A Heuristic Speech De-noising with the aid of Dual Tree Complex Wavelet Transform using Teaching-Learning Based Optimization

D. Yugandhar [#1], S.K. Nayak [*2]

[#] E.C.E Department, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India,
[1] yugndasari@gmail.com
[*] Electronic Science Department, Berhampur University, Berhampur, Odisha, India
[2] sknayakbu@rediffmail.com

Abstract— In our present work, we propose a nature inspired population based speech enhancement technique to find the dynamic threshold value using Teaching-Learning Based Optimization (TLBO) algorithm by using shift invariant property of dual tree complex wavelet transform (DT-ℂWT). The performance of these proposed methods are evaluated in terms of Perceptual Evaluation of Speech Quality (PESQ) and Peak Signal to Noise Ratio (PSNR). Speech quality of different speech waves are compared for two level wavelet packet decomposition and dual tree wavelet transform using soft threshold. The speech qualities of the waves are better than the other available articles in the literature.

Keyword- Speech enhancement, Dual tree complex wavelet transform, Teaching-Learning-based optimization, Perceptual Evaluation of Speech Quality

I. INTRODUCTION

In many speech related systems, the original speech signal is contaminated with some interference sources. The interference source, i.e the noise signal degrades the quality of the original speech. The speech signal is affected by wide-band noise in the form of white or colored noise and a period noise such as hum noise. The most common type of noise in communication channels is the additive wide band Gaussian noise. Speech enhancement aims at improving the performance of speech communication systems in noisy environments. Speech enhancement may be applied, for example, to a mobile radio communication system, a speech recognition system, a set of low quality recordings, or to improve the performance of aids for the hearing impaired. Several methods [1]-[6] have been proposed in the literature for the enhancement of degraded speech. A majority of these methods can be grouped into spectral processing and temporal processing methods. In the spectral processing methods, degraded speech is processed in the frequency domain mostly using Fourier transform for achieving speech enhancement. In the temporal processing methods, the processing is done in the time domain. Most speech enhancement methods improve the quality of the signal but degrade its intelligibility of the speech. Performance measures like PSNR and PESQ are widely used as the performance of the evolution criterion. For elimination of the Gaussian back ground noise in the communication channels, we have been implementing adaptive thresholding technique using TLBO optimization.

The TLBO algorithm [7]-[9] is a global optimization, population based iterative learning mechanism that exhibits some common characteristics with other evolutionary computation (EC) techniques like (GA) [10], Particle Swarm Optimization (PSO) [11], Differential Evolution (DE) [12], and Artificial Bee Colony (ABC) [13]. The TLBO algorithm does not require any algorithm-specific control parameters like mutation and crossover as in genetic algorithm. The TLBO methods provide the learning mechanism in adaptive models.

The organization of this paper is as follows. Section II presents speech de-nosing using wavelet thresholding. Section III discusses introduction of TLBO algorithm, Section IV discusses the proposed method of implementation. The simulation results and comparisons are given in section V. Finally the paper concludes with section VI.

II. SPEECH DE-NOISING USING WAVELET THRESHOLDING

The Fourier transform is the basic tool in frequency domain approach. Short time Fourier transform (STFT) [14] provides one of many ways to generate a time frequency analysis of signals. However, fixed time-frequency resolution of the STFT poses a serious constraint because the width of the window function is fixed. A variable window size is required to represent signals at different resolution. Another linear transform that provides such a time frequency analysis is known as continuous wavelet transform (CWT). This CWT gives a window function whose width increase in time while resolving the low-frequency contents, and decreases in time while resolving

D. Yugandhar et al. / International Journal of Engineering and Technology (IJET)

the high-frequency contents of a signal, which leads to good time-frequency resolution. The CWT of a function $f(t) \in L^2$ with respect to some analyzing wavelet ψ is given by [15]

$$W_\psi f(b, a) = \int_{-\infty}^{\infty} f(t) \, \overline{\psi}_{b,a}(t) \, dt \qquad (1)$$

where,

$$\psi_{b,a}(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a}), a > 0 \qquad (2)$$

The parameters b and a in (1) are the translation and dilation (scaling) parameters respectively. The normalization factor $\frac{1}{\sqrt{a}}$ is included so that $\|\psi_{b,a}\| = \|\psi\|$. The expression of the inverse transform is to reconstruct the original function from its integral wavelet transform is given by

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} db \int_{-\infty}^{\infty} \frac{1}{a^2} [W_\psi f(b, a)] \psi_{b,a}(t) \, da \qquad (3)$$

where, C_ψ is a constant that depends on the selection of the wavelet and is given by

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \qquad (4)$$

The condition above is known as the admissibility condition, restricts the class of functions like wavelets. Finding inverse wavelet transformation for synthesizing the original signal in (3) is quite cumbersome because it involves a two dimensional integration over the scale parameter a and the translation parameter b. If we consider scale parameter a to be of the form 2^{-s} and translation parameter b of the form $k2^{-s}$, where $k, s \in Z$, with these values of a and b, the discrete form of integral in (1) representation is known as discrete wavelet transform (DWT).

$$W_\psi f(k2^{-s}, 2^{-s}) = 2^{s/2} \sum_n f(n) \psi(2^s n - k) \qquad (5)$$

Let us consider the discrete function $f_m(n)$ which is delayed by m samples represented as $f_m(n) = f(n-m)$, and then its DWT is computed as given in (6). The important observation from (5)-(6) is that the DWT of a function shifted in time is quite different from the DWT of the original function. Therefore, the DWT transform is shift variant transformation. The wavelet transform can be regarded as a bank of band-pass filters with constant Q factor. The advantage of using variable size windows for different frequency bands is seen in the wavelet analysis.

$$\begin{aligned} W_\psi f_m(k2^{-s}, 2^{-s}) &= 2^{s/2} \sum_n f(n-m) \psi(2^s n - k) \\ &\approx 2^{s/2} \sum_n f(n) \psi[(2^s n - (k - m2^s)] \\ &\approx W_\psi f[(k - m2^s) 2^{-s}, 2^{-s}] \end{aligned} \qquad (6)$$

The wavelet analysis has a distinct ability to detect local features of the signal in both time and frequency, such as the plosive fine structures of the speech and other transient, instantaneous and dynamic speech components that contribute significantly to the quality of the speech. Therefore, wavelet transform can provide an appropriate model of speech signal for de-noising applications. One popular technique for wavelet based signal enhancement is the wavelet shrinkage algorithm [16]. Wavelet shrinkage is a simple de-noising method based on the thresholding of the wavelet coefficients. The estimate threshold defines the limit between the wavelet coefficients of the noise and those of the target signal. However, it is not always possible to separate the components corresponding to the target signal from those of noise by simple thresholding. Applying thresholding uniformly to all wavelet coefficients not only suppress additional noise but also some speech components like unvoiced ones which leads to the loss of perceptual quality of the filtered speech. In order to improve perceptual speech quality various thresholding and estimation techniques [17] have been proposed.

