International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 1, January - February 2017 ISSN 2278-6856

.

Novel Speech Signal Enhancement Techniques for Tamil Speech Recognition using RLS Adaptive Filtering and Dual Tree Complex Wavelet Transform

Vimala.C [1], Radha.V[2]

[2]Professor and Head, Department of Computer Science,
Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore – 641043, Tamil Nadu, India.

Abstract

A good speech signal enhancement technique must improve both quality and intelligibility of the enhanced signals for all types of environment conditions. However, the speech signal enhancement technique can reduce noise, but introduce its own distortion to the enhanced signals. Hence, it may or may not improve the quality and the intelligibility of the enhanced speech signals. The main objective of this paper is to propose suitable speech signal enhancement techniques that can improve both quality and intelligibility of the enhanced signals. In this research work, the combinational speech signal enhancement techniques are proposed using Dual Tree Complex Wavelet (DTCW) Transform and Recursive Least Squares (RLS) adaptive filtering. Three types of techniques are introduced by using the combination of DTCW and RLS adaptive filtering. The performances of the developed techniques are evaluated based on both subjective and objective speech quality measures. The experimental results prove that the proposed methods have provided better results in speech noise cancellation. Excellent results are achieved in improving the quality and intelligibility of the enhanced speech signal.

Keywords: Speech signal enhancement, Dual Tree Complex Wavelet (DTCW) Transform, RLS adaptive algorithm, Ideal Binary Mask (IBM), Phase Spectrum Compensation (PSC).

1. INTRODUCTION

In real time environment, the speech signals are corrupted by several forms of noise such as competing speakers, background noise, channel distortion and room reverberation etc. The presence of background noise in speech significantly reduces its quality and intelligibility of the signal. Therefore, enhancing the noisy speech signal is necessary for improving the perceptual quality. Speech signal enhancement is applied in many applications like telecommunications, speech and speaker recognition etc. [1]. Particularly, there is a huge need for speech signal enhancement in speech recognition system. This is because, speech recognition application may be developed in one environment and it can be operated in some other environment. In such cases, the mismatch between the training and testing conditions will be increased and the recognition performance will be decreased. Several techniques have been proposed for speech signal enhancement such as spectral subtraction, adaptive filtering, Kalman filtering, wavelet filtering and Ideal Binary Mask (IBM) etc.

The main objective of this paper is to implement efficient speech signal enhancement techniques which are suitable for different noisy conditions. The potent metrics of the DTCW transform has been considered and it is combined with RLS adaptive filtering, IBM and Power Spectrum Compensation (PSC) methods. Four types of noise (White, Babble, Mall and Car) and five types of SNR dB levels (-10dB, -5dB, 0dB, 5dB and 10dB) are involved in the proposed work and their performances are evaluated both subjectively and objectively. In this paper, apart from noise reduction, the improvement in the quality and intelligibility of the enhanced signal has been focused more. The proposed techniques have improved both intelligibility and the quality of the enhanced signal.

The paper is organized as follows. Section 2 discusses about the related works on wavelet transform. Section 3 explains the RLS adaptive filtering technique and section 4 discusses about the proposed technique using DTCW transform. In section 5, the experimental results are presented and the performance metrics used for the proposed work is explained in section 6. The overall discussions are summarized in section 7 and the conclusion and future work is given in section 8.

2. RELATED WORKS

Reshad Hosseini and Mansur Vafadust, (2008) have developed an almost perfect re-construction filter bank for non-redundant, approximately shift-invariant, complex wavelet transforms [2]. The proposed novel filter bank with Hilbert pairs wavelet filters does not have serious distributed bumps on the wrong side of power spectrum. The redundancy of an original signal is significantly

.
reduced and the properties of proposed filter bank can be exploited in different signal processing applications.

Slavy, G. Mihov et al. (2009) performed a de-noising of noisy speech signals by using Wavelet Transform [3]. The use of wavelet transform in de-noising and the speech signals contaminated with common noises is investigated. The authors state that, the wavelet-based de-noising with either hard or soft thresholding was found to be the most effective technique for many practical problems. The experimental results with large database of reference speech signals contaminated with various noises in several Signal-to-Noise Ratio (SNR) dB conditions are presented. The authors also insist that, the power spectrum estimation using a wavelet based de-noising may be applied as an important approach for better speech signal enhancement. The research work will be extended to be applied for the practical research on speech signal enhancement for hearing-aid devices.

Rajeev Aggarwal et al. (2011) have implemented a Discrete Wavelet Transform (DWT) based algorithm using both hard and soft thresholding for denoising [4]. Experimental analyzes is performed for noisy speech signals corrupted by babble noise at 0dB, 5dB, 10dB and 15dB SNR levels. Output SNR and MSE are calculated and compared using both types of thresholding methods. Experiments show that soft thresholding method was found to be better than a hard thresholding method for all the input SNR dB levels involved in the work. The hard thresholding method has extended a 21.79 dB improvement while soft thresholding has achieved a maximum of 35.16 dB improvement in output SNR.

Jai Shankar, B and Duraiswamy, K (2012), have proposed a de-noising technique based on wavelet transformation [5]. The noise cancellation method is improved by a process of grouping closer blocks. All the significant information resides in each set of blocks are utilized and the vital features are extracted for further process. All the blocks are filtered and restored in their original positions, where the overlapping is applied for grouped blocks. The experimental results have proved that the developed technique was found to be better in terms of both SNR and signal quality. Moreover, the technique can be easily modified and used for various other audio signal processing applications.

D. Yugandhar, S.K. Nayak (2016) have proposed 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 DTCWT [6]. The performance of the proposed methods is better in terms of PESQ and PSNR.
Pengfei Sun and Jun Qin (2017) have proposed a two-stage Dual Tree Complex Wavelet Packet Transform (DTCWPT) based speech enhancement algorithm, in which a Speech Presence Probability (SPP) estimator and a generalized Minimum Mean Squared Error (MMSE) estimator are developed [7]. In their work, to overcome the drawback of signal distortions caused by down sampling of Wavelet Packet Transform (WPT), a two-stage analytic decomposition concatenating Undecimated Wavelet Packet Transform (UWPT) and decimated WPT is employed. The process of RLS adaptive filtering technique is explained in the next section.

3. RLS Adaptive Filtering for Speech Signal Enhancement

RLS adaptive algorithm is a recursive implementation of the Wiener filter, in which the input and output signals are related by the regression model. RLS has the potential to automatically adjust the coefficients of a filter, even though the statistic measures of the input signals are not present [8]. In RLS algorithm, filter tap weight vector is updated by

$$w(n) = \overline{w}^T(n-1) + k(n)\overline{e}_{n-1}(n) \qquad (1)$$

The steps involved in RLS adaptive algorithm is given in the following algorithm and the variables used in the algorithm is illustrated in Table 1.

