signal [13,14].

3.1. Denoising by Thresholding

Thresholding is used in wavelet based denoising. In thresholding, it is assumed that high amplitude wavelet coefficients represent signal, and low amplitude coefficients represent noise. Thresholding process is applied by first computing the wavelet coefficients of input noisy speech, then filtering the coefficients to eliminate noise while preserving most of the original signal. After thresholding, inverse transform of denoised speech is retrieved [13].
Assume that a signal (s) has an additive Gaussian noise (n) having variance σ^2 and y be a finite length observation sequence of this noisy signal, so we can write,

$$y = s + n \qquad (3)$$

The goal is to recover the signal *s* from the noisy observation *y*. If a DWT matrix (W) applied to equation (1) in wavelet domain, then we can write Y = S + N, where $Y = W_y$, $S = W_s$, $N = W_n$. Then, the denoised signal $\hat{Y}$ can be estimated by using,

$$\hat{Y} = T(S, \lambda) \qquad (4)$$

Where T (.) denotes a thresholding function and λ denotes a threshold value. Hard and soft threshold rules are used in wavelet based signal noise reduction. In hard threshold, all wavelet coefficients below a certain threshold are set to be zero and other wavelet coefficients are not modified. The hard threshold method removes the noise by thresholding the detailed wavelet coefficients, while keeping the low-resolution (estimation) coefficients unaltered. [13,15]. It is expressed mathematically as,

$$T_h(X,\lambda) = \begin{cases} X; & for\ |X| > \lambda \\ 0; & for\ |X| \leq \lambda \end{cases} \qquad (5)$$

where X is the noisy wavelet coefficient and λ is the threshold proposed by Donoho and Johnstone [16]. Soft threshold sets all wavelets detail coefficients to zero whose total values are below a certain threshold and shrinks those above it. The thresholding result is equal to the value of a sign function which multiplies the subtraction value between a coefficient and threshold T. The sign function returns [1,0,-1] if the coefficient is greater than zero, equal 0, or less than zero, respectively,

$$T_s(X,\lambda) = \begin{cases} sign\{X\}.(|X| - \lambda); & for\ |X| > \lambda \\ 0; & for\ |X| \leq \lambda \end{cases} \qquad (6)$$

3.2. Threshold based Denoising Methods

Noise threshold estimation methods provide a representative threshold value (T) by considering the noise power (σ) to separate the undesired coefficients from the significant ones. Therefore, a low threshold value may not reduce the noise sufficiently while a higher value may destroy the details. The well-known threshold estimation methods named sqtwolog, rigrsure, heursure, and minimax are considered for this study.

This universal threshold method (sqtwolog) was proposed by Donoho and Johnstone [16,17] to remove noise. This method uses square root of logarithm to assess the threshold value (T_{sq}) where σ signifies the mean absolute deviation (MAD) and N is the length of the noisy signal. By assuming the noise as zero-mean Gaussian, T_{sq} is calculated as,

$$Tsq = \sigma\sqrt{2\log(N)} \qquad (7)$$

where $\sigma = (\text{Median}(|w|)/0.6745)$, in which w is the set of the noise wavelet coefficients. For colored-noisy signals, a scale-dependent threshold calculation is required to account for the different variances of the noise wavelet coefficients in each scale. In this case,

$$\text{Tsq}_\text{i} = \sigma_\text{i}\sqrt{2\log(\text{N}_\text{i})} \qquad (8)$$

Where noise standard deviation to scale i, $\sigma_\text{i} = \left(\frac{\text{Median}(|\text{w}_\text{i}|)}{0.6745}\right) = \frac{\text{MAD}_\text{i}}{0.6745}$, wi is the set of noise wavelet coefficients to scale i.

The Rigrsure, which is known as Steins unbiased risk estimator (SURE), is an adaptive threshold selection method which was investigated by Donoho and Jonstone [16,17] and it depends on Stein's impartial probability estimation guideline. This technique computes probability estimation first by using the given threshold (T), and after that minimizing the non-likelihood. Then, the threshold is acquired. The

diverse thresholds are picked in various scales, and the wavelet coefficients of comparing scales are downsized. w_b is the b^{th} coefficient wavelet square chosen from the vector w=[w_1 w_2 …w_N] and σ is the standard deviation of the noisy signal.

$$Tr_i = \sigma_i\sqrt{\mathrm{w_b}} \qquad (9)$$

The Heursure method is a combination of sqtwolog and rigrsure methods. The chosen threshold is the best predicable variable threshold. If the SNR is very small, the Sqtwolog method (Tsq) is used to choose the threshold; if SNR is large, Rigrsure form (Tri) will be utilized. Suppose s is the sum of the squares of the N wavelet coefficients, then $\mathrm{a} = (\mathrm{s} - \mathrm{N})/\mathrm{N}$, $\mathrm{b} = (\log_2\mathrm{N})^{3/2}\sqrt{\mathrm{N}}$. Threshold equation (Th) can be represented mathematically as [7]:

$$\mathrm{T_h} = \begin{cases} \mathrm{T_{sq}} & \mathrm{a} > \mathrm{b} \\ \min(\mathrm{T_{sq,}}\ \mathrm{T_r}) & \mathrm{a} \geq \mathrm{b} \end{cases} \qquad (10)$$