Assuming that $s(n)$ represents the clean speech with finite duration, $x(n)$ stands for the speech corrupted by white Gaussian noise $n(n)$ is having zero mean and variance σ^2 is shown in (7) and corresponding wavelet domain representations is shown in (8)

$$x(n) = s(n) + n(n) \tag{7}$$

$$X = S + N \tag{8}$$

If W denotes a wavelet transform matrix, (8) represents wavelet coefficients, where $X = W_X, S = W_s$, and $N = W_n$ are noisy speech, original speech and noise wavelet coefficients respectively. The de-noised signal $\hat{S}$ is obtained by soft thresholding noisy wavelet coefficients which can be represented as

$$\hat{S} = THR(X, T) \tag{9}$$

where $THR(.)$ denotes a thresholding function and T denotes the threshold value. Thresholding of coefficients can be in done in many ways. However there are two popular versions known as hard thresholding and soft thresholding. Hard thresholding sets any wavelet coefficient whose absolute value is less than or equal to threshold is to zero while the others are kept unchanged.

$$\hat{X} = \begin{cases} X & , if \ |X| > T \\ 0 & , if \ |X| \leq T \end{cases} \tag{10}$$

The soft thresholding is similar to the hard thresholding except that it either shrinks or kills (set to zero) coefficients based on the threshold condition given below.

$$\hat{X} = \begin{cases} \text{sgn}(X)(|X| - T) & , if \ |X| > T \\ 0 & , if \ |X| \leq T \end{cases} \tag{11}$$

The Donoho [18] showed that when noise dominates the observed data, the universal threshold method performs well and when the underlying signal dominates the observed data, the SURE method [19] performs better than universal Threshold method. Donoho and Johanstone [16] proposed the universal threshold value as $\sigma\sqrt{2\log(N)}$ and standard deviation $\sigma = MAD/0.6745$. MAD denotes the absolute median value on the first scale of the detailed wavelet coefficients and N is the length of the noisy signal. Shima [20], obtained threshold value basing upon the symmetric Kullback- Leibler divergence between the probability distributions of noise wavelet coefficients and noisy speech and finally obtained the improved threshold value using the segment Signal to Noise Ratio.

It can be observed that the soft thresholding method removes more noise components than the hard thresholding method which leads the signal degradation to higher range. So, an appropriate thresholding technique is to be selected for optimal enhancement of speech signals. Besides this, some other thresholding techniques like Stein's unbiased estimate selection rule, Heuristic threshold selection rule and Minimax performance threshold selection rule has been used in normal practice.

A. *The Dual Tree Complex Wavelet Transform (DT-$\mathbb{C}$WT)*

The wavelet transform suffers from four fundamental shortcomings namely oscillations, shift variance, aliasing and lack of directionality [21]. The DT-$\mathbb{C}$WT is a recent enhancement to the DWT with complex valued scaling function $\phi_c(t)$ and complex valued wavelet function $\psi_c(t)$ which can be represented as

$$\phi_c(t) = \phi_r(t) + j\phi_i(t) \tag{12}$$

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \tag{13}$$

where, $\psi_r(t)$ and $\phi_r(t)$ are real and even function of t and $\psi_i(t)$ and $\phi_i(t)$ are imaginary and odd function of t. However, $\psi_r(t)$ and $\psi_i(t)$ form a Hilbert transform pair for wavelet function and similarly $\phi_r(t)$ and $\phi_i(t)$ form Hilbert transform for scaling function. Hence, $\phi_c(t)$ and $\psi_c(t)$ are both analytic signal and supported on only one half of the frequency axis. The DT-$\mathbb{C}$WT introduced by Kingsbury, it consists of two real discrete wavelet transforms. The first DWT gives the real part of the transform while the second DWT gives the imaginary part. The two real DWTs use two different sets of perfect reconstruction (PR) filter which are jointly designed so that the overall transform is approximately analytic. The analysis filter banks used to implement

DT-$\mathbb{C}$WT of a signal $x(n)$ is shown in Fig. 1 [21]. The two real wavelet transforms use two different sets of filters, with each satisfying the PR conditions. The two sets of filters are jointly designed so that the overall transform is approximately analytic. The filters $h_0(n), h_1(n)$ denote the low-pass and high-pass conjugate quadrature filter (CQF) pair for the upper bank. The autocorrelation of the filter can be expressed as

$$\sum_n h_0(n)h_0(n+2k) = \delta(k) = \begin{cases} 1 \ for \quad k = 0 \\ 0 \ for \quad k \neq 0 \end{cases} \tag{14}$$

$$h_1(n) = (-1)^{(1-n)} h_0(n-1) \tag{15}$$

Similarly, the filters $g_0(n), g_1(n)$ form another CQF pair which denotes the low-pass and high –pass filters respectively for the lower filter bank. The DT-$\mathbb{C}$WT of an input real vector X can be represented by the rectangular matrix given below

$$F = \begin{bmatrix} F_h \\ F_g \end{bmatrix} \tag{16}$$

Let $w_h = F_h X$ and $w_g = F_g X$ represents real part and imaginary part of the DT-$\mathbb{C}$WT then $w_h + j w_g$ represents complex dual tree wavelet coefficients.
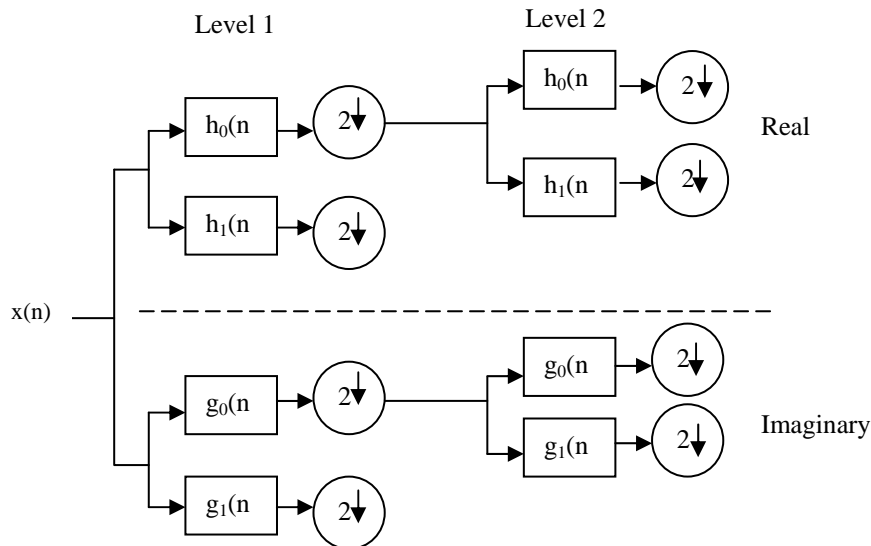

Fig. 1 Filter bank of Two Level Dual Tree Complex Wavelet Transform

III. INTRODUCTION OF TLBO ALGORITHM

In this paper, we have adapted TLBO algorithm shown in Fig. 2 [7] to find the better threshold value in different scales of DT-$\mathbb{C}$WT. The algorithm is based upon two phases like teacher phase and learner phase.