Algorithm of RLS adaptive filtering

Step 1: Initialize the algorithm by setting

$$\hat{w}(0) = 0,$$

$$P(0) - \delta^{-1} I, and$$

$$\delta = \begin{cases} \text{Small positive constant for high SNR} \\ \text{Large positive constant for low SNR} \end{cases}$$

Step 2: For each instant time, n=1,2,…, compute

$$k(n) = \frac{\lambda^{-1} P(n-1) u(n)}{1 + \lambda^{-1} u^H(n) P(n-1) u(n)}$$

$$y(n) = \hat{w}^H(n-1) u(n)$$

$$e(n) = d(n) - y(n)$$

$$\hat{w}(n) = \hat{w}(n-1) + k(n) e^*(n)$$

$$P(n) = \lambda^{-1} P(n-1) - \lambda^{-1} k(n) u^H(n) P(n-1)$$

When the input data characteristics are changed, the filter adapts to the new environment by generating a new set of coefficients for the new data [9]. Here, λ-1 denotes the reciprocal of the exponential weighting factor. The main advantage of RLS adaptive filtering is, it attempts to reduce the estimated error e(n). Therefore, output from the adaptive filter matches closely the desired signal d(n). The perfect adaptation can be achieved, when e(n) reaches

.
zero. In this work, the resultant enhanced signal y(n) produced by RLS filtering was found to be better in terms of quality and intelligibility.

Table 1: Variables used in RLS Algorithm

Variabl e	Description
N	Current algorithm iteration
u(n)	Buffered input samples at step n
P(n)	Inverse correlation matrix at step n
k(n)	Gain vector at step n
y(n)	Filtered output at step n
e(n)	Estimation error at step n
d(n)	Desired response at step n
Λ	Exponential memory weighting factor

Vimala.C and Radha.V have done a performance evaluation of the three adaptive filtering techniques, namely, Least Mean Squares (LMS), Normalized Least Mean Squares (NLMS) and RLS adaptive filtering techniques. These techniques are evaluated for Noisy Tamil Speech Recognition based on three performance metrics, namely, SNR, SNR Loss and MSE [10]. It is observed from the experiments that, RLS technique provides faster convergence and smaller error, but it increases the complexity when compared with LMS sand NLMS techniques. Based on the significant result achieved by the RLS adaptive filtering, the combinational techniques are proposed by using the DTCW transform based reconstruction methodology. The subsequent sections briefly explain the same in detail.

4. Proposed Technique using RLS Filtering and DTCW Transform

In signal processing, quality represents the naturalness of speech, and the intelligibility represents the understandability of text information present in the signal. However, removing noise and improving the perceptual quality and intelligibility of a speech signal, without altering the signal quality, is a crucial job. This is an important problem in any speech enhancement technique. The main objective of this research work is, to develop efficient speech enhancement techniques which can improve both quality and intelligibility of the enhanced speech signals. To meet this objective, two significant improvements are done with the existing RLS adaptive algorithm.

- Suitable square root correlation matrix and forgetting factor value is identified which can be applied for all type of noises and SNR dB levels, and
- The reconstruction methodology is applied to the resultant RLS signal using DTCW transform, to produce the perfect enhanced signal as like the original input signal.

Various initial square root correlation matrix values and RLS forgetting factor values have been evaluated. It is observed from the experiments that different forgetting factor value and square root correlation matrix need to be assigned for positive and negative SNR dB values. In such cases, these values to be assigned and their performances should be evaluated based on trial and error method. It is time consuming and not suitable for applying different noisy conditions. Therefore, the above mentioned two parameters are fine-tuned, to provide optimal values which are more suitable for both positive and negative SNR dB levels. The experiments are carried out under Matlab environment and the desired values for the above parameters are discovered. It is confirmed from the experimental outcome that, better results are obtained when the initial square root correlation matrix values are assigned as 2*eye(10) and RLS forgetting factor value is set to 1. After fine tuning these two parameters, the reconstruction methodology is implemented using DTCW transform. The advantages of using DTCW transform and the steps involved in the proposed technique are explained in the next section.

4.1 Advantages of using DTCW

The standard form of Discrete Wavelet Transform (DWT) is shift-variant, which is undesirable and does not provide perfect speech signal enhancement. To perform speech enhancement for severe noisy conditions, the simple DWT cannot produce the expected outcome. In such cases, there is a need for other alternative technique which can perform well under different noisy conditions. To overcome the shift-variance problem, noise reductions based on shift-invariant wavelet transforms, have been introduced by using DTCW transform [9]. It consists of two specifically designed DWTs, which are applied in parallel to the same input data and it is shown in Figure 1 [10]. DTCW transform is more attractive than the single DWT in terms of computational complexity, because it is equivalent to the two standard DWTs [13].

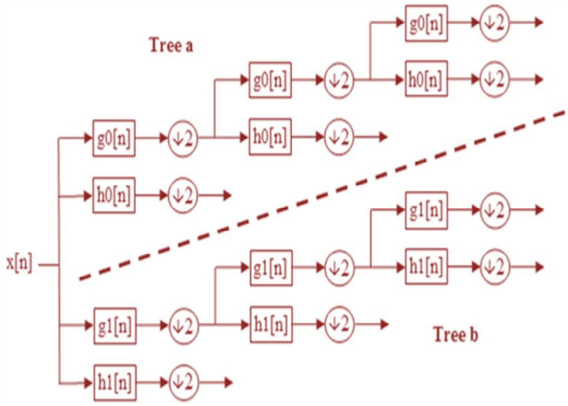

Figure 1 Structure of DTCW Transform

.
The sub-band signals of these two DWTs can be interpreted as the real and imaginary parts of a complex wavelet transform, which is nearly shift-invariant. In DTCW transform, the real and imaginary coefficients are calculated in *tree a* and *tree b* respectively [14]. In this research work, the enhanced signal produced by RLS adaptive filtering is reconstructed using DTCW transform. Figure 2 shows the proposed speech enhancement technique using RLS-DTCW Transform and its steps are briefly given in the following algorithm.

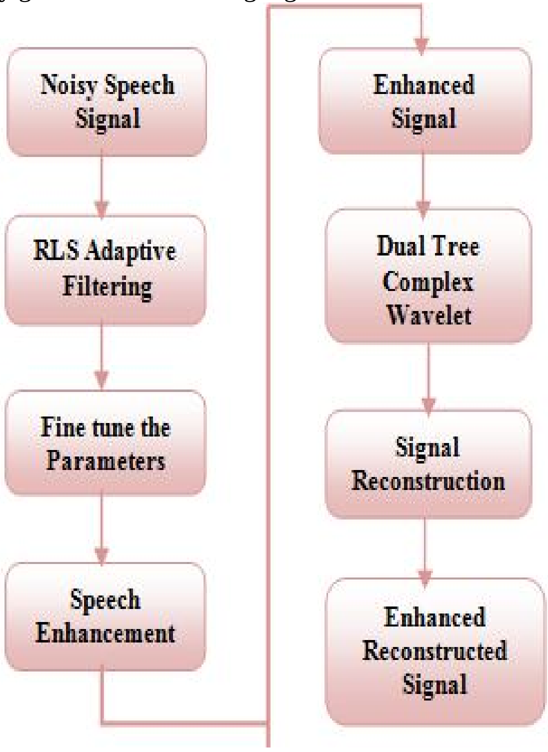

Figure 2 Proposed speech enhancement technique using RLS-DTCW Transform

As given in the algorithm, initially the noisy input signal is passed to the adaptive filtering. Later the parameters are fine tuned and the suitable values are assigned for speech enhancement. Subsequently, the output signal acquired from RLS filtering is further given for DTCW transform to perform reconstruction methodology. By using DTCW transform, additional information about the noisy input signal can be extracted, because it involves both real and imaginary coefficients. Therefore, the perfect reconstruction and better signal enhancement is achieved.

