International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I7P230

Volume 69 Issue 7, 233-243, July, 2021
©2021 Seventh Sense Research Group®

A Novel Speech Enhancement Solution Using Hybrid Wavelet Transformation Least Means Square Method

Jagadish S.Jakati[1], Shridhar S.Kuntoji[2]

[1]*Assistant Professor, Department of Electronics & Communication Engg, S. G. Balekundri Institute of Technology, Belagavi, VTU Research Scholar Karanataka (State), INDIA.*

[2]*Professor & HOD, Department of Electronics & Communication Engg,Basaveshwar Engineering College, Bagalkot, VTU Research Supervisor, Karanataka (State), INDIA.*

[1]jagadishjs30@gmail.com, [2]shridhar.ece@gmail.com

Abstract *- Currently, minimizing the noise in speech or audio signals is a challenging issue in the field of speech recognition, speech enhancement, and other speech communication applications. These applications have fascinated research community due to their diverse use in real-time, online and offline applications. Several approaches have been presented to enhance the quality of speech. Currently, the Wavelet Transformation based approach and Least Means Square based filtering schemes are extensively adopted in various researches. The existing techniques suffer from computational complexity and performance related issues. Thus, we focused on combining these schemes and presented a hybrid approach that uses wavelet packet transform and an adaptive LMS scheme. We present an extensive simulation study and comparative analysis by using the NOIZEUS speech corpus database. The experimental analysis shows a substantialaugmentation in the performance of speech enhancement.*

Keywords — *Speech enhancement, noise filtering, DWT, LMS, NOIZEUS,STFT, MOS, CEP, STOI*

I. INTRODUCTION

In the current living scenario, we face several types of background noises such as traffic noise and high volume speakers. Due to these noises, the original signal that is obtained in real-time environments is contaminated by these noises. Generally, the background noises are considered as broadband and non-stationary signal and the signal-to-noise ratio (SNR) of these signals can be very low. Moreover, these signal causes degradation of original signal which leads to the unintelligibility of speech signal which reduces the execution of speech coding and speech recognition systems. Thus, noise reduction has become an active area of research due to its significant applications in various real-time applications including hearing aids, hands-free communication modules, teleconferencing, etc. [1]. In the current research status, numerous algorithms have been developed to obtain the original signal from the noisy signals. This is obtained by suppressing the noise and minimizing the unwanted signals. Thus, speech enhancement is adopted to improve quality of audio signal along with intelligibility.

Currently, the speech enhancement methods are classified into three different categories as frequency domain methods, time domain methods, and time-frequency domain methods. The frequency domain techniques include wiener filtering, minimum mean square error (MMSE), and spectral subtraction methods. In [2] authors discussed new insights of the Wiener filter and suggested optimization of it to improve the performance of speech enhancement. Similarly, in [3] authors developed a weighted Weiner filter for single-channel noise reduction. Enzner et al. [4] discussed about MMSE based filtering with the help of Bayesian learning scheme. In [5] authors introduced a combined approach that uses spectral subtraction and Wiener filtering for speech enhancement. On the other side, the time domain methods include phase aware scheme, Kalman filtering and subspace, and many more. As in [6] authors developed a combined approach with the help of phase-aware and Kalman filtering schemes. Choi et al. [7] developed phase-aware approach with the help of a deep learning scheme. Similarly, the time-frequency domain schemes include transform based approaches such as Short Time Fourier Transform (STFT) and Wavelet Transform scheme [8, 9].

Similarly, the speech enhancement techniques can be classifed based on their use of microphone single channel and multi-channel speech enhancement techniques [10, 11]. The single channel speech enhancement schemes depend on the single microphone whereas the multichannel speech enhancement techniques utilize multiple microphones. Several studies have been developed based on these schemes which shows that the multichannel speech enhancement techniques show a significant performance improvement in improving the speech quality. Despite, notable advantages of multichannel schemes, the single channel speech enhancement techniques are easy to implement thus these techniques remain the active area of research. Taherian et al.

[12] developed single and multichannel speech enhancement techniques and used them for speech recognition.

Currently, wavelet transform based techniques are widely adopted in various speech processing tasks. Mavaddaty et al. [13] presented a wavelet packet transform strategy based on dictionary learning for speech enhancement. Similarly, Ram et al. [14] presented a DWT and RBF (radial basis function) network to improve the speech quality. However, these techniques suffer from poor reconstruction quality. In this work, we focus on the development of a novel approach for speech enhancement. Moreover, these schemes suffer from computational complexities. Main contribution of this work are as follows:

(a) First of all, we study about existing schemes about speech processing, and speech enhancement techniques.

(b) In next phase, we present a wavelet packet decomposition scheme to deal with the noise in speech signals. Later, the speech signal is processed through the thresholding process.

(c) Finally, the processed signal is passed through the LMS filtering scheme where we update the weights of signals to obtain the filtered signal.

The remainder of the manuscript is organized into 5 sections which are as follows: Section II describes the recent studies in this field of speech enhancement, section III presents the wavelet tree and LMS based approach to increase the quality of contaminated speech signal, section IV presents the experimental analysis and comparative analysis to report the performance improvement and finally, section V presents the concluding remarks and future scope of this research.

II. LITERATURE SURVEY

This section presents a brief literature review about existing techniques in this field of speech enhancement field. Chiea et al. [2] presented a unified approach for noise reduction for speech enhancement. This technique uses Wiener filter and binary mask which is considered as an optimal solution for these problems. In this work, the authors presented a cost function by considering two different design parameters.

Pardede et al. [5] focused on improving the speech quality for a secured communication because incorporating the additional security parameters may lead to the degrading of the signal. This scheme detects non-speech period with the help of a voice activity detector (VAD) which helps to find the noise estimate. Moreover, this system uses aamalgamation of Wiener filter and spectral subtraction approach to estimate the nose efficiently.

Dionelis et al. [6] introduced thekalman filtering scheme in modulation-domain which tracks the phase, spectral log-amplitudes of noise, and speech data. This amplitude and phase data is used to generate the speech data. Moreover, the Kalman filter helps to obtain the inter-framtemporal correlation and update the nonlinear relations of speech and noise.

Choi et al. [7] reported that conventional deep learning approaches focus on the magnitude estimation and reusing phase data for reconstruction where phase estimation becomes a tedious task. Thus authors developed a phase estimation scheme using a deep complex U-net model, later, a polar coordinate-wise making approach s considered to identify the distribution of masks. Finally, a novel loss function is designed to obtain the filtered signal.