A. Teacher phase

In the teacher phase of the algorithm the knowledge flow is from the teacher to the learners. During the teacher phase a teacher tries to increase the mean result of the entire class in the subject taught by him or her depending on his or her skills and knowledge. In the beginning of the algorithm let say M_j be the mean of each subject and T_i be the teacher i.e most learned person at any iteration. Now teacher T_i will improve existing mean M_j based on his skills and knowledge so that the new mean will be designated as M_{new}. The difference between the new mean and existing mean is given as

$$Difference\_Mean_i = r_i(M_{new} - T_F M_j) \tag{17}$$

where r_i is the random number which takes the values in the range 0 and 1, T_F known as teaching factor, it is not a parameter of TLBO algorithm. The value of T_F is not specified in the beginning of the execution of the

algorithm. The value of T_F is randomly selected by the algorithm itself with equal probability using the equation (18).
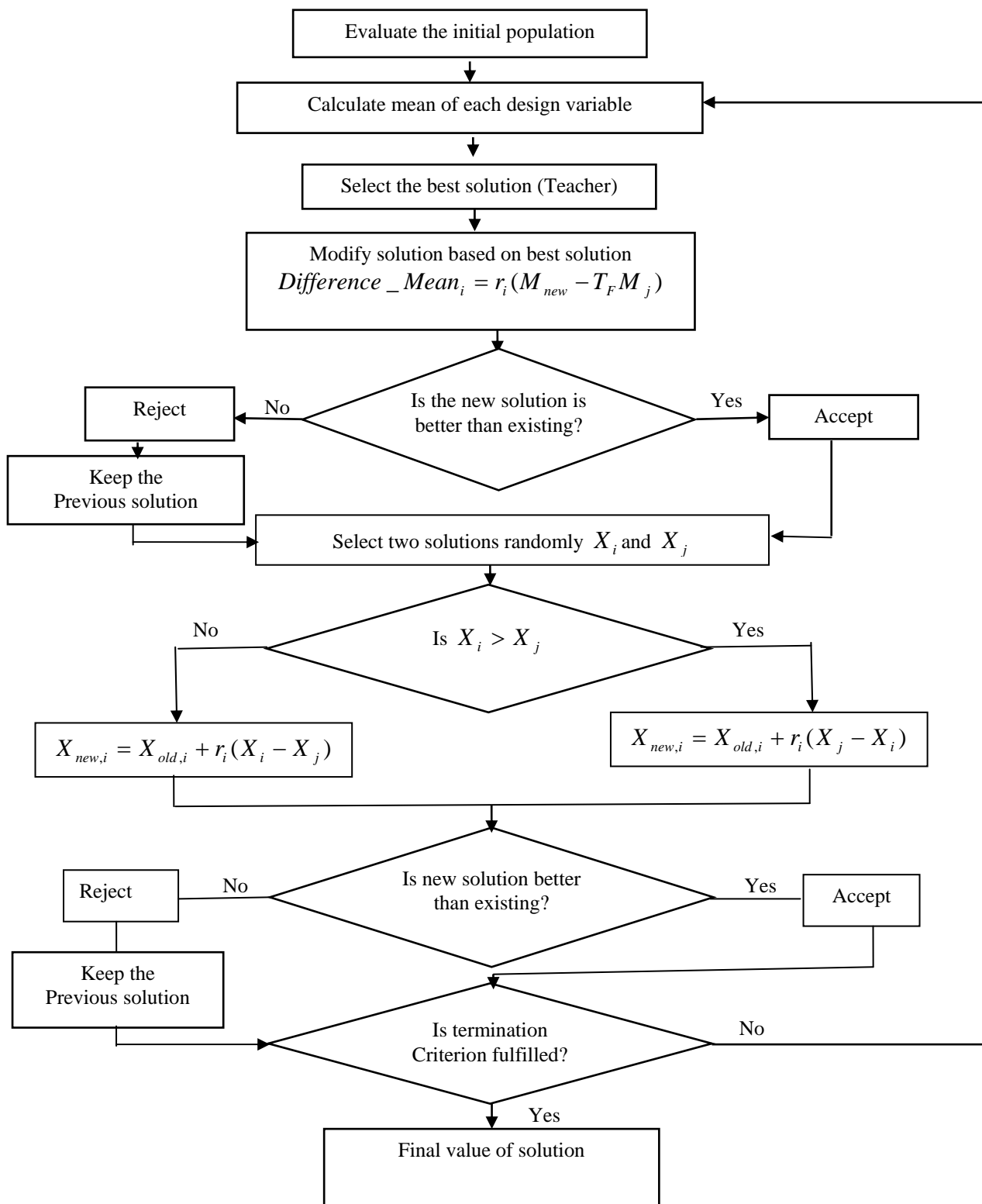

Fig. 2. Flowchart of TLBO algorithm.

$$T_F = round[1 + rand(0,1)] \tag{18}$$

From (18), it is concluded that the teaching factor value (T_F) is in between 1 and 2. Based on the $Difference\_Mean$, the existing solution is updated in the teacher phase using the equation given below

$$X_{new,i} = X_{old,i} + Differenc\_Mean_i \tag{19}$$

where $X_{new,i}$ is the updated solution value of existing solution $X_{old,i}$. Accept $X_{new,i}$ if it gives better fitness value. At the end of the teacher phase all the accepted function values are stored and these values become the input to the learner phase. The learner influence and knowledge depends on the outcome of the teacher phase.

B. Learner Phase

 It is the second part of the TLBO algorithm where Learners increase their knowledge through interaction among themselves. A learner interacts randomly with other learners to improve his or her knowledge. A learner can enhance their knowledge and able to learn new things if the other learner has more knowledge than him or her. Mathematically, the learning phenomenon of this learning phase is expressed below. At any iteration i , consider two different learners X_i and X_j where $i \neq j$

$$X_{new,i} = X_{old,i} + r_i(X_i - X_j), \text{if } F(X_i) > F(X_j) \tag{20}$$

$$X_{new,i} = X_{old,i} + r_i(X_j - X_i), \text{if } F(X_j) > F(X_i) \tag{21}$$

where, $F(X_i), F(X_j)$ are known as the learners knowledge or fitness value of the learners X_i and X_j respectively.

IV. PROPOSED METHOD OF IMPLEMENTATION

Different speech de-noising simulations have been carried out to verify the de-noising performance of the speech signal using TLBO algorithm. In these simulations, original speech signals from NOIZEUS speech database were considered. The input PSNR of 8.3 dB and 10 dB are added to the original speech signals to conduct two different simulations. The training process to achieve the optimized threshold value is shown in Fig. 3. The two threshold values obtained from the training process is used as the input to de-noise speech signals corrupted with input PSNR of 8.3 dB and 10 dB is shown in Fig. 4.

A. Implementation of TLBO: Training:

Step 1: Clean speeches from 'sp_1.wav' to 'sp_10.wav' were taken from the database.