The resultant signal has produced better signal enhancement and it is very much close to the original signal. The experimental results indicate that, the RLS adaptive filtering with DTCW transform has performed better when compared with the existing RLS adaptive filtering. The experimental result achieved by the developed technique is presented in the next section.

Steps involved in Proposed RLS-DTCW Transform

Step 1: Get the noisy speech signal as an input,
Step 2: Initialize RLS filtering,
Step 3: Fine tune the Initial square root correlation matrix inverse,
Step 4: Set RLS forgetting factor value to 1,
Step 5: Perform RLS filtering,
Step 6: Pass the resultant signal as an input to the DTCW transform,
Step 7: Initialize the DTCW transform for performing reconstruction methodology,
Step 8: Calculate the complex transform of a signal using two separate DWT decompositions (tree *a* and tree *b*),
Step 9: Extract the real coefficients using tree *a*,
Step 10: Extract the imaginary coefficients using tree *b*,
Step 11: Approximate shift-invariance, and

5. EXPERIMENTAL RESULTS

The perception of a speech signal is usually measured in terms of its quality and intelligibility. Quality is the subjective measure which reflects on individual preferences of listeners. Intelligibility is an objective measure which predicts the percentage of words that can be correctly identified by the listeners. ***It is noticed from the experimental results that, the resultant signals of RLS was found to be better in terms of both quality and intelligibility***. Since, the perfect reconstruction is done, the enhanced signals are found to be more clear and natural.

The experiments are done with 10 Tamil Spoken Digits uttered 10 times which are corrupted by four types of noise (White, Babble, Mall and Car noise) and five types of Signal-to-Noise Ratio (SNR) dB levels varying from -10dB to 10dB. The total dataset size is 2000 (10*10*4*5). Since the noisy dataset is not available for Tamil language, it is created artificially by adding noise from ***NOIZEUS*** database. In noisy environment, when the SNR dB level is less than 20 dB, the speech recognition will become a difficult problem. In this research work, even more critical situations are handled. Figure 3 shows the waveform representation of the enhanced signals corrupted by babble noise using proposed RLS-DTCW transform and the corresponding spectrograms are presented in Figure 4.

.

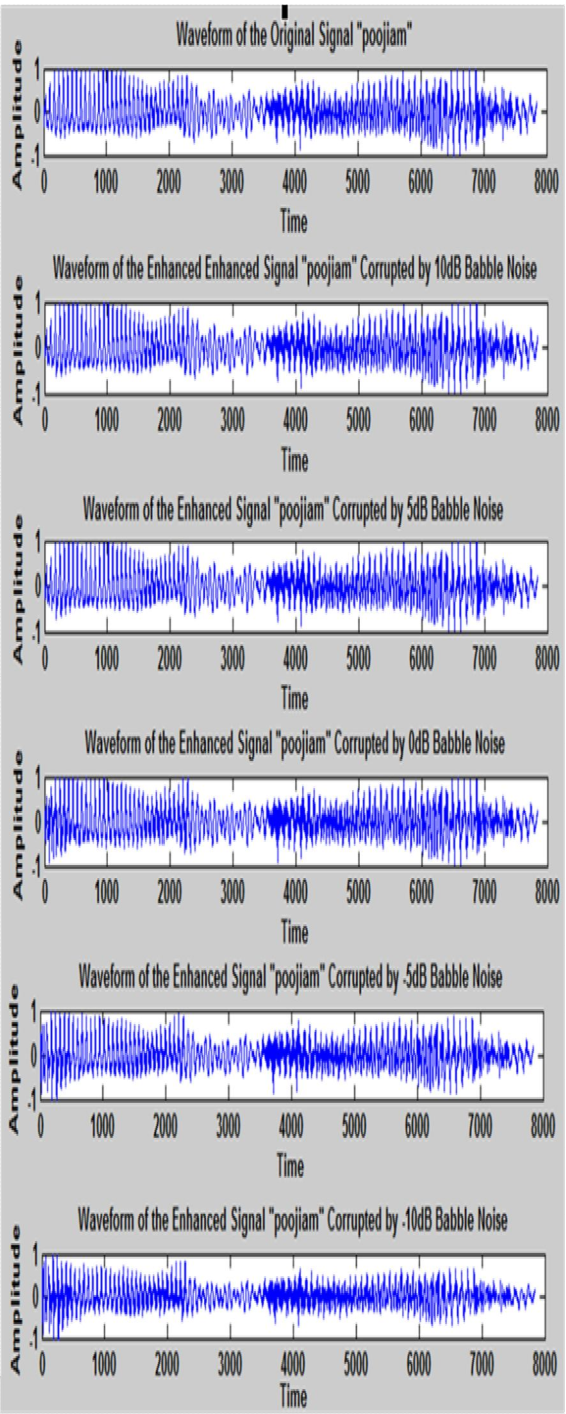

Figure 3 Waveform Representation of the Enhanced Signal using RLS-DTCW Transform Technique