Sharma et al. [16] developed a weighted sigmoid noise estimation approach to enhance the performance of speech enhancement. This approach adopts the speech presence probability. This single channel speech enhancement uses STFT transform, frequency spectrum, noise estimation, filtering. The current data is processed and a modified spectrum is generated. This is further processed through the inverse STFT.

Srinivasarao et al. [17] presented a hybrid wiener filtering approach for speech enhancement. This scheme uses a combination of wiener filter and Karhunen–Loéve Transform. The noisy speech is processed through the pre-processing phase which is further processed through the KLT transform, wiener filter, and inverse KLT transforms to obtain the filtered signal.

Yu et al. [18] developed a deep learning based scheme for speech reconstruction and further Kalman filter is applied to denoise the speech signal. In the first phase, two different deep neural networks are trained which are helpful to map the feature to the clean signal magnitudes and line spectrum frequencies. Further, line spectrum frequencies are transformed to apply Kalman filtering which generates the Kalman-filtered signal.

Roy et al. [19] focused on the LPC coefficient and used deep learning based approach. However, the existing deep learning approach generates bias estimates due to whitening filtering. Thus, this degrades the intelligibility of speech. In this work, the authors developed a deep learning approach to avoid the whitening filter. This scheme is called DeepLPC which jointly estimates the LPC power spectra of clean and noisy speech signals. Further, the obtained spectra are fed to inverse Fourier transform to produce the autocorrelation data. Later, this autocorrelation is realized by using Levinson-Durbin recursion to filter the noisy data. Similar to this, Roy et [21] again developed another deep learning approach. Authors reported that previous works consider augmented Kalman filter and temporal convolution network for prediction of LPC of clean and corrupted speech signals. To improve the performance further a new approach is developed called a multi-head attention network (MHANet) for LPC estimation.

Li et al. [20] focused on the development of a combined approach by using DNN and LSTM scheme. This approach adopts the progressive learning process. This deep learning model is called PL-CRNN which considers convolutional neural network and recurrent neural network.

Indra et al. [22] focused on wavelet transform and presented a novel approach for Tamil speech enhancement and speaker recognition system. The newly developed transform is known as modified Tunable-Q Wavelet Transform (TQWT). Compare to conventional continuous and discrete wavelet transform, this approach has the capability to refrain the Q factor which helps to minimize the redundancy. This approach achieves perfect reconstruction and also it is fully discrete in nature which leads to the perfect tuning of the Q-factor to minimize the redundancy. In this process, the speech data is divided into multiple parts and different Q factors are applied to each part.

Garg et al. [23] used data pre-processing and Bionic wavelet transform approach along with empirical mode decomposition (EMD) scheme. The complete process considers a clean signal which is later contaminated by adding additional noises. Later, an EMD scheme is implemented. Later, this signal is processed through the denoising phase which consists of Bionic wavelet transform, butterworth filter, and inverse bionic wavelet transform.

III. PRAPOSED METHOD

This section describe the complete proposed approach for speech signal denoising or speech signal denoising\enhancement by presenting a novel hybrid scheme using wavelet packet transform and Least-Mean-Squared Algorithm (LMS). First of all, we describe the wavelet packet transform followed by the LMS algorithm and later presented a combined approach to improve the signal quality. Generally, the noise is distributed in different time-frequency subspaces. The wavelet based approaches fail to decompose the signal in the higher frequency region where high-frequency noise rejection doesn't generate the optimal solution. Thus, we adopt the wavelet packet transform to deal with this issue. This wavelet of wavelet packet denoising approach is presented in three stages which include packet decomposition, coefficient quantization generated by the decomposition, and reconstruction of the signal. This denoising process follows several factors such as the selection of wavelet packets, determining the number of decomposition layers for wavelet packets, and threshold selection and estimation.

A. Wavelet packet transform

Let us consider that an orthonormal scaling function given as $\phi(t)$ and wavelet function is given as $\psi(t)$. These functions can be expressed as:

$$\begin{aligned}\phi(t) &= \sqrt{2}\sum_{k} h_{0k}\phi(2t-k)\\ \psi(t) &= \sqrt{2}\sum_{k} h_{1k}\phi(2t-k)\end{aligned} \qquad (1)$$

Where h_{0k} and h_{1k} represents the orthogonal filter coefficients. With the help of this, the wavelet packet function for $n = 0,1..$ can be expressed as:

$$\begin{aligned}w_{2n}(t) &= \sqrt{2}\sum_{k\in Z} h_{0k}w_n(2t-k)\\ w_{2n+1}(t) &= \sqrt{2}\sum_{k\in} h_{1k}w_n(2t-k)\end{aligned} \qquad (2)$$

Based on eq. (2),for $n = 0$, the wavelet packets are denoted as$w_0(t) = \phi(t)$, $w_1(t) = \psi(t). \{w_n(t)\}_{n\in Z}$which are determined with the help of $w_0(t) = \phi(t)$. Here, we assume that this scaling function is useful to construct the orthogonal wavelet basis. Thus, scaling function and wavelet function play an important role in wavelet packets. These function have orthogonality property over scale and translation which is expressed as

$$\langle w_n(t-k). w_n(t-l)\rangle = \delta_{kl}, \;\; where\ k,l \in Z$$

$$\langle w_{2n}(t-k). w_{2n+1}(t-l)\rangle = 0, \quad where\ n\ 0,1,2,.. \qquad (3)$$

During this process of wavelet decomposition, the scale scape is constructed using scaling function and wavelet space is constructed using wavelet functions. These spaces can be described as:

$$\begin{aligned}U_j^0 &= V_j\ where\ j \in Z\\ U_j^1 &= W_j\ where\ j \in Z\end{aligned} \qquad (4)$$

$V_j$denotes the scale space and W_j denotes the wavelet space. By applying the convolution of wavelet and scale space as $V - j = V_{j+1} \oplus W_{j+1}$ then we obtain

$$\begin{aligned}U_j^0 &= U_{j+1}^0 \oplus U_{j+1}^1\ where\ j \in Z\\ U_j^n &= U_{j+1}^{2n} \oplus U_{j+1}^{2n+1}\ where\ j \in Z, n \in Z^+\end{aligned} \qquad (5)$$

Where U_j^n represents the closed subspace of integrableand square space which is obtained by the linear grouping of wavelet packets after performing scaling and translation operations. Here, our main aim is to decompose the signal into multiple subspaces as $\{V_j\}_{j\in Z}$ and $\{W_j\}_{j\in Z}$ in the square space. The further decomposition can be expressed as:

$$\begin{aligned}W_j = U_j^1 &= U_{j+1}^2 \oplus U_{j+1}^3\\ U_{j+1}^2 &= U_{j+2}^4 \oplus U_{j+2}^5\\ U_{j+1}^3 &= U_{j+2}^6 \oplus U_{j+2}^7\end{aligned} \qquad (6)$$