Step 2: Noisy input signals are generated by adding input PSNR of 8.39dB and 10dB.

Step 3: Populations taken as 30 learners and two subjects are considered as design variables and 100 iterations were considered to train with TLBO algorithm.

Step 4: Each noisy speech signal is decomposed using two level DT-ℂWT whose analysis and synthesis filter coefficients have shown in Table I & II respectively. The inverse DT-ℂWT can be found after soft thresholding details wavelet coefficients by using (11).

Step 5: PESQ is selected as fitness function and the value is measured between corresponding de-noised and original signal of all the ten speech signals. The mean PESQ value of all ten signals is considered as fitness function to be maximized.

Step 6: Iterations 100 were conducted to get optimized threshold.

TABLE I Analysis dual tree complex wavelet coefficients

First Stage Real Coefficients		First Stage Complex Coefficients		Second Stage Real Coefficients		Second Stage Complex Coefficients	
LPF	HPF	LPF	HPF	LPF	HPF	LPF	HPF
0	0	0.011227	0	0.035164	0	0	-0.03516
-0.08839	-0.01123	0.011227	0	0	0	0	0
0.088388	0.011227	-0.08839	-0.08839	-0.08833	-0.1143	-0.1143	0.088329
0.69588	0.088388	0.088388	-0.08839	0.23389	0	0	0.23389
0.69588	0.088388	0.69588	0.69588	0.760272	0.587518	0.587518	-0.76027
0.088388	-0.69588	0.69588	-0.69588	0.587518	-0.76027	0.760272	0.587518
-0.08839	0.69588	0.088388	0.088388	0	0.23389	0.23389	0
0.011227	-0.08839	-0.08839	0.088388	-0.1143	0.088329	-0.08833	-0.1143
0.011227	-0.08839	0	0.011227	0	0	0	0
0	0	0	-0.01123	0	-0.03516	0.035164	0

TABLE II Synthesis dual tree complex wavelet coefficients

First Stage Real Coefficients		First Stage Complex Coefficients		Second Stage Real Coefficients		Second Stage Complex Coefficients	
LPF	HPF	LPF	HPF	LPF	HPF	LPF	HPF
0	0	0	-0.01123	0	-0.03516	0.035164	0
0.011227	-0.08839	0	0.011227	0	0	0	0
0.011227	-0.08839	-0.08839	0.088388	-0.1143	0.088329	-0.08833	-0.1143
-0.08839	0.69588	0.088388	0.088388	0	0.23389	0.23389	0
0.088388	-0.69588	0.69588	-0.69588	0.587518	-0.76027	0.760272	0.587518
0.69588	0.088388	0.69588	0.69588	0.760272	0.587518	0.587518	-0.76027
0.69588	0.088388	0.088388	-0.08839	0.23389	0	0	0.23389
0.088388	0.011227	-0.08839	-0.08839	-0.08833	-0.1143	-0.1143	0.088329
-0.08839	-0.01123	0.011227	0	0	0	0	0
0	0	0.011227	0	0.035164	0	0	-0.03516

Training phase

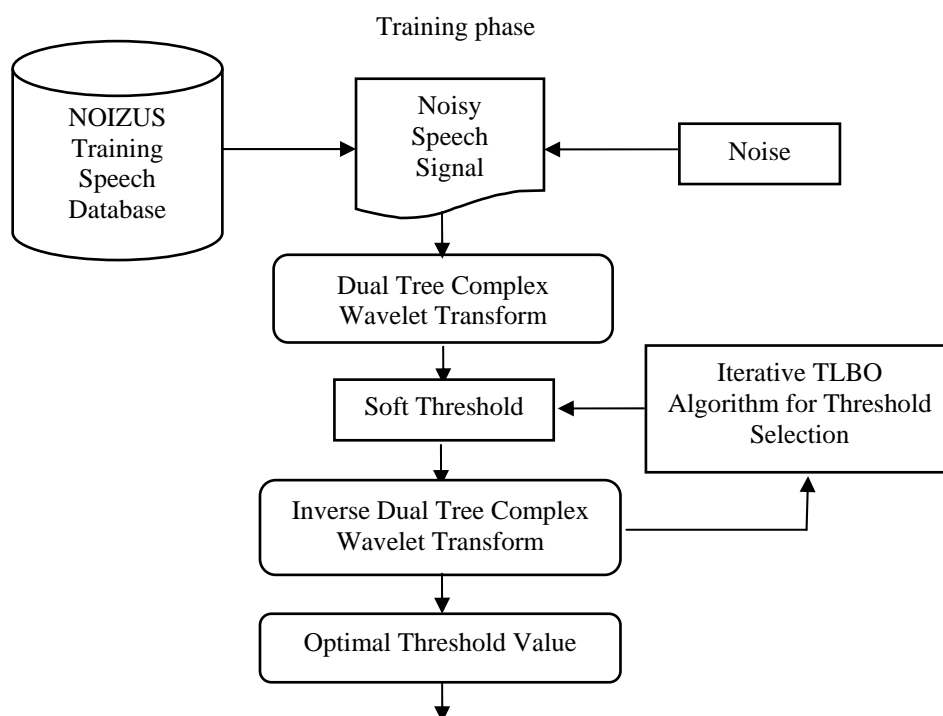
Fig 3: Architecture of proposed speech de-noising system for training

B.   De-noising of speech signals: Testing

Step 1: Noisy speech signals 'sp_11.wav' to 'sp_20.wav' are generated by considering input PSNR of 8.39dB and 10dB.

Step 2: The obtained optimum threshold values from the training process are used to de-noise the noisy speech signals using two level DT-ℂWT by soft thresholding details wavelet coefficients by using (11). The architecture of the proposed system for de-noising of speech signals with voice activity detection (VAD) has shown in Fig. 4. The VAD method proposed in [22][23] to identify both voice and silence zones has been implemented using Texas instruments TMS320C6713 floating point digital signal processor operating at 225MHz to check the performance of speech de-noising in real time.

Step 3: PSNR and PESQ performance measures were calculated between original and de-noised speech signals.

C.   Performance metrics of speech de-noising:

Many speech quality measures were proposed in the literature [24]. They are segmental Signal to Noise Ratio (SNR), Itakura-Saito distance measure, Cepstrum distance measures, output PSNR and PESQ.

1)   *PESQ:* Among all other objective measures mentioned above, the computational burden is more when PESQ measure is computed between original and de-noised signal. This measure is recommended by international telecommunication union standardization sector (ITU-T) for speech quality assessment of 3.2 kHz handset telephony as described [24] , the PESQ is computed as a linear combination of the average asymmetrical disturbance values A_{ind} and the average disturbance value (D_{ind}) which can be expressed as

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \qquad (22)$$

where, $a_0 = 4.5$, $a_1 = -0.1$ and $a_2 = -0.0309$ are constants. The parameters in the above equation were optimized for speech processed through networks.