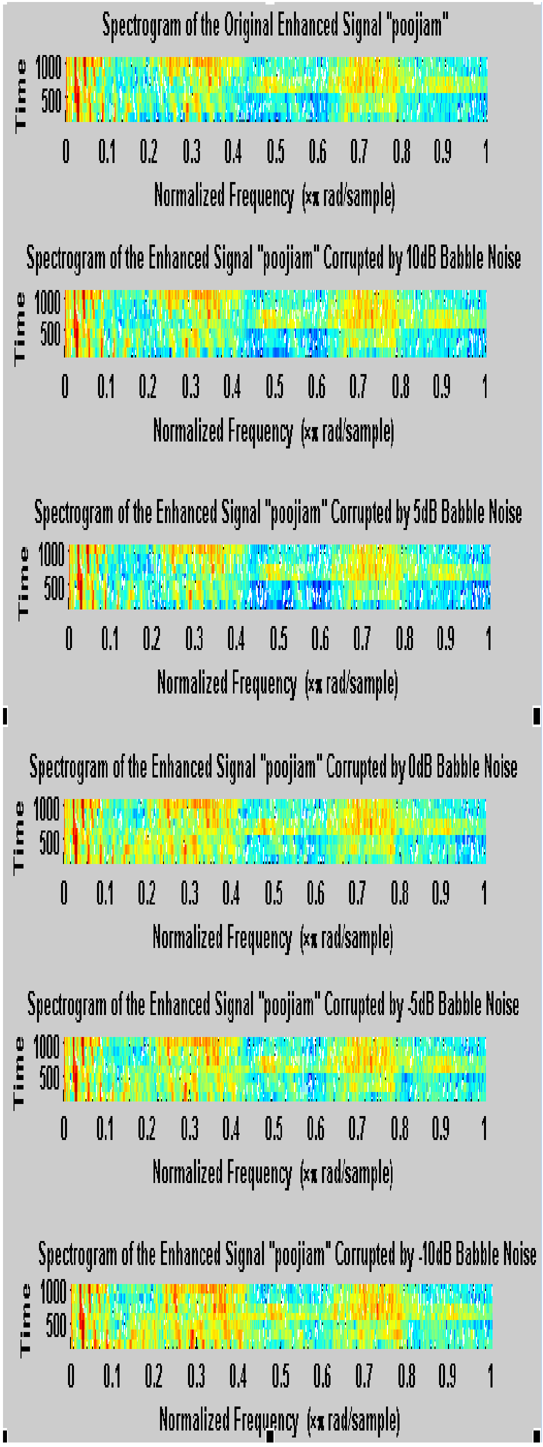

Figure 4 Spectrogram Representation of the Enhanced Signal using RLS-DTCW Transform Technique

As discussed in the previous section, next to RLS, the IBM and PSC methods have produced better results. Based on the improvements achieved with the combination of RLS and DTCW transform, the IBM and PSC methods are also considered for improving their performance. Therefore, these two methods are also improved by combining with the RLS-DTCW transform. To accomplish this task, the filtered signal using IBM and PSC methods are passed to the RLS and DTCW transform. The performance improvement of the IBM and PSC methods are assessed in two ways:

- By applying the DTCW reconstruction

.
methodology alone (IBM-DTCW), (PSC-DTCW), and

- By applying both DTCW and RLS adaptive Filtering technique (IBM-RLS-DTCW), (PSC-RLS-DTCW).

The overall approach of the proposed work is given in Figure 5. It is observed from the experimental outcomes that, there was a reasonable improvement achieved by using DTCW transform for both IBM and PSC methods. However, there was a significant performance improvement obtained while using both RLS and DTCW transform rather than using DTCW transform alone. Particularly, the PESQ, MOS values has been increased and the WSS and MSE values have been reduced extensively. Performance evaluation of RLS-DTCW, IBM-RLS-DTCW and PSC-RLS-DTCW techniques based on speech signal quality measures are discussed below.

6. Performance Evaluations based on Speech Signal Quality Measures

The developed speech signal enhancement techniques are evaluated by using both subjective and objective speech quality measures. In this work, six types of objective quality measures and one subjective quality measure is considered.

6.1 Objective Speech Quality Measures

Objective metrics are evaluated, based on the mathematical measures. The objective quality measures used in this work are as follows:

- Perceptual Evaluation of Speech Quality (PESQ),
- Log Likelihood Ratio (LLR),
- Weighted Spectral Slope (WSS)
- Segmental SNR (SegSNR),
- Output SNR, and
- Mean Squared Error (MSE).

6.1.1 Perceptual Evaluation of Speech Quality (PESQ)

PESQ is the most sophisticated and accurate speech signal quality measure. It is recommended by ITU-T for speech quality assessment of 3.2 kHz narrow-band handset for telephony and speech codec applications. To compute PESQ, the difference between the original and the enhanced signals are computed and averaged over time. The prediction of subjective quality rating between 1.0 and 4.5 will be produced. The higher value represents the better quality of the enhanced signal.

6.1.2 Log Likelihood Ratio (LLR)

LLR is computed with respect to the difference between the target and the reference signals in frame-by-frame analysis. LLR computation requires the corresponding original speech signal as the reference signal for comparing with the target signal and it is given by

$$LLR\ (\bar{a}_p, \bar{a}_c) = \log\left(\frac{\bar{a}_p R_c \bar{a}_p^T}{\bar{a}_c R_c \bar{a}_c^T}\right) \quad (2)$$

where, $\bar{a}_c$ is the LPC vector of the original speech frame, $\bar{a}_p$ is the LPC vector of the enhanced speech frame, R_c is the autocorrelation matrix of the original speech signal.

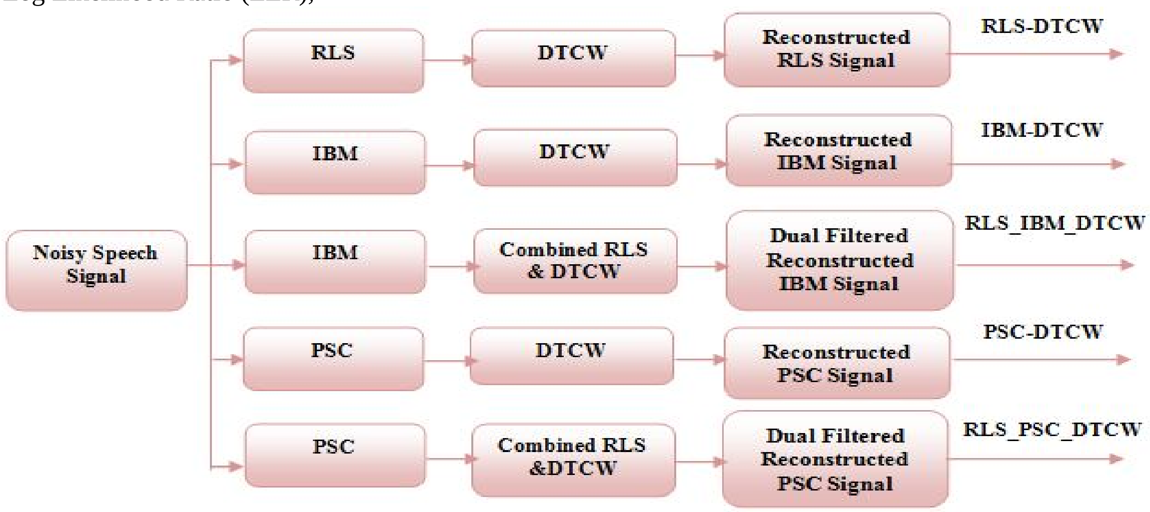

Figure 5 Overall Approach of the Proposed Work

6.1.3 Weighted Spectral Slope (WSS)

WSS is measured based on the comparison of the smoothed spectra from the clean and distorted speech samples. The spectral slope is obtained as the difference between the adjacent spectral magnitudes in decibels. WSS computation is given by

.

$$WSS = \frac{1}{M}\sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W_{WSS}(j,m)(S_c(j,m)-S_p(j,m))^2}{\sum_{j=1}^{K} W_{WSS}(j,m)} \quad (3)$$

where, $W_{WSS}(j,m)$ are the weights, K=25, M is the number of data segments, and $S_c(j,m)$ and $S_p(j,m)$ are the spectral slopes for the j^{th} frequency band of the clean and enhanced speech signals, respectively.