This decomposition further can be realized as:

$$\begin{aligned}W_j &= U_{j+1}^2 \oplus U_{j+1}^3\\ W_j &= U_{j+2}^4 \oplus U_{j+2}^5 \oplus U_{j+2}^6 \oplus U_{j+2}^7\\ &\vdots\\ W_j &= U_{j+k}^{2^k+1} \oplus U_{j+k}^{2^k+1} \oplus \ldots \oplus U_{j+k}^{2^{k+1}-1}\end{aligned} \qquad (7)$$

Finally, the wavelet packet coefficients can be expressed as follows:

$$d_k^{j+1,2n} = \sum_l h_{0(2l-k)}\, d_l^{j,n}$$
$$d_k^{j+1,2n+1} = \sum_l h_{0(2l-k)}\, d_l^{j,n} \qquad (8)$$

Where $d_k^{j+1,n} = \sum_k \left[\sum_l h_{0(l-2k)}\, d_k^{j,2n} + h_{l(l-2k)} d_k^{j,2n+1}\right]$

Further, we consider thresholding based scheme approach which is applied on the obtained coefficients. These thresholds functions are categorized soft and hard thresholding. According to thresholding approach, the wavelet coefficients which are greater than the threshold λ then it sets all other to zero. This is defined as:

$$f_h(x) = \begin{cases} x, if\ |x| \geq \lambda \\ 0, \quad otherwise \end{cases} \qquad (9)$$

Similarly, the soft thresholding function helps to shrink the coefficient towards the zero. This function is expressed as:

$$f_h(x) = \begin{cases} x - \lambda, if\ |x| \geq \lambda \\ 0, \quad if\ |x| < \lambda \\ x + \lambda, if\ |x| \leq \lambda \end{cases} \qquad (10)$$

B. LMS Algorithm

In this subsection, we describe the Least Means Square (LMS) algorithm for speech signal quality enhancement. Currently, adaptive filtering based schemes are widely adopted in various speech processing systems. In this field of adaptive filtering, the LMS algorithm plays an important role due to its nature of efficient performance and low cost.Generally, the time step $n$is selected as the initial setting of coefficient parameters for any filtering scheme where later these time steps are updated in each new detection.

Let us consider that the estimated signal is denoted as $d[n]$which is obtained from the $X[n]$ a series of observation. The new updated signal can be given as: $d[n]' = w^T x[n]$. In conventional methods, the value of filtering coefficient w is updated at each time step n. This process of updating the filtering coefficients can be given as:

$$w[n+1] = w[n] = ue[n]x[n] \qquad (11)$$

Where e denotes the error denoted as$e[n] = d[n] - w^T[n]x[n]$.$u$represents the step size. In this approach, selection of step size is a crucial task which is obtained by converging the weight of a solution to approximate the optimal Wiener filtering. The approximated range of step size can be estimated as:

$$0 < u < \left[\frac{2}{\left(P * R_X(0)\right)}\right] \qquad (12)$$

Where P represents the filtering order, and $R_X(0)$ denotes the autocorrelation sequence at zero signal input. Below given figure depicts the process of LMS adaptive filtering which contains input signal $x(n)$, $d(n)$ represents the expected response, $y(n)$ represents the output signal, $v(n)$ is the interference noise and $e(n)$ is the error signal which is the difference between $d(n)$ and $y(n)$.

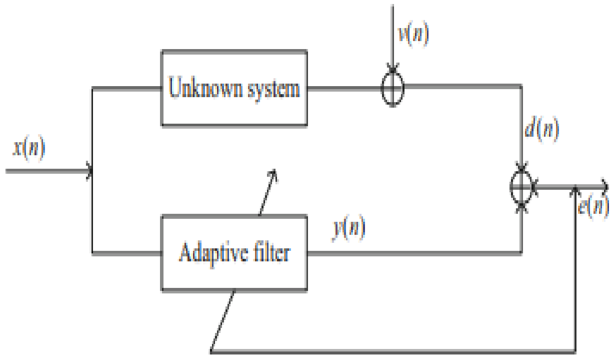


Fig.1. LMS adaptive filtering.

C. Hybrid model of WPT and LMS algorithm

This section describes the proposed hybrid mode of wavelet packet transform and LMS algorithm. According to the proposed approach, first of all, we process the speech signal through the wavelet packet transform where it generates the tree structure of wavelet packets. Below given figure depicts a three level wavelet packet decomposition structure.

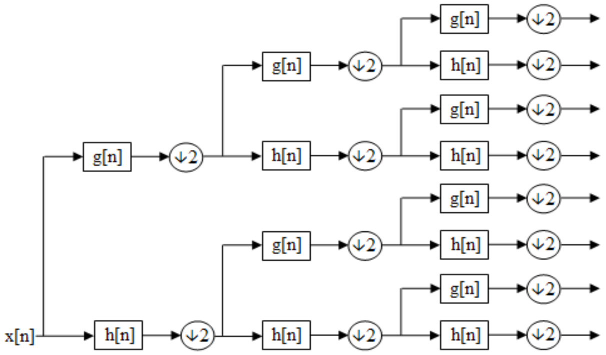


Fig.2. Wavelet packet decomposition

This decomposed data is processed through the thresholding model and wavelet filtered audio signal is obtained as an intermediate phase. Later, this intermediate audio is processed through the LMS algorithm and final filtered output is generated.

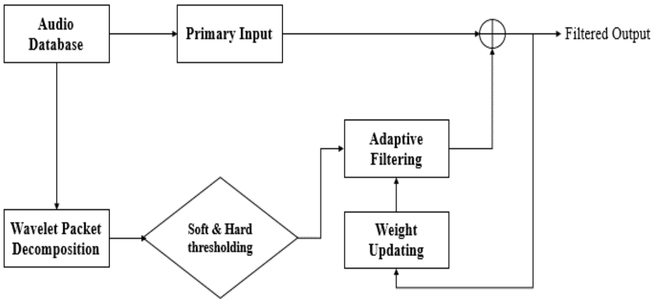


Fig.3. Architecture of proposed hybrid WPT and LMS approach for speech filtering.