2)   *Output PSNR:* Here, we have been evaluating the performance of the algorithm by considering output PSNR. It is the ratio between the power of output signal and the power of noise in decibel scale. PSNR is most commonly used to measure the quality of reconstructed signal.

$$Output\,PSNR\,in\,dB = 10*\log_{10}\frac{\sigma_{\hat{s}}^2}{\sigma_e^2}, \quad where \quad \sigma_e^2 = \sum_{i=1}^{M}\left(S(n)-\hat{S}(n)\right)^2 \qquad (23)$$

where $M, S(n), \hat{S}(n), \sigma_{\hat{s}}^2$, and σ_e^2 are the length of the signal, original speech signal, de-noised speech signal, power of de-noised signal and noise power respectively.
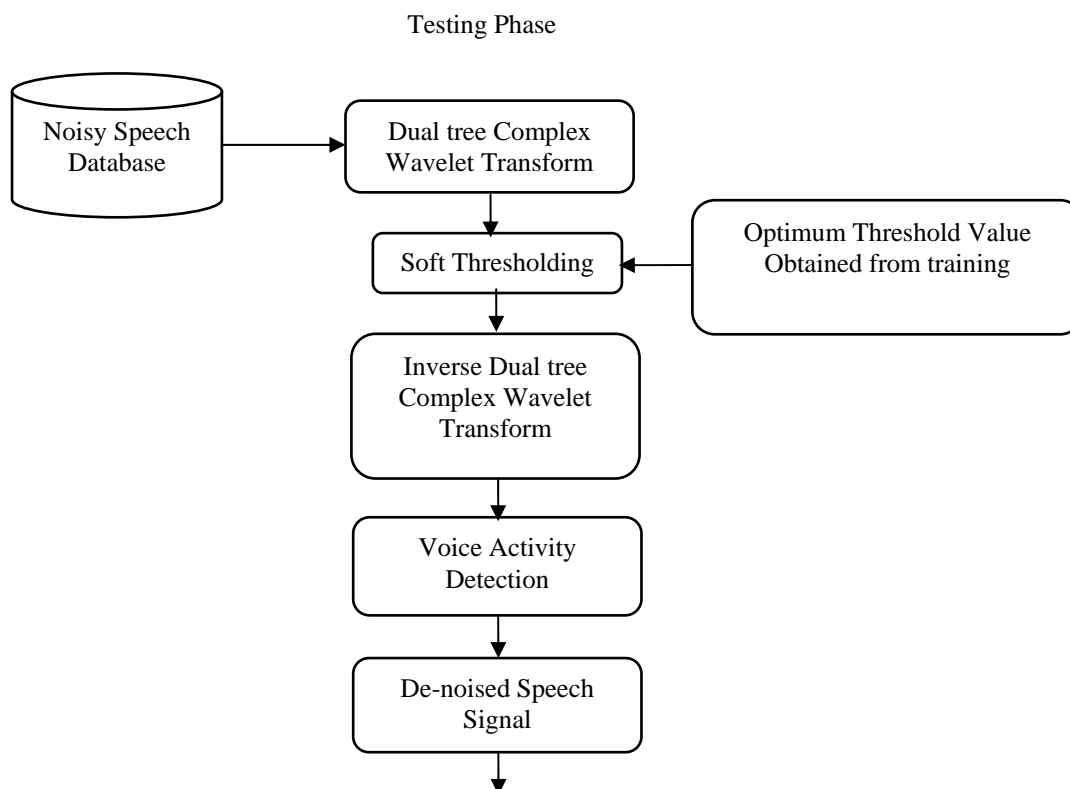
Testing Phase

Fig 4: Architecture of proposed system for speech de-noising with VAD

V.   SIMULATION RESULTS AND COMPARISONS

For simulation purpose we have been considering a population size of 30 learners, solution length of 2 and 100 iterations. An extensive evaluations and comparisons have done by taking two, three and four levels of decomposition of a signal in the dual tree complex wavelet domain. Two level DT-ℂWT gives good result comparatively other levels of decompositions. The PSNR and PESQ measures were given for input PSNR of 8.39dB and 10dB shown in the Table III and Table IV and the corresponding speech signal outputs were shown in Fig. 5 & Fig. 6 respectively. The de-noised speech signal using TLBO preserves both voiced and unvoiced speech comparatively using conventional DT-ℂWT were shown in Fig. 5(d) and 6(d) for input PSNR of 8.39dB and 10dB respectively. The output of de-noised speech signals after making silence zones to zero are shown in Fig. 5(f) and 6(f). The output of VAD using TMS320C6713 is also shown in Fig. 7. Comparisons charts of PSNR and PESQ when input noise SNR of 10 dB was shown in Fig. 8 & Fig. 9.

TABLE III Output PSNR and PESQ of all three methods when input PSNR is 8.39 dB

S.No	Speech File Name	Wavelet Packet		Dual Tree Wavelet		TLBO	
		PSNR	PESQ	PSNR	PESQ	PSNR	PESQ
1	Sp_11.wav	8.2528	2.0183	6.5668	2.1776	10.4938	2.3770
2	Sp_12.wav	7.8354	2.1370	6.8485	2.2693	9.0912	2.3796
3	Sp_13.wav	8.2303	1.9996	6.4831	2.1334	10.7449	2.5225
4	Sp_14.wav	8.4649	1.8012	7.7529	1.9851	10.7499	2.3006
5	Sp_15.wav	8.4862	2.0020	7.3168	2.1416	11.2276	2.3510
6	Sp_16.wav	7.8576	1.9549	7.4149	2.1532	8.3971	2.3534
7	Sp_17.wav	7.4630	1.9989	6.6714	2.1175	8.9746	2.2946
8	Sp_18.wav	8.3997	2.0700	8.0891	2.2741	8.8466	2.4239
9	Sp_19.wav	8.8636	2.0483	8.1149	2.1779	10.4765	2.3725
10	Sp_20.wav	6.9259	1.9248	5.7347	2.0050	8.8908	2.4487