The Minimax threshold selection method uses a fixed threshold. The minimax method finds thresholds over a given arrangement of functions of the greatest Mean Square Error (MSE) by using Minimax rule which is used in statistics to design estimators. In this strategy, the threshold value will be chosen by acquiring a minimum error between wavelet coefficient of noise signal and original signal. The method finds optimal thresholds [8]. The threshold value (Tm) is given,

$$\mathrm{T_m} = \begin{cases} \sigma_\mathrm{i}(0.3936 + 0.10829\log_2\mathrm{N}) & \text{for } \mathrm{N} > 32 \\ 0 & \text{for } \mathrm{N} < 32 \end{cases} \qquad (11)$$

where $\sigma_\mathrm{i} = (1/0.6745).\mathrm{median}(|\mathrm{w_i}|)$ wi represents the wavelet coefficient vector at scale i and N represents the length of the signal [7].

3.3. The Maximal Overlap Discrete Wavelet Transform Technique

The Modwt is a linear filtering operation that transforms into coefficients related to variations over a set of scales. It is used to inspect the scale-dependent signal behaviors. The Modwt is an undecimated modification to DWT and therefore it is time shift-invariant method, i.e., a translation in the signal will result in a translation of wavelet coefficients by the same amount. In DWT (or Modwt), a digital signal S(n) is decomposed into its detailed (cD1(n)) and approximation (smoothed) (cA1(n)) coefficients using high-pass filter (HiF-D) and low-pass filter (LoF-D), respectively. Consequently, the cD1(n) contains higher frequency components, while the cA1(n) has low-pass frequencies. Inversely, that is possible to perform the original signal from the approximations and details coefficients. The Modwt differs from the DWT in that it is a highly redundant, non-orthogonal transform. The Modwt offers several advantages over the DWT. The redundancy of the Modwt facilitates alignment of the decomposed wavelet and scaling coefficients at each level with the original time series, thus enabling a ready comparison between the series and its decomposition. Besides, the redundancy of the Modwt wavelet coefficients increases the effective degrees of freedom on each scale and thus decreases the variance of certain wavelet-based statistical estimates. The Modwt aligns wavelet coefficients at each time point with the original data index so one can simultaneously analyze localized signal variation with respect to scale and time, and the temporal relation to events. The Modwt retains down-sampled values at each level of the decomposition that would be otherwise discarded by the DWT. The Modwt is well-defined for all sample sizes N, whereas for a complete decomposition of *J* levels the DWT requires *N* to be a multiple of 2^J [9,10].

4. EXPERIMENTAL RESULTS

The performance of the thresholding methods referenced in this paper was evaluated using signal-to-noise ratio (SNR) and mean squared error (MSE) values. SNR is the most widely used objective measure method to evaluate speech quality. The goal is to minimize reconstructed error variance and maximize SNR. The 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. It is usually calculated in terms of decibel (dB) as in equation (12),

$$SNR = 10\log_{10}\frac{\sum_n s^2(n)}{\sum_n [s(n)-\hat{s}(n)]^2} \qquad (12)$$

MSE is a distance measure between the processed speech ($\hat{s}(n)$) and the clean speech (s(n)) and It is computed as in equation (13),

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\left(s(n)-\hat{s}(n)\right)^2 \qquad (13)$$

4.1. Test Setup

During the experimental work, NOIZEUS speech corpus [19,20] was used for the noisy input speech dataset. It contains 30 phonetically balanced speech sentences pronounced by three male and three female speakers. The NOIZEUS speech files were corrupted by eight different artificially added background noises including car, train, restaurant, babble (crowd of people), airport, street, train station and exhibition at different levels of SNRs. The corpus was sampled at 8 kHz. In our evaluation, we have tested the performance of the designed algorithms on such noisy speech samples degraded for all these noise types and Awgn (additive white Gaussian noise) degraded input speech dataset at initial noise SNR levels of 0dB, 5dB, 10dB and 15dB. During the tests, the input noisy speech signal was decomposed by applying different threshold selection methods to the wavelet coefficients. These were sgtwolog, modwtsqtwolog, heursure, rigrsure, and minimax thresholding methods, with hard or soft threshold techniques. Because the noises were artificially added to the clean speech signal in NOIZEUS corpus, the clean speech sample (Figure 1) of a noisy speech was used as reference for the evaluation of the performance of denoising procedure during all of the applied scenarios.