IV. RESULTS AND DISCUSSION

In this segmentwe describe the experimental exploration of the proposed speech enhancement technique and compared the obtained performance with existing techniques. For this experiment, we have considered the NOIZEUS speech corpus database [14,29]. This dataset comprises a total of 30 sentences which are generated by three female and three male participants. These datasets samples are deterioratedby eight types of noise which are generated at varied SNR levels. These noises include airport, restaurant, train, car, babble, and exhibition hall and railway station noise. The outcome of the proposed hybrid approach is quanitfied in terms of Perceptual Evaluation of Speech Quality (PESQ), cepstrum distance measures (CEP),Mean Opinion Score (MOS), Frequency weighted SNR, Short Time Objective Ineligibility measure (STOI), Mean of SNR, Means of Segmented SNR, signal distortion (Csig), Cbak which is a compound estimate for noise distortion, a compoundestimate for overall speech quality (Covrl), Mean LLR, and Itakura-Saito. Given the table below describes the evaluation parameters.

Table.1. Description of performance evaluation parameters

Parameter	Measuring quantity	Range	Description
PESQ	Speech Quality	-0.5 - 4.5	Higher the value implies better quality
MOS	Speech Quality	1-5	Higher is better
CEP	Error	[0-10]	Minimum is better
Frequency weighted SNR	Speech quality	Generally the average range of 10 to 35 dB	Higher is better
SIG	distortion	1-5	Higher is better
BAK	distortion	1-5	Higher is better
OVRL	Speech quality		

These parameters can be computed as follows:

- PESQ: the PESQ measurement is a complex parameter and recommended by ITU-T to assess the speech quality. The PESQ is computed based on the linear combination of average asymmetrical disturbance A_{ind} and average disturbance D_{ind}. This can be computed as:

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \tag{13}$$

Where a_0, a_1 and a_2 are the three constant parameters whose values are 4.5, -0.1and -0.0309

- Log-likelihood ratio (LLR): the LLR is computed as

$$d_{LLR}(\vec{a}_p, \vec{a}_c) = \log\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T}\right) \tag{14}$$

Where $\vec{a}_c$ denotes the LPC vector obtained from original speech frame, $\vec{a}_p$ denotes the LPC vector obtained from the enhanced speech frame and R_c represents the autocorrelation of original speech signal.

- Itakura-Saito (IS): the IS parameter can be computed as follows:

$$d_{IS}(\vec{a}_p, \vec{a}_c) = \frac{\sigma_c^2}{\sigma_p^2}\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T}\right) + \log\left(\frac{\sigma_c^2}{\sigma_p^2}\right) - 1 \tag{15}$$

Where σ_c is the LPC gain of clean signal whereas and σ_p represents the LPC gains of enhanced speech signal.

- Cepstrum coefficients: the CC can be obtained as follows:

$$d_{CEP}(\vec{c}_c, \vec{c}_p) = \frac{10}{10 \log 10} \sqrt{2 \sum_{k=1}^{p} \left[c_c(k) - c_p(k)\right]^2} \tag{16}$$

Where $\vec{c}_c$ denotes the LPC gain of clean signal and $\vec{c}_p$represents the LPC gains of enhanced speech signal.

We apply this proposed scheme of the considered data on varied types of noise at varied SNR levels. Below given figure 2, 3,4 and 5 shows the outcome of speech filtering for airport, AWGN, Babble and restaurant noises for 0dB, 5 dB, 10dB and 15 dB, respectively. Moreover, we presented the spectrogram of noisy and recovered signal.

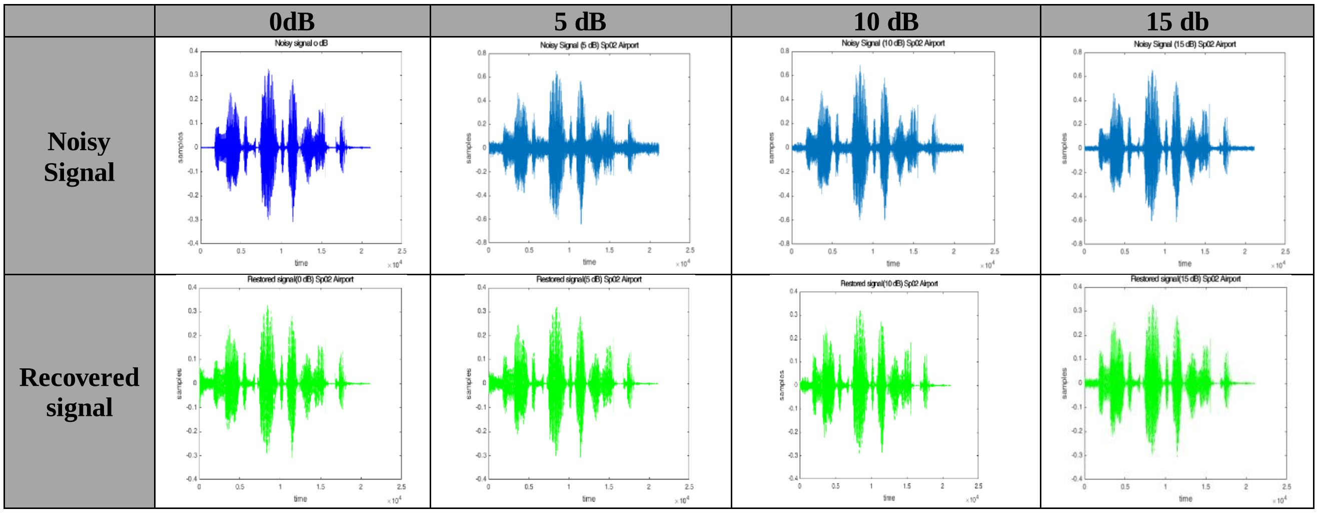


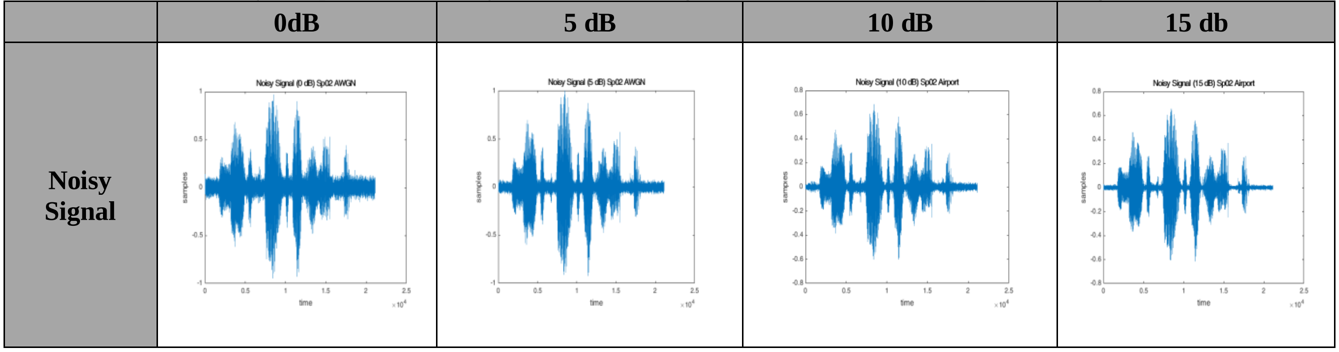
Figure 4. Illustration of noisy and recovered signal for airport noise.