6.1.4 Segmental SNR (SegSNR)

SegSNR represents the average measurements of SNR over short good frames. The SegSNR computation is given by

$$SNRseg = \frac{10}{M}\sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n)-\hat{x}(n))^{2'}} \quad (4)$$

where, $x(n)$ is the input signal, $\hat{x}(n)$ is the processed enhanced signal, N is the frame length and M is the number of frames in the signal.

6.1.5 Output SNR

SNR is defined as the power ratio between the clean signal and the background noise. The SNR can be computed for both input and output signals. SNR is defined by

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \quad (5)$$

where, P_s and P_n represents the average power of speech signal and noisy signal respectively. An output SNR represents the relationship between the strength of the original and the degraded speech signal expressed in decibels and it is computed after applying the speech signal enhancement techniques. Ideally, the greater SNR indicates that the speech is stronger than the noise. An efficient technique should improve the value of the output SNR for the enhanced signal.

6.1.6 Mean Squared Error (MSE)

The MSE measure is defined by

$$MSE = \frac{1}{MN}\sum_{l-0}^{M-1}\sum_{k-0}^{N-1}\left[X_k^2(l) - \hat{X}_k^2(l)\right]^2 \quad (6)$$

where, x_k^2 is the short time MSS of the clean signal, $\hat{x}_k^2$ is the estimated MSS, N is the total number of frequency bins. The small values of MSE show the better estimate of the true MSS.

6.2 Subjective Speech Quality Measure

Subjective quality evaluations are performed by involving a group of listeners to measure the quality of the enhanced speech. The process of performing MOS is described below.

Mean Opinion Score (MOS)

MOS predicts the overall quality of an enhanced signal, based on human listening test. In this work, instead of using a regular MOS, the composite objective measures introduced by Yang Lu and Philipos, C. Loizou (2008) is implemented [15]. The authors have derived new accurate measures from the basic objective measures, which are obtained by using multiple linear regression analysis and nonlinear techniques. It is time consuming and cost effective but provides more accurate estimate of the speech quality, so it is considered in this research work. Separate quality ratings for both signal and background distortions are used and it is shown in Table.2. To calculate the MOS, the listeners have to rate the particular enhanced speech signal, based on the overall quality. The overall quality is measured by calculating the mean value of signal and background distortions (1= bad, 2=poor, 3= fair, 4= good and 5= excellent). In this research work, the MOS is calculated by performing listening test from 20 different speakers (10 males and 10 females). The listeners were asked to rate the speech sample under one of the five signal quality categories.

Table 2: Signal and Background Distortion Scale Rating

Rating	Signal Distortion Scale	Background Distortion Scale
5	Purely Natural, No degradation	Not perceptible
4	Fairly Natural, Slight degradation	Somewhat noticeable
3	Somewhat natural, Somewhat degraded	Noticeable but not intrusive
2	Fairly unnatural, Fairly degraded	Fairly Noticeable, Somewhat Intrusive
1	Quite unnatural, Highly degraded	Quite Noticeable, Highly intrusive

The experimental results obtained by the adopted techniques and the performance evaluations are presented in the next section. Tables 3,4,5 and 6 illustrate the performance evaluation of the proposed speech signal enhancement techniques for white, babble, mall and car noise respectively (for SNR dB types -10 dB,-5 dB, 0dB, 5 dB and 10 dB).

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 1, January - February 2017 ISSN 2278-6856

.

Table 3: Performance Evaluations of the Proposed Speech Signal Enhancement Techniques for White Noise

SNR dB Types	Metrics	RLS	RLS-DTCW	IBM	IBM-DTCW	RLS-IBM-DTCW	PSC	PSC-DTCW	RLS-PSC-DTCW
-10 dB	**PESQ**	3.48	**4.03**	2.32	2.38	**3.94**	1.17	1.19	**3.28**
	LLR	0.86	**0.39**	0.42	3.09	**0.40**	2.41	1.99	**1.24**
	WSS	9.41	**1.20**	63.88	57.53	**0.48**	143.84	143.63	**4.68**
	SegSNR	6.54	**10.85**	-1.85	-1.97	**1.78**	-0.11	0.52	**-0.60**
	Output SNR	4.73	**9.38**	-1.98	-2.19	**-0.55**	0.16	0.77	**1.80**
	MSE	0.14	**0.08**	0.31	0.30	**0.25**	0.24	0.21	**0.19**
	MOS	3.89	**4.63**	2.92	1.53	**4.55**	-0.43	0.53	**3.57**
-5 dB	**PESQ**	3.71	**4.31**	2.50	2.47	**4.25**	1.56	1.58	**3.17**
	LLR	0.52	**0.38**	0.46	3.07	**0.28**	1.92	1.61	**1.26**
	WSS	4.92	**0.55**	46.74	50.08	**0.22**	102.55	101.79	**5.34**
	SegSNR	11.31	**17.65**	-1.82	-1.99	**2.63**	1.66	2.15	**-0.56**
	Output SNR	8.83	**12.92**	-2.03	-2.23	**-0.52**	1.89	2.38	**2.62**
	MSE	0.09	**0.05**	0.31	0.30	**0.25**	0.20	0.18	**0.17**
	MOS	4.28	**4.87**	3.17	1.66	**4.87**	0.74	1.33	**3.46**
0 dB	**PESQ**	3.86	**4.40**	2.62	2.56	**4.34**	1.98	2.00	**3.06**
	LLR	0.29	**0.36**	0.45	2.99	**0.14**	1.51	1.24	**1.26**
	WSS	1.85	**0.17**	42.72	47.46	**0.15**	66.22	66.21	**6.22**
	SegSNR	16.02	**22.42**	-1.86	-1.92	**3.29**	3.25	3.58	**-0.52**
	Output SNR	13.54	**14.71**	-2.08	-2.17	**-0.48**	3.42	3.78	**3.23**
	MSE	0.05	**0.04**	0.31	0.30	**0.25**	0.16	0.15	**0.16**
	MOS	4.54	**4.95**	3.30	1.79	**5.02**	1.82	2.10	**3.37**
5 dB	**PESQ**	4.00	**4.45**	1.55	2.65	**4.31**	2.49	2.55	**2.98**
	LLR	0.12	**0.35**	0.77	2.82	**0.10**	1.19	0.92	**1.23**
	WSS	0.64	**0.06**	111.47	42.11	**0.22**	42.94	42.41	**6.37**
	SegSNR	20.03	**25.51**	-0.92	-1.95	**3.64**	4.17	4.41	**-0.48**
	Output SNR	17.84	**15.62**	-0.93	-2.20	**-0.43**	4.25	4.56	**3.53**
	MSE	0.03	**0.04**	0.27	0.30	**0.25**	0.15	0.14	**0.16**
	MOS	4.75	**4.99**	1.64	1.98	**5.01**	2.79	2.88	**3.32**
10 dB	**PESQ**	4.28	**4.49**	0.71	2.68	**3.01**	2.98	4.27	**2.91**
	LLR	0.04	**0.35**	1.87	2.71	**0.68**	0.93	0.10	**1.27**
	WSS	0.27	**0.03**	226.67	39.58	**28.55**	29.86	0.24	**7.08**
	SegSNR	23.07	**27.51**	-0.06	-1.88	**4.82**	4.61	3.80	**-0.40**
	Output SNR	21.19	**16.13**	-0.06	-2.14	**4.92**	4.63	-0.36	**3.66**
	MSE	0.02	**0.04**	0.24	0.30	**0.13**	0.14	0.24	**0.15**
	MOS	4.82	**4.93**	-0.86	2.08	**3.47**	3.60	4.98	**3.24**