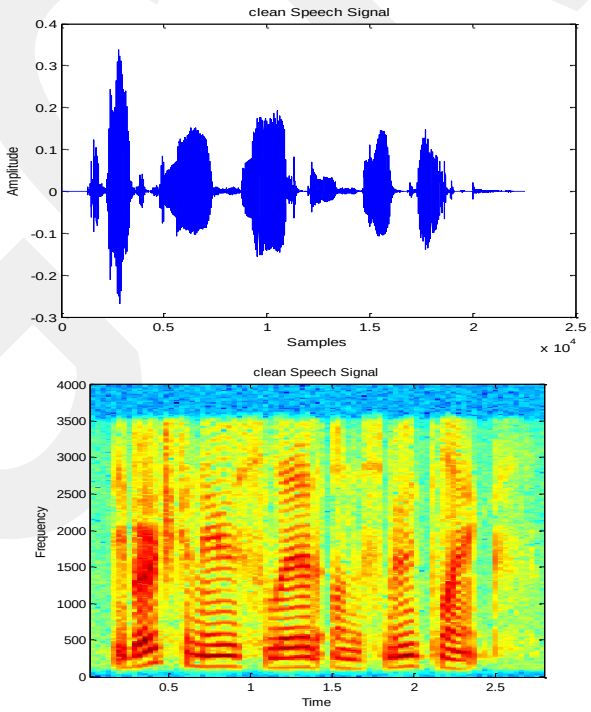

Figure 1*. a) Clean speech signal, b) Spectrogram of Clean speech signal*

The software algorithm for denoising of speech through discrete wavelet transform was implemented and tested in MATLAB (Vers. R2016) platform. We used the 'wden' function which returns denoised version of an input signal by thresholding the wavelet coefficients [18]. Then, the performance of the threshold

selection methods was examined in different types of scenarios. We carried out two different types of experiments to evaluate the effectiveness of threshold selection methods in denoising of speech from different kinds of environmental background noises. First, performances of different threshold methods in denoising process were tested as described in Section 4.2. After obtaining outstanding SNR and MSE results for the Modwt filtering method in the first experiment results, we evaluated the performance of the Modwt method together with other wavelet families as told in Section 4.3.

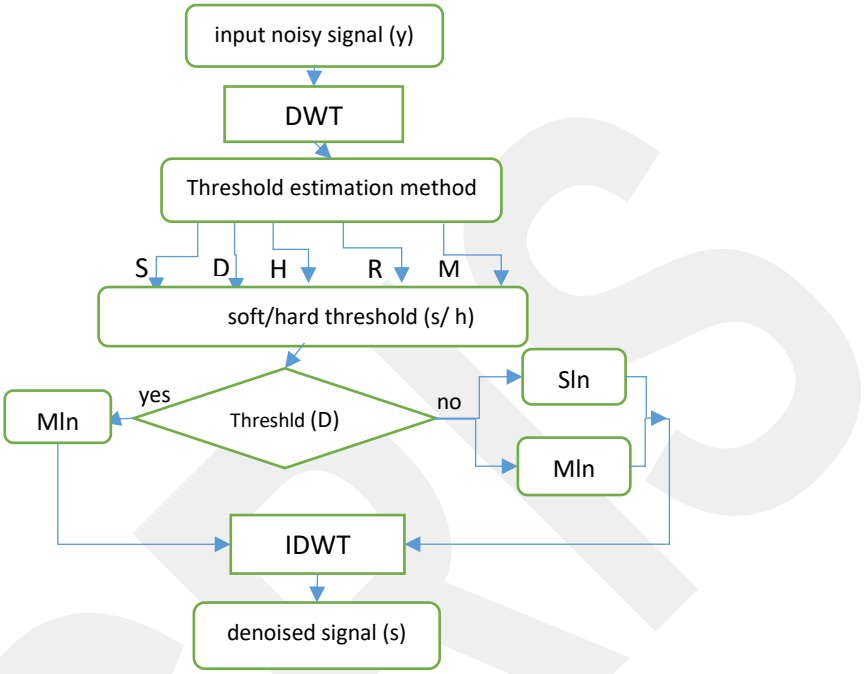

***Figure 2**. Flowchart of denoising senarios during the tests*

The meanings of three-letter abbreviations corresponding to the "test scenarios" during the tests can be explained as follows; The first letter represents the name of the threshold selection method including [Sqtwolog (S), Modwtsqtwolog (D), Heirsure (H), Rigrsure (R) and Minimax (M)], the second letter represents soft (s) or hard (h) thresholding, and the last letter represents sln (S) or mln (M) scaling (i.e. RsM means the senario with Rigrsure-soft thresohld-Mln scale). Threshold's rescaling method 'sln' is a model with unscaled noise and performs threshold rescaling using a single estimation of level noise based on the first-level coefficients, 'mln' is a model with nonwhite noise and performs the threshold rescaling using a level-dependent estimation of the level noise. The flowchart of the scenarios is seen in Figure 2. In Table 1 and 2, bold emphasized values show the best two values in a column.

4.2. Test Results for Denoising of Speech with Discrete Wavelet Transform

The input signals presented in Table-1 includes the noisy speeches at SNR levels of 5dB. During these tests, Daubechie family [21] wavelet filter (db5) at fifth decomposition level was selected. (increasing of the decomposition level didn't affect the results very much as also implied in [22]). To evaluate the system performance, objective tests such as SNR and MSE values were employed. Hard and soft threshold methods were used for thresholding the wavelet coefficients. The input noisy speech signal was decomposed by applying different threshold selection methods to the wavelet coefficients named sgtwolog, modwtsqtwolog, heursure, rigrsure, and minimax. We tested the effectiveness of Modwt technique in terms of other well-known techniques. Because 'wden' function only supports the sqtwolog universal thresholding method for Modwt, we used 'Modwtsqtwolog' option in 'wden' function. For each input noisy signal, wavelet denoising procedure was applied under different scenarios in Figure 2 and SNR and MSE measurements were obtained as can be seen in Table 1 and Table 2, respectively.