D. Yugandhar et al. / International Journal of Engineering and Technology (IJET)

TABLE IV Output PSNR and PESQ of all three methods when input PSNR is 10dB

S.No	Speech File Name	Wavelet Packet		Dual Tree Wavelet		TLBO	
		PSNR	PESQ	PSNR	PESQ	PSNR	PESQ
1	Sp_11.wav	9.5020	2.1522	7.5974	2.2336	11.3361	2.4478
2	Sp_12.wav	9.0322	2.2494	7.9362	2.3724	9.8432	2.4920
3	Sp_13.wav	9.2396	2.1186	7.3644	2.2002	11.6156	2.6381
4	Sp_14.wav	9.6371	1.8958	8.8275	2.0598	11.5351	2.3868
5	Sp_15.wav	9.8315	2.1351	8.2933	2.2298	12.4893	2.4873
6	Sp_16.wav	9.0295	2.0332	8.5034	2.2188	9.1216	2.4487
7	Sp_17.wav	8.5217	2.1056	7.3212	2.2384	9.5314	2.4022
8	Sp_18.wav	9.4570	2.1581	8.8905	2.2755	9.4834	2.4777
9	Sp_19.wav	9.9850	2.1090	8.9986	2.2198	11.1622	2.4421
10	Sp_20.wav	8.1596	2.0022	6.5909	2.0446	9.4689	2.5932
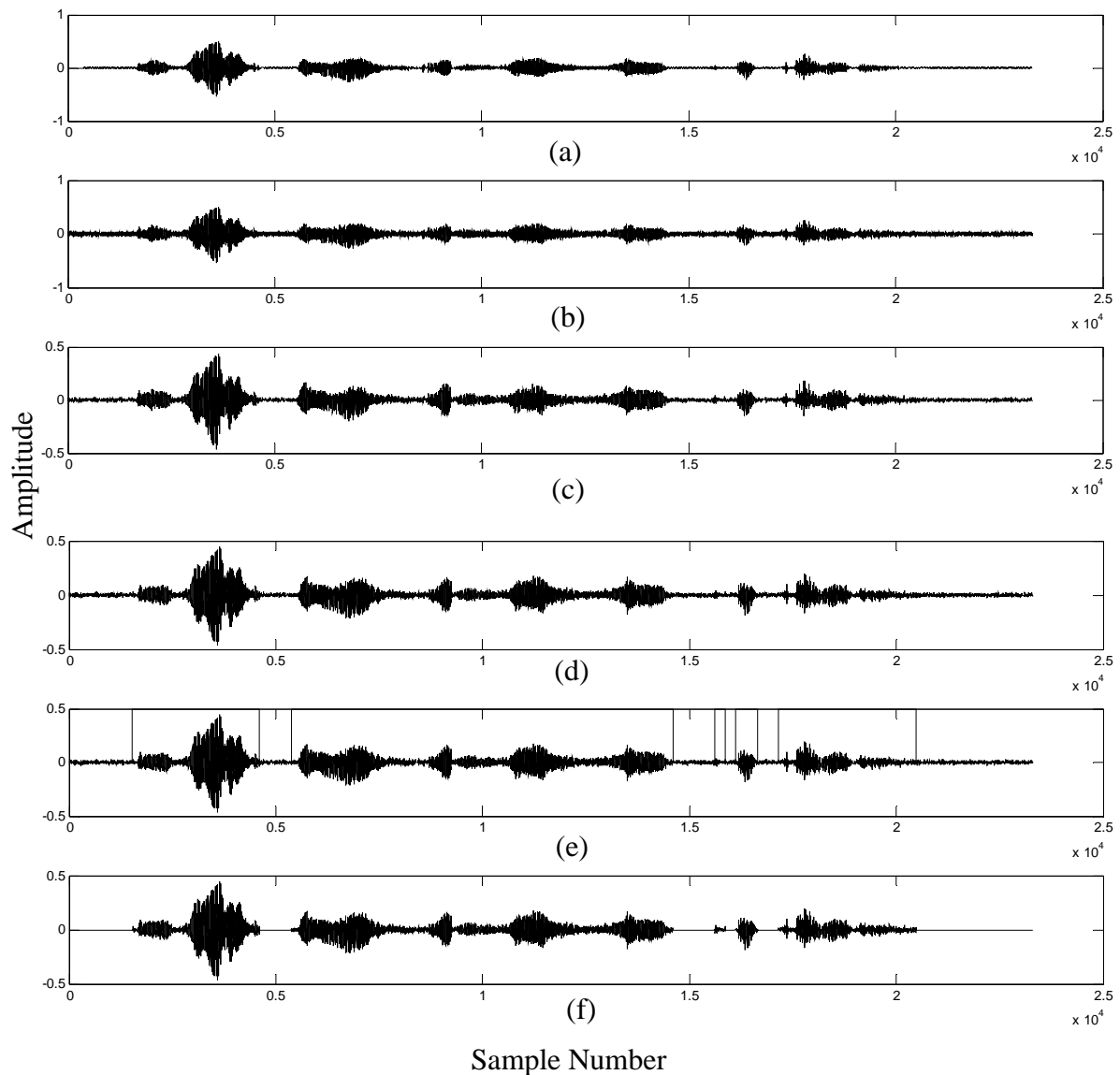
Fig. 5: TLBO with DT-ℂWT output for speech corrupted with white noise with input PSNR 8.39 dB: (a) Original speech signal (b) Noisy speech signal (c) De-noised speech using DT-ℂWT. (d) De-noised speech using TLBO with DT-ℂWT (e) Voiced and silence zones of TLBO output (f) TLBO voice activity detection output making silence zones zero
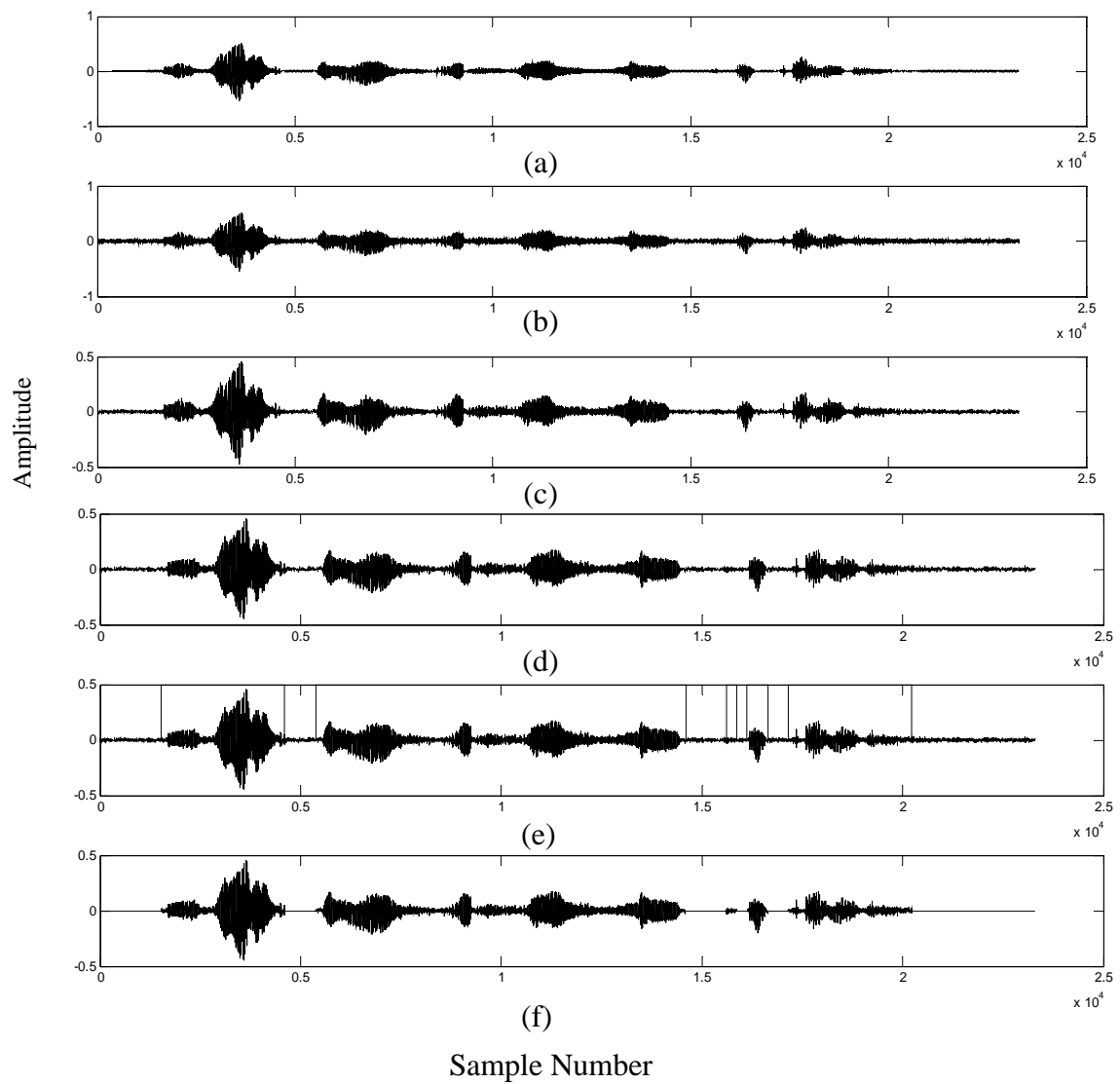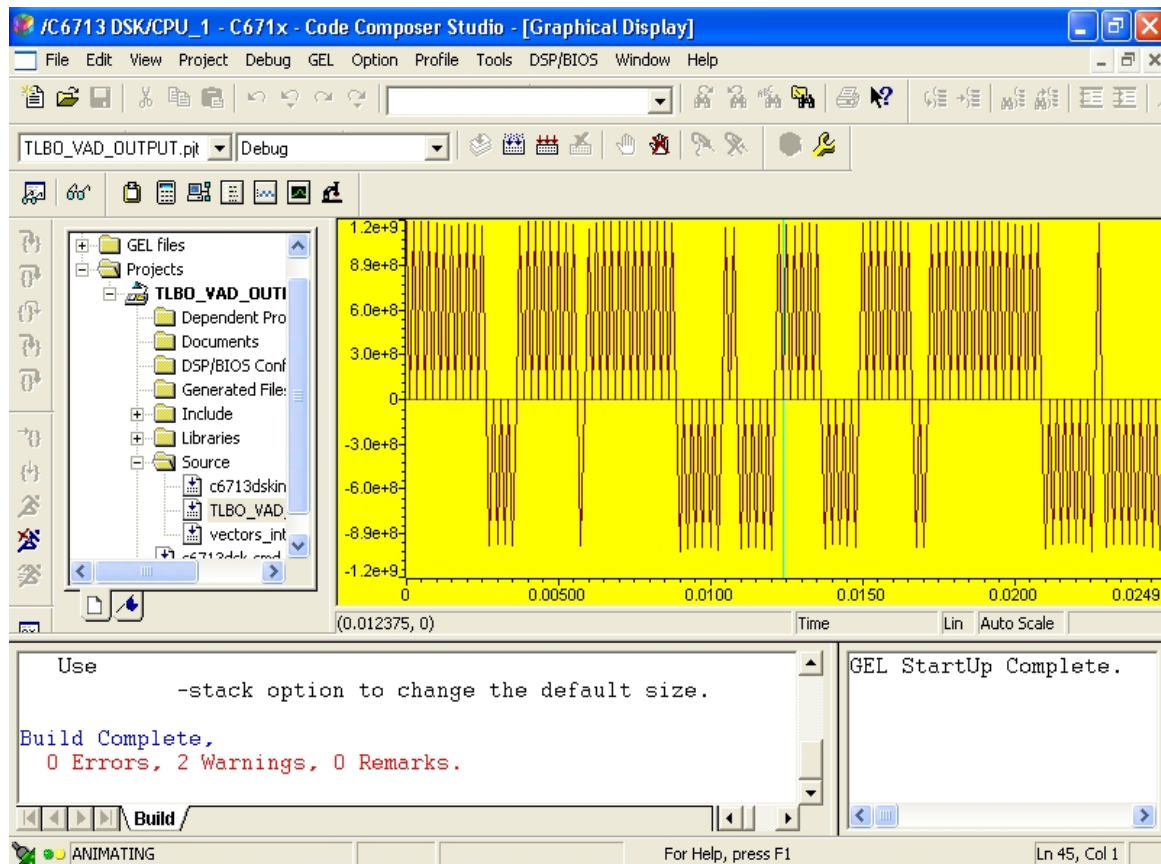
Fig. 6: TLBO with DT-ℂWT output for speech corrupted with white noise with input PSNR 10dB: (a) Original speech signal (b) Noisy speech signal (c) De-noised speech using DT-ℂWT. (d) De-noised speech using TLBO with DT-ℂWT (e) Voiced and silence zones of TLBO output (f) TLBO voice activity detection output making silence zones zero.