.

Table 4: Performance Evaluations of the Proposed Speech Signal Enhancement Techniques for Babble Noise

SNR dB Types	Metrics	RLS	RLS-DTCW	IBM	IBM-DTCW	RLS-IBM-DTCW	PSC	PSC-DTCW	RLS-PSC-DTCW
-10 dB	PESQ	3.50	**4.15**	2.14	2.40	**4.13**	0.71	0.68	**3.34**
	LLR	0.29	**0.40**	0.37	2.95	**0.39**	1.81	1.71	**1.30**
	WSS	12.09	**0.63**	71.04	64.33	**0.85**	168.65	172.15	**8.66**
	SegSNR	8.81	**14.63**	-1.66	-1.58	**2.64**	-0.99	-0.96	**-0.25**
	Output SNR	5.35	**12.30**	-1.78	-1.77	**-0.23**	-0.99	-0.98	**2.65**
	MSE	0.13	**0.06**	0.30	0.29	**0.24**	0.27	0.26	**0.17**
	MOS	4.18	**4.73**	2.76	1.57	**4.71**	-0.52	0.06	**3.56**
-5 dB	PESQ	3.71	**4.27**	2.39	2.39	**4.23**	0.86	0.82	**3.28**
	LLR	0.10	**0.38**	0.40	2.85	**0.23**	1.47	1.40	**1.24**
	WSS	5.89	**0.17**	61.80	62.29	**0.28**	143.74	145.92	**7.85**
	SegSNR	13.01	**24.94**	-1.52	-1.67	**3.01**	0.73	0.77	**-0.26**
	Output SNR	9.10	**15.81**	-1.69	-1.89	**-0.24**	0.88	0.91	**3.02**
	MSE	0.09	**0.04**	0.30	0.29	**0.24**	0.22	0.21	**0.17**
	MOS	4.49	**4.84**	3.01	1.63	**4.89**	0.15	0.52	**3.54**
0 dB	PESQ	3.96	**4.35**	2.51	2.50	**4.29**	1.22	1.19	**3.23**
	LLR	0.05	**0.36**	0.40	2.80	**0.10**	1.19	1.10	**1.25**
	WSS	2.85	**0.07**	56.75	57.78	**0.17**	105.80	106.41	**7.34**
	SegSNR	17.19	**28.16**	-1.52	-1.55	**3.04**	2.15	2.20	**-0.27**
	Output SNR	13.51	**16.42**	-1.70	-1.77	**-0.25**	2.44	2.50	**3.04**
	MSE	0.05	**0.04**	0.30	0.29	**0.24**	0.18	0.18	**0.16**
	MOS	4.74	**4.91**	3.14	1.77	**4.99**	1.07	1.24	**3.50**
5 dB	PESQ	4.18	**4.42**	1.57	2.55	**4.31**	1.83	1.80	**3.08**
	LLR	0.03	**0.36**	0.81	2.73	**0.07**	0.92	0.84	**1.25**
	WSS	1.22	**0.03**	123.44	53.88	**0.13**	67.26	67.99	**7.78**
	SegSNR	21.14	**29.31**	-1.05	-1.60	**3.36**	3.42	3.51	**-0.31**
	Output SNR	17.82	**16.60**	-1.15	-1.81	**-0.28**	3.65	3.79	**3.32**
	MSE	0.03	**0.03**	0.28	0.29	**0.24**	0.16	0.15	**0.16**
	MOS	4.93	**4.97**	1.52	1.87	**5.02**	2.21	2.13	**3.38**
10 dB	PESQ	4.33	**4.48**	0.81	2.63	**4.30**	2.34	2.31	**3.04**
	LLR	0.02	**0.35**	1.58	2.67	**0.08**	0.78	0.69	**1.24**
	WSS	0.44	**0.02**	239.94	45.49	**0.14**	43.56	42.76	**7.33**
	SegSNR	24.07	**30.21**	-0.05	-1.79	**3.67**	4.29	4.43	**-0.31**
	Output SNR	21.23	**16.67**	-0.03	-2.02	**-0.29**	4.40	4.61	**3.58**
	MSE	0.02	**0.03**	0.24	0.29	**0.24**	0.15	0.14	**0.15**
	MOS	4.96	**4.99**	-0.66	2.03	**5.01**	3.02	2.80	**3.36**

.

Table 5: Performance Evaluations of the Proposed Speech Signal Enhancement Techniques for Mall Noise