Figure 3 shows the graph of values in Table 1. When we applied the Modwt method together with sqtwolog thresholding, we obtained higher SNR than other methods. As can be seen from the results in Table 1 and Figure 3, the highest point of SNR was obtained in the scenario based on the combination of the Modwt method with sqtwolog thresholding (D), hard threshold selection (h) and mln scaling (M), (the scenario: DhM), for the input noisy signal option Awgn. From the evaluation of the results in Table 1 and Table 2, it is concluded that the application of the Modwt technique with sqtwolog threshold method (the scenario DhM) for the noisy input signals as low as SNR level 5 dB gives the best SNR (and least MSE) result for all different kinds of noisy input speech signals when compared to other scenarios. In Figure 4, denoised form of a car noisy input signal at SNR level of 5dB with the scenario (SsS) can be seen. Here, the input signal was decomposed just using sqtwolog (without Modwt). In this test condition, output SNR level of 6.60dB was obtained. Figure 5 shows the denoised signal output of a car noisy input signal at SNR level of 5dB and its denoised form with the scenario (DhM) can be seen. The input signal was decomposed using modwtsqtwolog. In this test condition, output SNR level of 8.45dB was obtained. To summarize, from the results in Table 1, we can say that there is a performance increase up to about %24 with the Modwt technique (for the input train noisy speech) when we compared with closest SNR value.

Table 1. *Results of SNR values for 5dB noisy input speech signal*

Thresold type	Scenario Type	Input background Noises (SNR level at 5dB)								
		car	airport	babble	exhibition	restaurant	station	Street	Train	whitenoise
Sqtwolog (S)	SsS	6.599453	**6.599284**	6.038119	6.054357	6.569429	6.395572	5.996394	4.625853	3.832965
	SsM	0.827104	0.814066	0.979351	1.815115	1.07446	0.807469	1.497182	1.019166	2.006832
	ShS	5.174613	4.659435	4.654143	5.620194	4.77907	6.06793	5.357301	6.452877	6.550975
	ShM	2.262341	2.121267	2.053034	3.341008	2.517587	2.232352	3.112913	2.650479	4.046997
Modwt sqtwolog (D)	DsM	4.566519	4.609105	4.673867	5.87466	5.061075	4.603288	5.272652	5.030505	6.063724
	DhM	**8.446362**	**8.13396**	**6.71972**	**7.431758**	**7.513789**	**8.57759**	**7.31659**	**8.908714**	**10.195995**
Heirsure (H)	HsS	4.889845	4.568035	4.599193	5.891204	4.693976	5.06769	5.163509	7.203949	9.41544
	HsM	5.164751	5.079769	**6.259527**	6.752426	**6.730659**	6.897959	**6.758369**	7.115178	8.497333
	HhS	4.430483	4.357516	4.364076	4.465338	4.373811	4.620351	4.400491	5.862836	7.788144
	HhM	5.564955	5.60122	4.93193	4.840682	4.844311	5.670954	4.846422	6.470235	8.654882
Rigrsure (R)	RsS	4.889845	4.568035	4.599193	5.891204	4.693976	5.219849	5.163509	**7.161943**	**9.580408**
	RsM	6.735506	6.437465	6.259527	6.752426	6.730659	7.132055	6.758369	7.074018	8.480397
	RhS	4.430483	4.357516	4.364076	4.465338	4.373811	4.562203	4.400491	4.893766	7.089057
	RhM	5.597294	5.428461	4.93193	4.840682	4.844311	5.597028	4.846422	5.372702	8.086111
Minimax (M)	MsS	**6.937342**	6.428099	6.034069	**7.047137**	6.556032	**7.408714**	6.609635	6.237572	5.599136
	MsM	2.125373	2.101124	2.245213	3.275954	2.483061	2.109691	2.962915	2.446835	3.592945
	MhS	4.817294	4.506366	4.54746	5.209974	4.611494	5.400822	5.030778	6.589164	8.118678
	MhM	4.212545	4.160218	3.789047	4.504452	4.31733	4.393178	4.598494	4.811363	6.304121