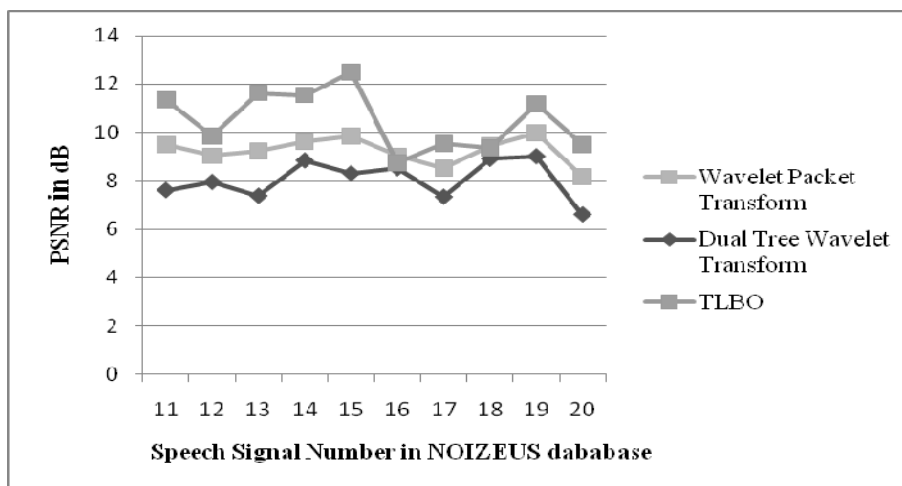Figure 7: VAD output using TMS320C6713 digital signal processor
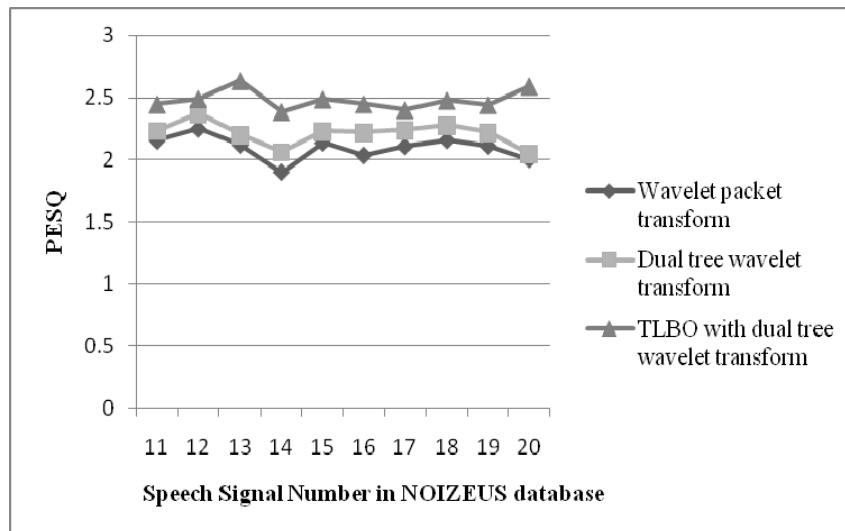