For this airport noise scenario, we measured the performance by several parameters as mentioned in table 1. The obtained performance is given in table 2 for airport noise.

Table.2. speech denoising performance for airport noise.

Airport	0 dB	5 dB	10 dB	15 dB
Raw PESQ	3.8157008	3.7607	3.7851	3.8503
MOS	3.9558386	3.8912	3.9202	3.9952
CEP	0.6889	0.6958	0.7356	0.6886
Freq weighted SNRseg	29.34	28.6853	28.3375	28.9006
STOI	0.9972	0.9877	0.9861	0.9915
snr_mean	17.46	17.4742	17.4712	17.4717
segsnr_mean	25.58	24.7125	23.7843	24.0221
Csig	5	5	5	5
Cbak	5	4.9589	4.9181	4.9642
Covl	4.6174	4.5693	4.5957	4.6509
Mean LLR	0.0516	0.0438	0.0423	0.0371
Itakura-Saito	0.2063	0.2164	0.1858	0.1606

Similarly, we present the noisy and recovered signal for AWGN noise as depicted in figure 3.



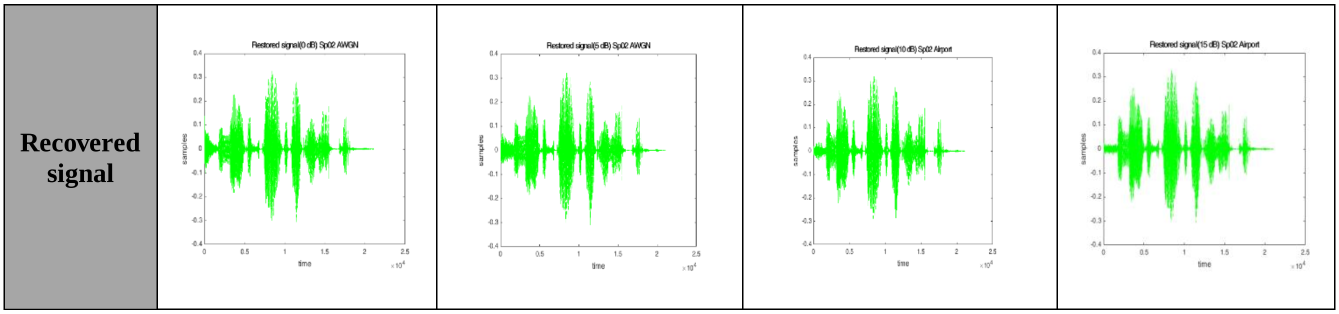


Figure 3. Illustration of noisy and recovered signal for AWGN noise.

For this experiment also, we measure the performance by adding AWGN noise at different SNR levels. The obtained denoising performance is presented in table 3.

Table.3. Speech denoising performance for AWGN noise

AWGN (dB)	0	5	10	15
Raw PESQ	4.1312	3.8706	3.8337	3.9149
MOS	4.2772	4.0179	3.9765	4.0661
CEP	0.3571	0.7183	0.7642	0.735
Freq weighted SNRseg	32.2546	30.8526	30.5706	30.5832
STOI	0.9998	0.9968	0.9942	0.994
snr_mean	17.4736	17.4763	17.4562	17.4386
segsnr_mean	31.0867	29.4284	28.4479	28.281
Csig	5	5	5	5
Cbak	5	5	5	5
Covl	4.8913	4.6456	4.6141	4.6922
Mean LLR	0.0361	0.0878	0.0902	0.0753
Itakura-Saito	0.1405	0.2752	0.288	0.2528

Below given figure 4 depicts the outcome of noisy and recovered signal by using proposed speech enhancement approach for babble noise at different SNR levels.

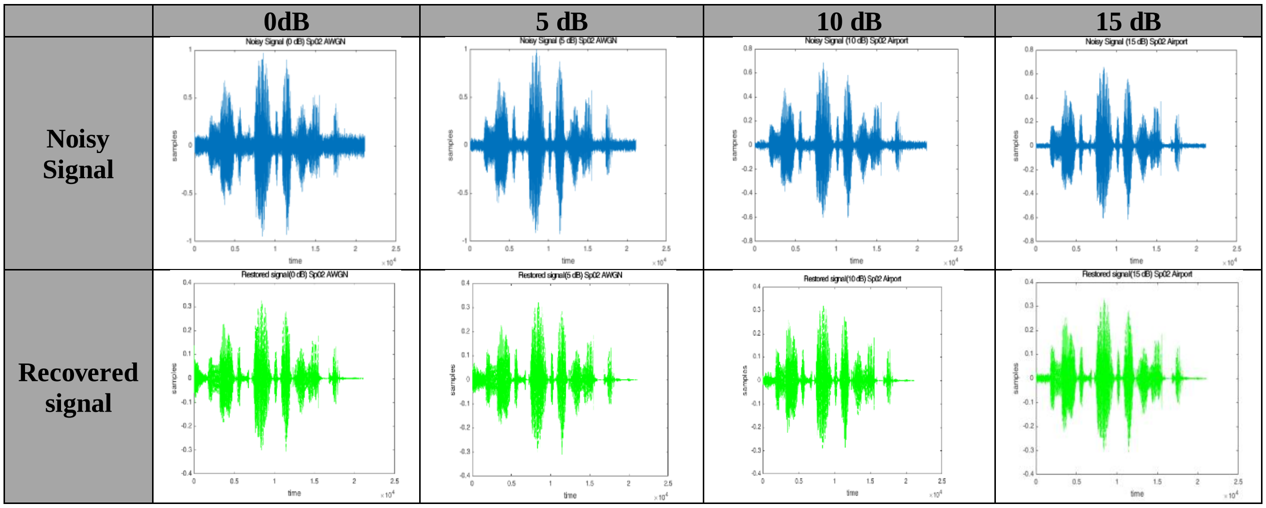


Figure 4. Illustration of noisy and recovered signal for Babble noise.