Table 2. *Results of MSE values for 5dB noisy input speech signal*

sp01_5db noise		Input background Noises (SNR level at 5dB)								
		car	airport	babble	exhibition	restaurant	station	street	train	whitenoise
sqtwolog	SsS	0.000256	0.000256	0.000291	0.00029	0.000258	0.000268	0.000294	0.000403	0.000484
	SsM	0.000966	0.000969	0.000933	0.00077	0.000913	0.000971	0.000828	0.000924	0.000736
	ShS	0.000355	0.0004	0.0004	0.00032	0.000389	0.000289	0.00034	0.000265	0.000259
	ShM	0.000694	0.000717	0.000729	0.000542	0.000655	0.000699	0.000571	0.000635	0.00046
Modwt sqtwolog	DsM	0.000408	0.000404	0.000398	0.000302	0.000364	0.000162	0.000347	0.000367	0.000289
	DhM	0.000167	0.00018	0.000249	0.000211	0.000207	0.000405	0.000217	0.00015	0.000112
heirsure	HsS	0.000379	0.000408	0.000405	0.000301	0.000397	0.000364	0.000356	0.000223	0.000134
	HsM	0.000356	0.000363	0.000277	0.000247	0.000248	0.000239	0.000247	0.000227	0.000165
	HhS	0.000421	0.000429	0.000428	0.000418	0.000427	0.000403	0.000424	0.000303	0.000195
	HhM	0.000325	0.000322	0.000375	0.000383	0.000383	0.000317	0.000383	0.000263	0.000159
rigrsure	RsS	0.000379	0.000408	0.000405	0.000301	0.000397	0.000351	0.000356	0.000225	0.000129
	RsM	0.000248	0.000265	0.000277	0.000247	0.000248	0.000226	0.000247	0.000229	0.000166
	RhS	0.000421	0.000429	0.000428	0.000418	0.000427	0.000409	0.000424	0.000379	0.000228
	RhM	0.000322	0.000335	0.000375	0.000383	0.000383	0.000322	0.000383	0.000339	0.000182
minimax	MsS	0.000237	0.000266	0.000291	0.000231	0.000258	0.000212	0.000255	0.000278	0.000322
	MsM	0.000717	0.000721	0.000697	0.00055	0.00066	0.000719	0.000591	0.000665	0.000511
	MhS	0.000386	0.000414	0.00041	0.000352	0.000404	0.000337	0.000367	0.000256	0.00018
	MhM	0.000443	0.000448	0.000489	0.000414	0.000433	0.000425	0.000405	0.000386	0.000274