SNR dB Types	Metrics	RLS	RLS-DTCW	IBM	IBM-DTCW	RLS-IBM-DTCW	PSC	PSC-DTCW	RLS-PSC-DTCW
-10 dB	PESQ	3.26	**4.29**	2.26	2.40	**4.27**	0.89	0.88	**3.29**
	LLR	0.37	**0.60**	0.36	2.91	**0.59**	2.13	1.81	**1.28**
	WSS	14.23	**0.77**	66.56	57.64	**1.18**	183.17	183.70	**7.47**
	SegSNR	4.47	**10.41**	-2.04	-1.78	**2.26**	-2.09	-1.65	**-0.20**
	Output SNR	3.03	**8.10**	-2.15	-2.04	**-0.17**	-2.15	-1.73	**2.21**
	MSE	0.17	**0.09**	0.31	0.30	**0.24**	0.31	0.29	**0.18**
	MOS	3.93	**4.73**	2.90	1.63	**4.72**	-0.78	0.09	**3.53**
-5 dB	PESQ	3.59	**4.21**	2.61	2.51	**4.07**	1.13	1.12	**3.34**
	LLR	0.22	**0.48**	0.40	2.92	**0.48**	1.73	1.38	**1.37**
	WSS	6.62	**0.37**	51.74	53.40	**1.40**	124.22	125.95	**8.08**
	SegSNR	11.34	**18.63**	-1.59	-1.87	**2.52**	1.29	1.44	**-0.17**
	Output SNR	7.22	**14.09**	-1.76	-2.12	**-0.19**	1.27	1.47	**2.49**
	MSE	0.11	**0.05**	0.30	0.30	**0.24**	0.21	0.20	**0.18**
	MOS	4.33	**4.73**	3.27	1.75	**4.62**	0.33	0.91	**3.53**
0 dB	PESQ	4.18	**4.44**	2.63	2.61	**4.28**	1.28	1.26	**3.20**
	LLR	0.09	**0.40**	0.39	2.83	**0.16**	1.43	1.20	**1.30**
	WSS	2.31	**0.04**	43.23	42.87	**0.16**	104.77	106.06	**7.10**
	SegSNR	17.89	**25.84**	-1.79	-1.89	**3.51**	2.29	2.49	**-0.27**
	Output SNR	15.30	**16.11**	-2.03	-2.18	**-0.26**	2.42	2.64	**3.36**
	MSE	0.04	**0.04**	0.31	0.30	**0.24**	0.18	0.17	**0.16**
	MOS	4.89	**4.96**	3.36	1.95	**4.96**	0.91	1.25	**3.46**
5 dB	PESQ	4.14	**4.38**	1.55	2.66	**4.24**	2.07	2.09	**3.26**
	LLR	0.03	**0.38**	0.85	2.78	**0.12**	1.01	0.79	**1.29**
	WSS	1.51	**0.04**	125.52	43.22	**0.46**	56.96	58.21	**7.67**
	SegSNR	20.21	**27.81**	-1.01	-1.83	**3.14**	3.67	3.78	**-0.22**
	Output SNR	16.14	**16.58**	-1.08	-2.09	**-0.22**	3.69	3.86	**3.08**
	MSE	0.04	**0.03**	0.28	0.30	**0.24**	0.16	0.15	**0.16**
	MOS	4.90	**4.92**	1.45	2.01	**4.94**	2.45	2.47	**3.50**
10 dB	PESQ	4.33	**4.44**	0.70	2.65	**4.29**	2.57	2.53	**3.17**
	LLR	0.02	**0.37**	1.86	2.70	**0.08**	0.79	0.62	**1.30**
	WSS	0.74	**0.02**	231.57	41.37	**0.33**	40.37	39.92	**7.29**
	SegSNR	23.35	**29.41**	-0.12	-1.90	**3.45**	4.28	4.41	**-0.29**
	Output SNR	20.10	**16.79**	-0.07	-2.15	**-0.28**	4.27	4.49	**3.35**
	MSE	0.02	**0.03**	0.25	0.30	**0.24**	0.15	0.14	**0.16**
	MOS	4.97	**4.98**	-0.90	2.06	**5.00**	3.25	3.04	**3.43**

.

Table 6: Performance Evaluations of the Proposed Speech Signal Enhancement Techniques for Car Noise

SNR dB Types	Metrics	RLS	RLS-DTCW	IBM	IBM-DTCW	RLS-IBM-DTCW	PSC	PSC-DTCW	RLS-PSC-DTCW
-10 dB	**PESQ**	3.61	**3.74**	2.17	2.34	**3.68**	1.63	1.63	**2.93**
	LLR	0.05	**0.28**	0.38	2.77	**0.40**	0.72	0.62	**1.37**
	WSS	8.08	**1.12**	68.65	56.13	**1.60**	153.95	153.91	**8.83**
	SegSNR	7.87	**9.26**	-1.59	-1.71	**1.76**	-0.46	-0.57	**-0.35**
	Output SNR	4.05	**8.68**	-1.75	-1.90	**-0.31**	-0.31	-0.43	**1.80**
	MSE	0.15	**0.09**	0.30	0.29	**0.24**	0.25	0.25	**0.19**
	MOS	4.41	**4.45**	2.79	1.67	**4.34**	1.43	1.51	**3.19**
-5 dB	**PESQ**	3.84	**3.88**	2.46	2.49	**3.88**	2.33	2.33	**2.99**
	LLR	0.04	**0.32**	0.35	2.67	**0.22**	0.70	0.56	**1.32**
	WSS	4.26	**0.27**	51.15	50.25	**0.50**	92.25	91.43	**8.24**
	SegSNR	15.02	**17.53**	-1.61	-1.71	**2.92**	1.75	1.72	**-0.35**
	Output SNR	8.52	**14.06**	-1.78	-1.92	**-0.32**	1.92	1.91	**2.93**
	MSE	0.09	**0.05**	0.30	0.29	**0.24**	0.19	0.19	**0.17**
	MOS	4.63	**4.56**	3.19	1.87	**4.60**	2.65	2.54	**3.27**
0 dB	**PESQ**	4.04	**4.16**	2.57	2.58	**3.85**	2.80	2.77	**3.02**
	LLR	0.02	**0.34**	0.34	2.60	**0.14**	0.71	0.55	**1.28**
	WSS	2.07	**0.09**	46.61	46.48	**0.27**	63.56	63.50	**8.32**
	SegSNR	20.28	**24.93**	-1.61	-1.71	**3.55**	3.37	3.43	**-0.40**
	Output SNR	13.29	**16.07**	-1.78	-1.92	**-0.37**	3.55	3.66	**3.50**
	MSE	0.05	**0.04**	0.30	0.29	**0.24**	0.16	0.15	**0.16**
	MOS	4.82	**4.77**	3.33	2.02	**4.62**	3.32	3.10	**3.32**
5 dB	**PESQ**	4.18	**4.34**	1.29	2.61	**3.84**	3.00	2.93	**3.04**
	LLR	0.02	**0.34**	1.41	2.56	**0.11**	0.72	0.55	**1.26**
	WSS	0.91	**0.04**	130.82	43.74	**0.22**	45.69	45.85	**8.23**
	SegSNR	23.29	**28.34**	-1.29	-1.74	**3.77**	4.24	4.36	**-0.43**
	Output SNR	17.61	**16.52**	-1.46	-1.95	**-0.40**	4.36	4.55	**3.66**
	MSE	0.03	**0.03**	0.29	0.29	**0.24**	0.15	0.14	**0.15**
	MOS	4.94	**4.91**	0.75	2.08	**4.63**	3.65	3.35	**3.34**
10 dB	**PESQ**	4.31	**4.42**	0.38	2.62	**3.83**	3.14	3.07	**2.99**
	LLR	0.02	**0.34**	2.13	2.54	**0.11**	0.73	0.56	**1.26**
	WSS	0.42	**0.02**	221.13	41.94	**0.20**	34.31	33.90	**7.52**
	SegSNR	24.99	**29.70**	-0.05	-1.86	**3.86**	4.64	4.81	**-0.42**
	Output SNR	21.01	**16.65**	-0.02	-2.07	**-0.39**	4.70	4.94	**3.73**
	MSE	0.02	**0.03**	0.24	0.30	**0.24**	0.14	0.13	**0.15**
	MOS	5.05	**4.98**	-1.30	2.11	**4.62**	3.87	3.54	**3.31**

.