For babble noise, we assess the execution of proposed approach based on several parameters. The obtained performance is presented in below given table 4.

Table.4. Speech denoising performance for babble noise

Babble (dB)	0	5	10	15
Raw PESQ	3.7916	3.8295	3.7402	3.8593
MOS	3.9278	3.9717	3.8665	4.0053
CEP	0.6798	0.6925	0.7404	0.6831
Freq weighted SNRseg	29.6167	28.5633	28.5814	28.8296
STOI	0.9937	0.9931	0.988	0.9948
snr_mean	17.4727	17.4747	17.4715	17.4705
segsnr_mean	26.332	24.3719	23.9753	23.9464
Csig	5	5	5	5
Cbak	5	4.9718	4.9053	4.9664
Covl	4.5901	4.6287	4.5552	4.6628
Mean LLR	0.057	0.039	0.0443	0.0333
Itakura-Saito	0.2133	0.2079	0.194	0.1574

Finally, figure 5 depicts the outcome for restaurant noise using proposed speech filtering scheme.

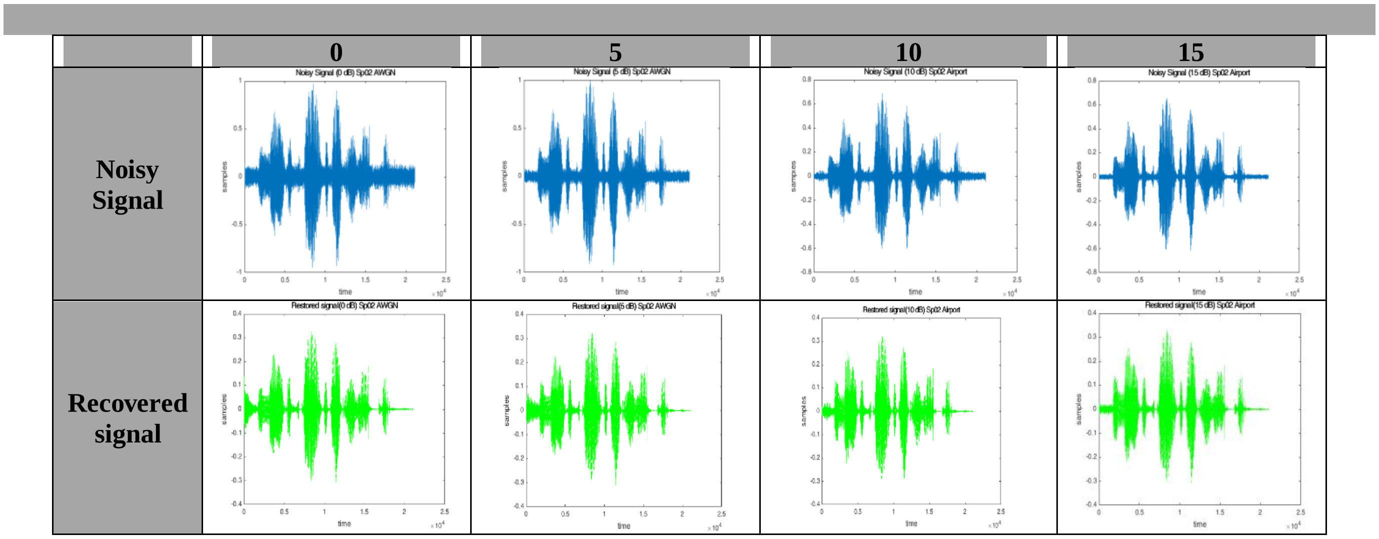


Figure 5. Illustration of noisy and recovered signal for restaurant noise.

Finally, we quanitify the speech enhancement performance for restaurant noise. The obtained outcome for this analysis is presented in below given table 5.

Table.5. Speech denoising performance for restaurant noise

Restaurant	0 dB	5 dB	10 dB	15 dB
Raw PESQ	3.8221	3.771	3.8453	3.8915
MOS	3.9632	3.9035	3.9896	4.0408
CEP	0.7134	0.8562	0.7612	0.7034
Freq weighted SNRseg	29.5957	28.1679	28.3895	28.968
STOI	0.9952	0.9929	0.9924	0.9957
snr_mean	17.4631	17.4715	17.4698	17.4705
segsnr_mean	26.2839	23.7732	23.9071	23.9906
Csig	5	5	5	5
Cbak	5	4.9048	4.9515	4.9872
Covl	4.612	4.5721	4.6417	4.6894
Mean LLR	0.0625	0.0548	0.0412	0.0369
Itakura-Saito	0.1909	0.1946	0.1687	0.1447

Noise	[24]		[25]		[26]		[27]		Proposed	
	R	σ	R	σ	R	σ	R	σ	R	σ
Airport	0.8905	0.1201	0.8605	0.1406	0.9001	0.1105	0.8812	0.1324	0.9261	0.1110
Babble	0.9113	0.1391	0.8822	0.2124	0.8604	0.2310	0.9014	0.1411	0.9352	0.1241
Car	0.9552	0.1301	0.8960	0.1502	0.9041	0.1412	0.8715	0.1721	0.9587	0.1025
Exhibition	0.9298	0.1042	0.9215	0.1102	0.9125	0.1215	0.9035	0.1265	0.9505	0.1045
Restaurant	0.9021	0.1296	0.8881	0.1568	0.8602	0.1815	0.8689	0.1678	0.9206	0.1135
Station	0.8892	0.1504	0.9035	0.1356	0.9165	0.1263	0.8705	0.1701	0.9152	0.1335
Street	0.9701	0.0958	0.9258	0.1192	0.9058	0.1295	0.8902	0.1401	0.9785	0.0904
Train	0.9056	0.1487	0.8891	0.1881	0.9165	0.1418	0.8905	0.1775	0.9365	0.1234

Further, we compared the outcome of proposed approach in terms of average correlation and standard deviation withexisting speech enhancement techniques as mentioned in [24,28,30]. The correlation can be computed as:

$$R = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_i (y_i - \bar{y})^2 (x_i - \bar{x})^2}} \qquad (17)$$

Where x_i denotes the average objective MOS of noise condition and $y_i$represents the subjective MOS, x represents the average of all values of x_i and y is the average of all y_i. Similarly, we considered standard deviation of error which represents the difference between subjective and objective MOS. This is computed as follows:

$$\sigma = \sigma_s \sqrt{1 - R^2} \qquad (18)$$

Below given table 6shows the comparative analysis in terms of correlation and standard deviation.