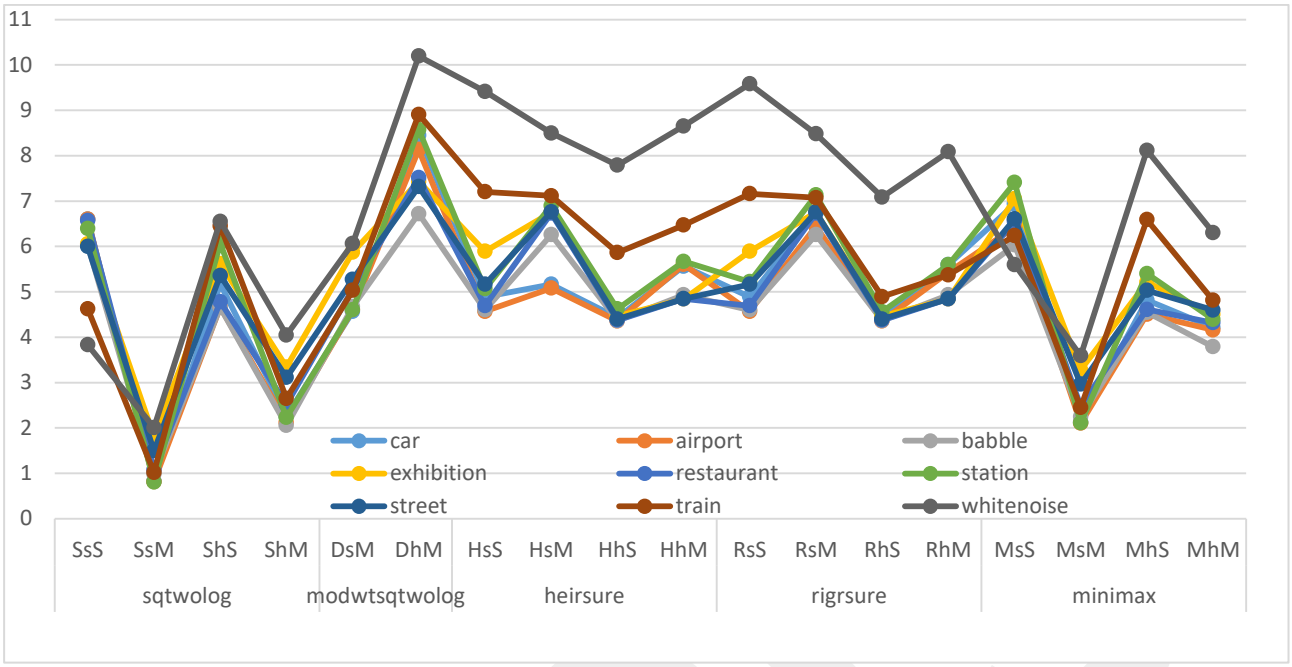

Figure 3. *Comparison of wave SNR*

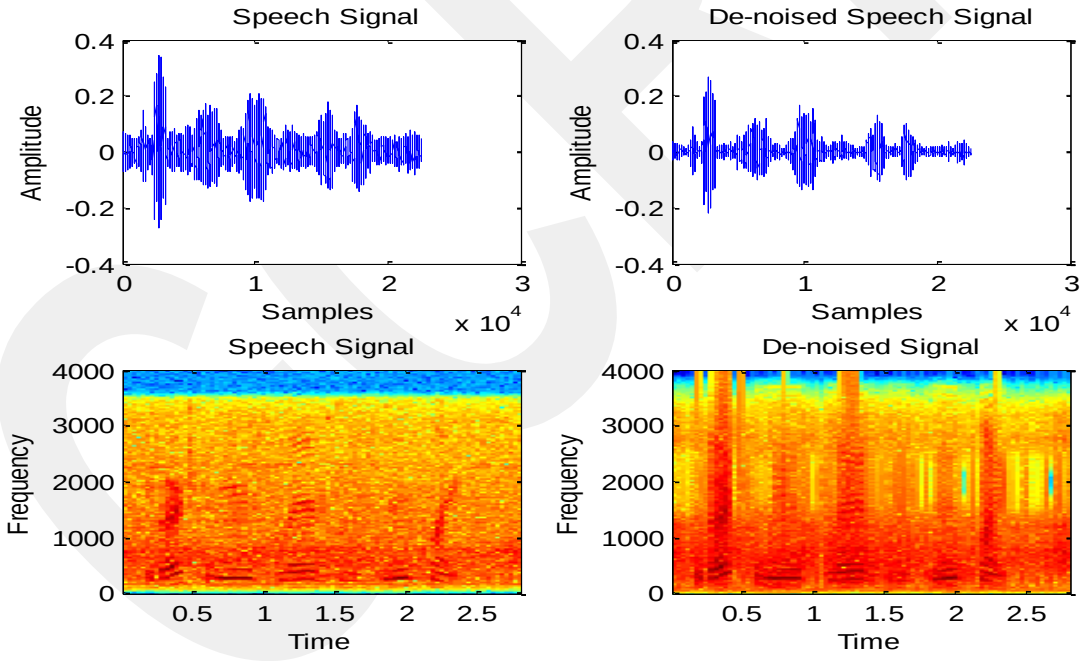

Figure 4. *5dB car noise speech signal and its denoised form during the scenario SsS*

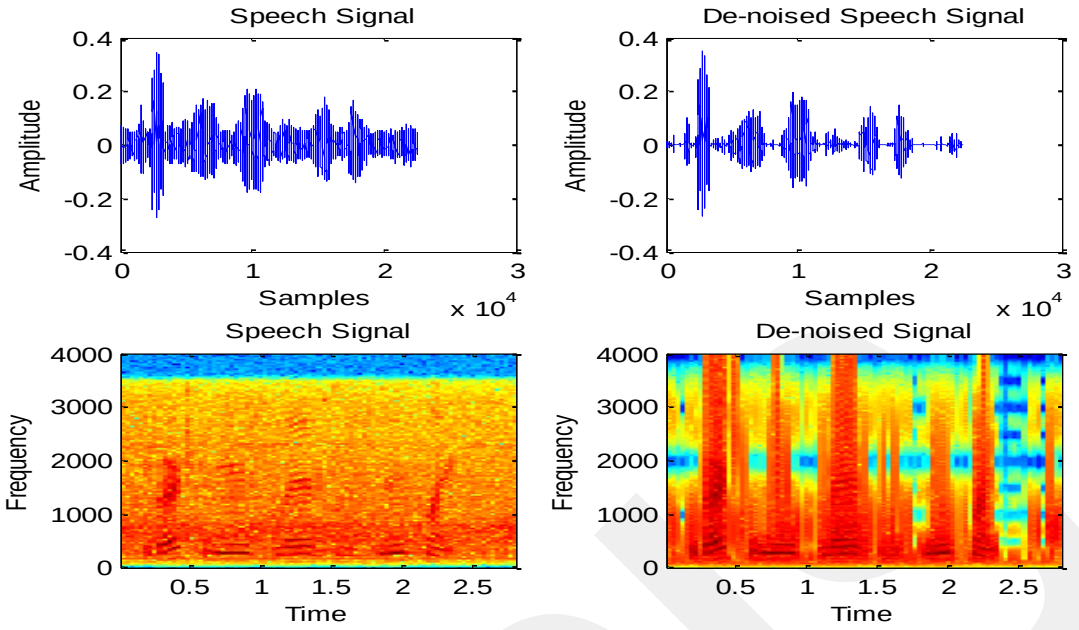

Figure 5*. 5dB car noise speech signal and its denoised form during the scenario DhM.*