7. DISCUSSIONS

The experimental results have proved that the RLS-DTCW has ***produced maximum PESQ, MOS, SegSNR, LLR and Output SNR.*** Moreover, ***the proposed technique has reduced WSS, and MSE values when compared to the existing RLS adaptive filtering technique***. The IBM and PSC techniques also ***provided more significant performance improvements in terms of PESQ, MOS, SegSNR and Output SNR values***. Moreover, very good results are achieved in reducing WSS and MSE also. To make clear idea about the improvements achieved by the proposed techniques (shown in Tables 3,4,5 and 6), Table 7 illustrates the difference obtained by the proposed techniques based on both subjective and objective speech quality measures. The average values of improvement obtained for five types of SNR dB levels for four type of noise are presented in the Table. The highest difference achieved by the proposed techniques for each speech quality measure is highlighted.

It is evident from the above Table that, the maximum difference of **3.59 and 2.02** are achieved for **PESQ and LLR** respectively. The **SegSNR** value has been increased up to **11.93**, **output SNR** value has been increased up to **6.95** and the **MOS** value has been improved up to **5.9**. Above all, the **WSS** value has been reduced from **231.57 to 3.33**. Therefore, it is clear from the above results that, the RLS-DTCW, IBM-RLS-DTCW and PSC-RLS-DTCW have obtained very good results. These techniques were found to be better for speech signal enhancement and have improved the quality and intelligibility of the enhanced signals.

8. CONCLUSION

The main objective of this research work is to implement efficient speech signal enhancement techniques which are more suitable for different noisy conditions. In this research work, apart from noise reduction, the improvement in the quality and intelligibility of an enhanced signal is more concentrated. Three types of speech signal enhancement techniques were introduced by using the combination of RLS adaptive filtering and DTCW transform. The proposed techniques are evaluated based on subjective and objective speech quality measures. All the three techniques developed in this paper were found to be good and has achieved better speech signal enhancement. The proposed reconstruction methodology has significant improvement in terms of improving both speech quality and intelligibility of the enhanced signals. These three methods can be used for speech signal enhancement or they can be applied as a front-end processor for noisy speech recognition. The future work is to evaluate the proposed technique for noisy Tamil speech recognition.

References

[1] Lawrence R. Rabiner, B. H. Juang, and Yegnanarayana, “Fundamentals of Speech Recognition”, Pearson Education India, 2008, ISBN: 9788-1775- 85605.

[2] Reshad Hosseini and Mansur Vafadust, “Almost Perfect Reconstruction Filter Bank for Non-redundant, Approximately Shift-Invariant, Complex Wavelet Transforms”, Journal of Wavelet Theory and Applications, Vol. 2(1), pp. 1–14, 2008, ISSN: 0973-6336.

[3] Slavy. G. Mihov, Ratcho. M. Ivanov and Angel. N. Popov, “Denoising Speech Signals by Wavelet Transform”, Annual Journal of Electronics, 2009, ISSN: 1313-1842.

[4] Rajeev Aggarwal, Sanjay Rathore, Jai Karan Singh, Mukesh Tiwari, Vijay Kumar Gupta and Anubhuti Khare, “Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold”, International Journal of Computer Applications, Vol. 20(5), pp. 14-19, 2011, ISSN: 0975 – 8887.

[5] B. Jai Shankar and K. Duraiswamy, “Audio Denoising using Wavelet Transform”, International Journal of Advances in Engineering & Technology (IJAET), Vol. 2(1), pp. 419-425, 2012, ISSN: 2231-1963.

[6] D. Yugandhar, S.K. Nayak, “A Heuristic Speech De-noising with the aid of Dual Tree Complex Wavelet Transform using Teaching-Learning Based Optimization”, Vol. 8(5), pp. 1967-1980, 0975-4024 (Online) 2319-8613 (Print).

[7] Pengfei Sun and Jun Qin, “Speech Enhancement via Two-Stage Dual Tree Complex Wavelet Packet Transform with a Speech Presence Probability Estimator”, The Journal of the Acoustical Society of America, **141**, 808 (2017).

[8] https://www.mathworks.com/help/dsp/ug/rls-adaptive-filters.html

[9] http://www.itu.dk/stud/speciale/segmentering/Matlab6p5/help/toolbox/filterdesign/adaptiv2.html.

[10] C. Vimala and V. Radha, “Optimal Adaptive Filtering Technique for Tamil Speech Enhancement”, International Journal of Computer Applications (IJCA), Vol. 41(17), pp. 23-29, 2012, ISSN: 0975–8887.

[11] Hua Ye, Guang Deng, Stefan, J. Mauger, Adam, A. Hersbach, Pam, W. Dawson and John, M. Heasma, "A Wavelet-Based Noise Reduction Algorithm and

.

Its Clinical Evaluation in Cochlear Implants", PLoS ONE, Vol. 8(9), pp. 1-10, 2013, DOI: 10.1371/ journal.pone.0075662.

[12] N.G. Kingsbury," Complex wavelets for shift invariant analysis and filtering of signals", Journal of Applied and Computational Harmonic Analysis, Vol.10(3), pp. 234–253, 2001, ISSN: 1063-5203.

[13] Ivan, W. Selesnick, Richard, G. Baraniuk, and Nick, G. Kingsbury, "The Dual-Tree Complex Wavelet Transform", IEEE Signal Processing Magazine, Vol. 22(6), pp. 123–151, 2005, doi: 10.1109/MSP. 2005.1550194.

[14] Hosseini and Mansur Vafadust, "Almost Perfect Reconstruction Filter Bank for Non-redundant, Approximately Shift-Invariant, Complex Wavelet Transforms", Journal of Wavelet Theory and Applications, Vol. 2(1), pp. 1–14, 2008, ISSN: 0973-6336.

[15] Yi Hu and Philipos, C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms", Speech Communication, Vol. 49(7), pp. 588–601, 2007.

AUTHOR

Dr. Vimala.C done her Ph.D in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women. She has more than 2 years of teaching experience and 3 years of research experience. She worked as a Project Fellow for the UGC Major Research project.
Her area of specialization includes Speech Recognition, Speech Synthesis and speech signal enhancement. She has 17 publications at National and International level conferences and journals.

Dr.V.Radha, Professor and Head in Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India. She has more than 26 years of teaching experience and 16 years of Research Experience.
Her Area of Specialization includes Image Processing, Optimization Techniques, Speech Signal Processing, Data Mining & Data Warehousing and RDBMS. She has authored more than 100 papers published in refereed International journals and Conferences. She has obtained funding projects from UGC-MRP in the field of speech signal processing. She is a Reviewer of American Journal Operational Research and American Journal of Signal Processing. She is an Editor in Chief of International Journal of Computational Science and Information Technology (IJCSITY). She is the member of many international bodies such as IAENG, AIRCC, IACSIT etc.. Visited countries such as the USA and Singapore.