Similarly, we further measured the mean objective score outcome in terms ofPESQ, STOI,Csig, Cbak, Covl, and SegSNR. The obtained performance is compared with various state-of-art techniques as mentioned in [21].

Technique	Csig	Cbak	Covl	PESQ	STOI	SegSNR
LSTM-CKFS[21]	2.63	2.55	2.42	1.99	77.58	6.54
EEUE-FCNN[21]	2.76	2.66	2.56	2.05	79.45	6.93
Deep Xi-KF[21]	3.11	2.83	2.72	2.16	81.89	7.14
Deep X-Resnet TCN MMSE LSA[21]	3.38	3.02	2.81	2.22	82.05	7.67
DeepLPCResNet[21]	3.49	3.17	2.95	2.35	84.71	8.78
Deep LPC MHANet [21]	3.66	3.32	3.14	2.59	88.41	9.21
Proposed	4.88	4.5	4.56	4.2	98.22	12.58

Finally, we present the spectrogram analysis for clean noisy and recovered audio samples. For this analysis, we have used multiple Window SavitzkyGolay(SG) filter approach. Below given figure 6 depicts the outcome of histogram of original signal.

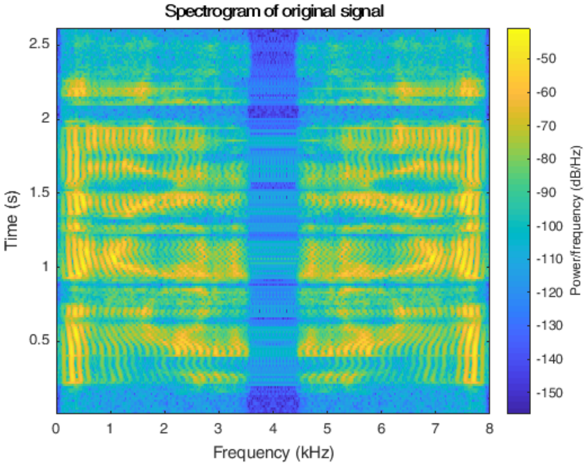


Fig.6. Spectrogram for original signal (sp02)

Further, we obtained the histogram of noisy and recovered speech signals for varied types of noises. These outcomes of spectrogram is depicted in figure 7.

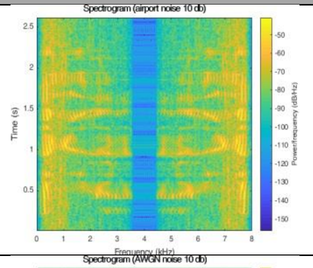
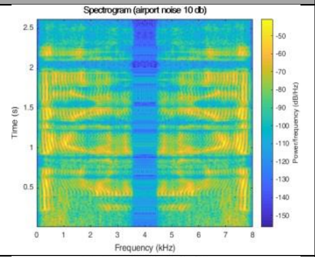
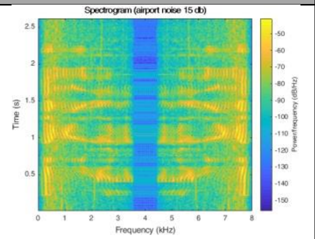
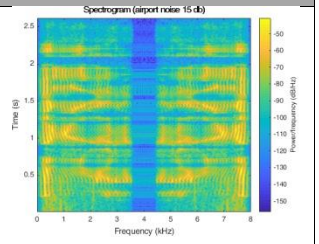
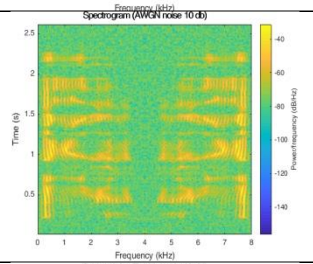
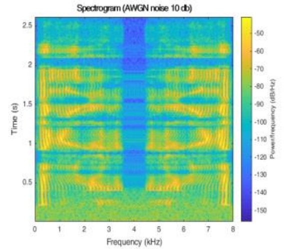
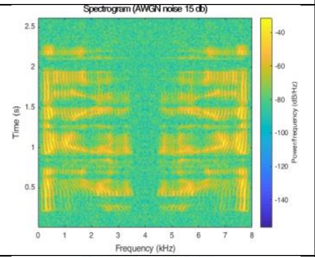
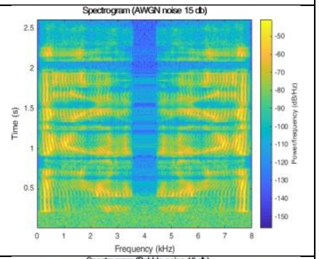
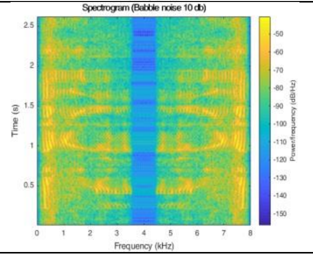
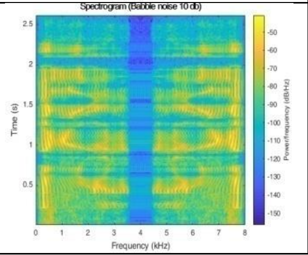
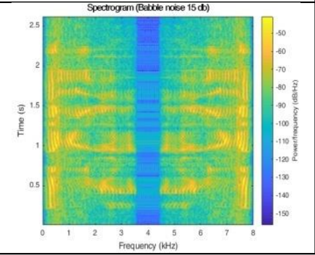
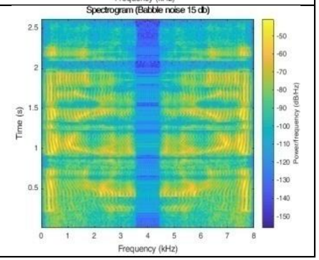
Noise Type	10 db		15 db	
	Noisy signal	Recovered signal	Noisy signal	Recovered signal
Airport				
AWGN				
Babble				

Fig.7. spectrogram analysis for diverse types of noises at varied SNR levels.

The complete experimental analysis shows a significant improvement in the speech enhancement by using proposed scheme when compared with existing techniques.

IV. CONCLUSION

This article mainly focuses on the speech enhancement technique to improve the performance of speech processing based applications. We took the advantage of transform domain and adaptive filtering schemes and presented a novel hybrid approach to deal with speech enhancement related issues. The complete proposed solution is developed into two phases, first of all, we present a wavelet packet transform scheme followed by the thresholding mechanism to minimize the noise. Further, this data is processed through adaptive LMS filtering. The experimental analysis shows a substantial enhancement in the performance when compared with the existing scheme.