4.3. Test Results for Denoising of Speech with Maximal Overlap Discrete Wavelet Transform

The performance of the Modwt method together with other wavelet families were evaluated. During the tests, we applied scenario (DhM) to different kinds of wavelet families (haar, db, sym, coif) by selecting the Modwtsqtwolog as threshold method in Matlab. The implemented speech denoising algorithm was evaluated for the input noisy signals at four different SNR levels (0dB, 5dB, 10dB, 15dB) from NOIZEUS database. The results of tests and the used wavelet functions during these tests are shown in Table 3-6. For all types of noises, the Modwt results had again superior performance to other results. The outputs of the scenarios using the Modwt have an outstanding performance over the other methods for different kinds of noisy input signals. Table-3 shows the SNR results of 0dB, 5dB and 10dB input noisy speeches for different kinds of wavelets. It can be seen from Table-3 that sym5 wavelet gave higher SNR results for airport, car, babble and train noisy signal inputs at 0dB input SNR level. Haar wavelet gave higher SNR for restaurant noisy input speech. The results for SNR level 5dB can also be seen in Table 3. For 5dB input SNR level, even if some results were very close each other, we obtained better output SNR at sym5 wavelet for airport, babble and restaurant noisy speech signals. We got higher SNR for car noise input at coif3 wavelet and for train noise at sym15 wavelet. For the noisy speech signals at SNR level of 10dB, the wavelet family coif4 gives better SNR result for airport noisy speech signal. For noise types babble and restaurant, db5 gives better result than other wavelet families. For the car and train noisy speech types, higher SNR value is obtained for sym15 wavelet. By evaluating the results of Table-3, even if we got higher SNR by changing the type of wavelet family, we can say that the effect of wavelet family type on performance is about maximum %2. Therefore, it does not seem critical on SNR performance.

***Table 3**. Wavelet family comparison at 0 dB, 5 dB and 10dB input SNR levels*

Wavelet type	Input noise level	Input noise type				
		airport	car	babble	train	restaurant
haar	0 dB	3.921514	4.953874	3.89721	4.663745	**2.926489**
	5 dB	7.378904	7.696322	6.339718	8.297602	7.430855
	10 dB	10.790927	10.822377	10.874318	11.584533	10.734209
db5	0 dB	4.051875	5.293733	3.844242	4.69209	2.515858
	5 dB	8.13396	8.446362	6.721453	8.908714	7.513789
	10 dB	12.014205	11.593715	**11.980205**	12.693516	**11.560045**
db10	0 dB	3.963526	5.293786	3.690797	4.666482	2.421438
	5 dB	8.11928	8.490163	6.682319	8.956868	7.425254
	10 dB	11.886973	11.544245	11.835829	12.832093	11.528413
db15	0 dB	3.937564	5.210047	3.573668	4.544382	2.335742
	5 dB	8.040169	8.473918	6.608608	8.906647	7.352999
	10 dB	11.897426	11.565462	11.715849	12.853611	11.461991
sym5	0 dB	**4.224198**	**5.418761**	**3.989513**	**4.783298**	2.528579
	5 dB	**8.205666**	8.482623	**6.744722**	8.968946	**7.632089**
	10 dB	11.974079	11.59391	11.921995	12.706739	11.521217
sym10	0 dB	4.056841	5.338404	3.812885	4.781361	2.430903
	5 dB	8.137401	8.46651	6.705735	8.959106	7.493683
	10 dB	12.032354	11.617111	11.938741	12.914808	11.518183
sym15	0 dB	4.224187	5.408234	3.800507	4.737344	2.374583
	5 dB	8.173276	8.503411	6.652598	**8.98189**	7.439504
	10 dB	12.007936	**11.699244**	11.838824	**12.99832**	11.508912
coif3	0 dB	4.106695	5.371772	3.921855	4.765497	2.507153
	5 dB	8.210693	**8.522588**	6.736789	8.943089	7.571337
	10 dB	12.014915	11.610166	11.92487	12.766572	11.531545
coif4	0 dB	4.086557	5.386845	3.890359	4.745099	2.448227
	5 dB	8.201464	8.521521	6.717993	8.935498	7.535606
	10 dB	**12.050653**	11.623474	11.908661	12.857187	11.553395
coif5	0 dB	4.090118	5.366035	3.838619	4.749305	2.435685
	5 dB	8.182087	8.509159	6.699481	8.945622	7.504453
	10 dB	12.029903	11.633294	11.882987	12.913901	11.544564

5. CONCLUSION

Wavelet transform is a modern technology method for speech signal enhancement. In this study, we measured the performance of wavelet families according to different noisy speech signals. The aim of this work was to remove many kinds of background noises from an input speech signal by using wavelet transform method. In the scope of this study, an algorithm was designed in Matlab by the authors to try different wavelet thresholding senarios proposed in this paper for time series analysis of input noisy signals. Then, we tried different thresholding methods for input noisy signals and evaluated the denoising performance of popular wavelet families. We measured the performance of the thresholding methods by evaluating the SNR and MSE values. When we selected Modwt thresholding method together with sqtwolog threshold, we obtained higher SNR than other methods. From the evaluation of the results, it is concluded that the test scenario with Modwt (DhM) gives the best SNR result for all different kinds of noisy speech signals. The Modwt has many advantages over the DWT as implied in this paper. It is a highly redundant and nonorthogonal transform method. This redundancy facilitates alignment of the decomposed wavelet and scaling coefficients at each level, and also decreases the variance of wavelet-based statistical estimates.