Figure 8: Comparison graph of PSNR for input PSNR 10 dB

Figure 9: Comparison graph of PESQ for input PSNR 10 dB

VI. CONCLUSION

TLBO algorithm is one of the meta- heuristic algorithm has been proposed recently for solving engineering optimization problems. The soft thresholding shrinks wavelet coefficients based on the threshold value leads to decrease in the signal energy as well as noise. Therefore, the output PSNR is less than the input PSNR in the case of wavelet packets and complex dual tree wavelet transform. The threshold obtained by TLBO algorithm preserves the signal energy as well as the quality of the speech signal leads to increase in output PSNR as compared to other two methods. TLBO method of de-noising is able to provide 15%, 18% improvement in average output PSNR and average PESQ than wavelet packet transform and 32%, 12% improvement in average output PSNR and average PESQ than conventional dual tree complex wavelet transform. The Listening tests using DSK 6713 processor show that the de-noised speech signal obtained by TLBO preserving both voiced and unvoiced sounds.

ACKNOWLEDGMENT

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal processing.*, vol. 27, no. 2, pp. 113-120, 1979.
[2] P.C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, First edition, Boca Raton, FL,2007.
[3] M. Berouti, R.Schwartz and J.Makhoul, "Enhancement of speech corrupted by acoustic noise," *IEEE Proceedings of International Conference on Acoust., Speech, Signal process.,* 1979, vol. 4, pp. 208-211.
[4] P.Lockwood, and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars," *Speech Communication*, vol. 11, no. 3, pp. 215-228, 1992.
[5] S. Kamath and P.A. Loizou, "A multi band spectral subtraction methods for enhancing speech corrupted by colored noise," *IEEE Proceedings of International Conference Acoust., Speech, Signal process.*, vol. 4 , pp. IV-4164, 2002.
[6] Y. Ephraim and D. Malah, "Speech Enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal process.*, vol. 33, no. 2, pp. 443-445, 1985
[7] R.V.Rao, V.J.Savsani and D.P.Vakharia, "Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems," *Computer Aided Design*, vol. 43, no. 3, pp. 305-315, 2011.
[8] R.V.Rao, V.J.Savsani and D.P.Vakharia, "Teaching–learning-based optimization: an Optimization method for continuous non-linear large scale problems," *Information Sciences,* vol. 183, no. 1, pp. 1-15, 2012.
[9] R.V.Rao and V. Patel, "An elitist teaching–learning-based optimization algorithm for solving complex constrained optimization problems," *International Journal of Industrial Engineering Computations*, vol. 3, no. 4, pp. 535-560, 2012.
[10] K.F. Man, K.S.Tang and S.Kwong, "Genetic algorithms: concepts and application [in engineering design]," *IEEE Trans. Industrial Electronics*, vol. 43, no. 5, pp. 519-534, 1996.
[11] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *IEEE International Conference on Neural Network.*, vol. 4, pp. 1942-1948, 1995.
[12] S. Das, P.N. Suganthan, "Differential Evolution: A survey of the State-of-the-Art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4-31, 2010.
[13] F.S. Abu-Mouti and M.E. EI-Hawary, "Overview of Artificial Bee Colony (ABC) algorithm and its applications*," IEEE International System Conference*, pp. 1-6, 2012.

[14] I.Y. Soon and S.N. Koh, "Speech enhancement using a 2-D Fourier Transform," *IEEE Trans. on Speech and audio process.*, vol. 11, no. 6, pp. 717-724, 2003.

[15] J.C. Goswami, and A.K. Chan, *Fundamentals of wavelets theory, algorithms & applications*, John wiley & sons, Inc., Publication, 2011.

[16] D.L. Donoho and I.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage,' *American Statistical Assoc.*, vol. 90, no. 432, pp. 1200-1224, 1995.

[17] J.H. Chang, S. Gazor, N.S. Kim and S.K. Mitra, "Multiple statistical models for soft decision in noisy speech enhancement," *Pattern Recognition.*, vol. 40, no. 3, pp. 1123—1134, 2007.

[18] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[19] Y. Hu and P.C. Loizou, "Speech enhancement based on wavelet thresholding the multi-taper spectrum," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 1, pp. 59-67, 2004.

[20] S. Tabibian, A. Akbari and B. Nasersharif, "Speech enhancement using a wavelet thresholding method based on symmetric Kullback-Leibler divergence," *Signal process.*, vol. 106, pp. 184-197, 2015.

[21] D. Heric, and B. Potocnik, "Image enhancement by using directional wavelet transformation*," Journal of Computing and Information Technology*, vol. 14, no. 4, pp. 299–305, 2006.

[22] Y. Ghanbari, M.R. Karami-Mollaei, "A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets," *Speech Commun.*, vol. 48, No. 8 pp. 927–940, 2006.

[23] J. Ramírez, J.C. Segura, C. Benítez, A. Torre, A.J. Rubio, "A new Kullback–Leibler VAD for speech recognition in noise," *IEEE Signal Process. Lett.*, Vol. 11, No. 2), pp. 266–269, 2004.

[24] Y. Hu and P.C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transactions on Audio, Speech and language processing*, vol. 16, no. 1, pp. 229- 238, 2008.

AUTHOR PROFILE

D.Yugandhar completed his M. Tech. from JNTU, Hyderabad in 2007 and registered for Ph. D. in Berhampur University. He has 11 years of experience in the industry and 13 years of teaching experience in various engineering colleges. Presently he is working as Associate Professor in the Dept. of ECE of Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India. His interesting areas of research are Signal processing and Image Processing.

Dr. S. K. Nayak completed his M. Tech. from IISC, Bangalore in 1990 and Ph. D. degree from Berhampur University in 1996. He is now working as Professor in the Dept. of Electronic Science, Berhampur University, Berhampur. His interesting areas of research are embedded systems, Image Processing and Signal processing.