ACKNOWLEDGMENT

Authors would like to sincerely thank Dr. S. S. Injaganeri**,** Principal, Basaveshwar Engineering College, Bagalkot, Karnataka and Dr. B.R.Patagundi, Principal, S. G. Balekundri Institute of Technology, Belagavi, Karnataka and dear staff member of BEC, Bagalkot and SGBIT, Belagavi for their kind support in carrying out this research work.

REFERENCES

[1] Dash, T. K., & Solanki, S. S., Comparative study of speech enhancement algorithms and their effect on speech intelligibility. In 2017 2nd International conference on communication and electronics systems (ICCES) (2017) 270-276. IEEE.

[2] Chiea, R. A., Costa, M. H., &Barrault, G., New insights on the optimality of parameterized Wiener filters for speech enhancement applications. Speech Communication, 109(2019) 46-54.

[3] Andersen, K. T., &Moonen, M., Robust speech-distortion weighted interframe Wiener filters for single-channel noise reduction. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(1)(2017)97-107.

[4] Enzner, G., &Thüne, P., Robust MMSE filtering for single-microphone speech enhancement, In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2017) 4009-4013. IEEE.

[5] Pardede, H., Ramli, K., Suryanto, Y., Hayati, N., &Presekal, A., Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter. Electronics, 8(8)(2019) 897.

[6] Dionelis, N., & Brookes, M., Phase-aware single-channel speech enhancement with modulation-domain Kalman filtering. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(5) 937-950PDCA12-70 data sheet,OptoSpeedSA, Mezzovico, Switzerland., (2018).

[7] Choi, H. S., Kim, J. H., Huh, J., Kim, A., Ha, J. W., & Lee, K., Phase-aware speech enhancement with deep complex u-net. In International Conference on Learning Representations., (2018).

[8] Bhowmick, A., & Chandra, M., Speech enhancement using voiced speech probability based wavelet decomposition. Computers & Electrical Engineering, 62(2017)706-718.

[9] Benesty, J., & Cohen, I., Single-channel speech enhancement in the STFT domain. In Canonical Correlation Analysis in Speech Enhancement (2018)37-57. Springer, Cham.

[10] Taherian, H., Wang, Z. Q., Chang, J., & Wang, D., Robust speaker recognition based on single-channel and multi-channel speech enhancement. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28(2020) 1293-1302.

[11] Wang, Z. Q., Wang, P., & Wang, D., Complex spectral mapping for single-and multi-channel speech enhancement and robust ASR. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28(2020)1778-1787.

[12] Taherian, H., Wang, Z. Q., Chang, J., & Wang, D., Robust speaker recognition based on single-channel and multi-channel speech enhancement. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28(2020)1293-1302.

[13] Mavaddaty, S., Ahadi, S. M., &Seyedin, S., Speech enhancement using sparse dictionary learning in wavelet packet transform domain. Computer Speech & Language, 44 (2017) 22-47.

[14] Ram, R., &Mohanty, M. N., Use of radial basis function network with discrete wavelet transform for speech enhancement. International Journal of Computational Vision and Robotics, 9(2) (2019) 207-223.

[15] ITU-T P.835, Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, ITU-T Recommendation (2003) 835.

[16] Sharma, N., Singh, M. K., Low, S. Y., & Kumar, A., Weighted Sigmoid-Based Frequency-Selective Noise Filtering for Speech Denoising. Circuits, Systems, and Signal Processing, 40(1)(2021) 276-295.

[17] Srinivasarao, V., &Ghanekar, U., Speech intelligibility enhancement: a hybrid wiener approach. International Journal of Speech Technology, 23(3)(2020) 517-525.

[18] Yu, H., Zhu, W. P., Ouyang, Z., & Champagne, B., A hybrid speech enhancement system with DNN based speech reconstruction and Kalman filtering. Multimedia Tools and Applications, 79(43)(2020) 32643-32663.

[19] Roy, S. K., Nicolson, A., &Paliwal, K. K.. DeepLPC: A deep learning approach to augmented Kalman filter-based single-channel speech enhancement. IEEE Access.

[20] Li, A., Yuan, M., Zheng, C., & Li, X., Speech enhancement using progressive learning-based convolutional recurrent neural network. Applied Acoustics, 166(2020) (2021), 107347.

[21] Roy, S. K., Nicolson, A., &Paliwal, K. K., DeepLPC-MHANet: Multi-Head Self-Attention for Augmented Kalman Filter-based Speech Enhancement. IEEE Access., (2021).

[22] Indra, J., Kiruba Shankar, R., Kasthuri, N., &GeethaManjuri, S., A Modified Tunable–Q Wavelet Transform Approach for Tamil Speech Enhancement. IETE Journal of Research, (2020),1-14.

[23] Garg, A., &Sahu, O. P., A hybrid approach for speech enhancement using Bionic wavelet transform and Butterworth filter. International Journal of Computers and Applications, 42(7)(2020), 686-696.

[24] Zhou, Weili; Zhu, Zhen., A novel BNMF-DNN based speech reconstruction method for speech quality evaluation under complex environments. International Journal of Machine Learning and Cybernetics, (), –. doi:10.1007/s13042-020-01214-3., (2020).

[25] Fu SW, Tsao Y, Hwang HT et al., Quality-net: an end-to-end non-intrusive speech quality assessment model based on BLSTM. arXiv preprint arXiv:1808.05344., (2018).

[26] Soni MH, Patil HA Novel subbandautoencoder features for non-intrusive quality assessment of noise suppressed speech. In: 2016 conference of the international speech communication association on interspeech. IEEE, (2016) ,3708–3712

[27] Rajesh KD, Arun K.,Non-intrusive speech quality assessment using multi-resolution auditory model features for degraded narrowband speech. IET Signal Proc 9 (2015) 638–646.

[28] JagadishS.Jakati and ShridharS.Kuntoji., Speech Enhancement Using Novel Time-Frequency Analysis Techniques: A Survey on Comparison., International Journal of Advanced Trends in Computer Science and Engineering, 9(4)(2020)4229-4234.

[29] JagadishS.Jakati and ShridharS.Kuntoji., Efficient Speech De-noising Algorithm using Multi-levelDiscrete Wavelet Transform and Thresholding,International Journal of Emerging Trends in Engineering Research, 8(6)(2020) 2472-2480.

[30] JagadishS.Jakati and ShridharS.Kuntoji, A Noise Reduction Method Based on Modified LMS Algorithm of Real-time Speech Signals,WSEAS TRANSACTIONS on SYSTEMS and CONTROL, 16(2021)162-170.