The application of the Modwt method to all these noisy speech signals to measure its thresholding performance was a new technique that had not been found in any previous works in literature. This study represented the effectiveness of this technique on different types of noisy speech signals by trying many kind of scenario during the tests. From the results, it can be concluded that the Modwt thresholding method can be used to obtain high quality speech enhancement of noisy speech signals. For a future work on the development of wavelet thresholding methods, a combination of Modwt and adaptive filtering methods for denoising of an input noisy speech signal could be tried. With an adaptive filter, a data dependent definition of a threshold function would minimize the changing performance of Modwt against

different kind of noises. Besides, the effectiveness of Modwt method on any other input noisy signal could be evaluated in a feature work study.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors

REFERENCES

[1] Chavan MS, Mastorakis N, "Studies on implementation of Harr and Daubechies wavelet for denoising of speech signal", *International journal of circuits, systems and signal processing*, 4 (3), 83-96, (2010).

[2] Morlet J, Arens G, Fourgeau E, and Giard D, "Wave propogation and sampling theory. Complex signal and scattering in multilayered media", *Geophysics*, 47(2), 203–221, (1982).

[3] Burrus CS, Gopinath RA, Guo H, Odegard JE, Selesnick IW, "Introduction to wavelets and wavelet transforms: a primer (Vol. 1)". New Jersey: *Prentice hall*, (1998).

[4] Debnath L, Shah FA, "Wavelet transforms and their applications" (pp. 12-14). Boston: *Birkhäuser*, (2002).

[5] Meyer Y, "Wavelets and operators" (Vol. 1). *Cambridge university press*., (1992).

[6] Luna AEV, Nuñez AJ, Lucero DS, Lima CMO, Soto JGA, Gil AF, Alarcon MM, "De-noising audio signals using Matlab wavelets toolbox In Engineering education and research using Matlab". *InTech*., (2011).

[7] Munegowda BK, "Performance and Comparative Analysis of Wavelet Transform in Denoising Audio Signal from Various Realistic Noise". Doctoral dissertation, *Napier University*, Edinburgh, Scotland, United Kingdom, (2016).

[8] Verma N, Verma AK, "Performance analysis of wavelet thresholding methods in denoising of audio signals of some Indian Musical Instruments". *International Journal of Engineering Science and Technology*, 4 (5), 2040-2045, (2012).

[9] Cornish CR, Bretherton CS, "Maximal Overlap Wavelet Statistical Analysis with Application to Atmospheric Turbulence". *Journal of Boundary-Layer Meteorology*, 119, 339–374, (2006).

[10] Percival DB, Burnell AC, Walden AT, "Wavelet Methods for Time Series Analysis". *Cambridge University Press*, (2000).

[11] Polikar R, "The wavelet tutorial". (1996).

[12] Yuan X, "Auditory model-based bionic wavelet transform for speech enhancement". Doctoral dissertation, *Marquette University*, (2003).

[13] Venkateswarlu SC, Reddy AS, Prasad KS, "Speech Enhancement in terms of Objective Quality Measures Based on Wavelet Hybrid Thresholding the Multitaper Spectrum". *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 5 (1), 201-219, (2016).

[14] Kim E, "A Wavelet Transform Module for a Speech Recognition Virtual Machine". Doctoral dissertation, *Minnesota State University*, Mankato, (2016).

[15] Kumari VSR, Devarakonda DK, "A wavelet based denoising of speech signal". *International Journal of Engineering Trends and Technology (IJETT),* Vol.5, 107-115, (2013).

[16] Donoho DL, Johnstone IM, "Ideal Denoising In an Orthonormal Basis Chosen From A Library of Bases". *Comptes Rendus De L Academie Des Sciences Serie I-Mathematique*, 319 (12), 1317-1322, (1994).

[17] Donoho DL, Johnstone IM, "Ideal spatial adaptation by wavelet shrinkage". *Biometrika*, 81 (3), 425–455, (1994).

[18] Misiti M, Misiti Y, Oppenheim G, Poggi JM, "Wavelet toolbox." *The MathWorks Inc*., Natick, MA, 15, 21, 1996.

[19] Loizou P, "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms". *Speech Communication*, 49, 588-601, (2017).

[20] Hirsch HG, Pearce D, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions". *ISCA Tutorial and Research Workshop (ITRW) ASR2000*, September 2000.

[21] Daubechies I, "Orthonormal bases of compactly supported wavelets". *Comm. Pure Appl. Math*., 41, 909-996, (1988).

[22] Mihov S.G., Ivanov, R.M., & Popov, A. N., "Denoising Speech Signals by Wavelet Transform", *Annual Journal of Electronics*, ISSN 1313-1842, pp:69-72, (2009).

[23] Sumithra, A., & Thanushkodi, B. "Performance evaluation of different thresholding methods in time adaptive wavelet based speech enhancement". *IACSIT Int. J. Eng. Technol*, 1(5), 42-51. (2009).

[24] Sanam, T.F. & Shahnaz, C. "A semisoft thresholding method based on Teager energy operation on wavelet packet coefficients for enhancing noisy speech", *EURASIP Journal on Audio, Speech, andMusic Processing* 2013:25, pp.2-15, https://doi.org/10.1186/1687-4722-2013-25.

[25] Du, L., Xu, R., Xu, F., et al, "Research on key parameters of speech denoising algorithm based on wavelet packet transform", *IEEE 3rd International Conference on Computer Science and Information Technology*, 2010, pp:551-556, https://doi.org/ / 10.1109/ICCSIT.2010.5564729.