International Journal of Informatics and Communication Technology (IJ-ICT)
Vol.9, No.2, August 2020, pp. 63~72
ISSN: 2252-8776, DOI: 10.11591/ijict.v9i2.pp63-72

Speech enhancement using modified wiener filter based MMSE and speech presence probability estimation

V. Vijayasri Bolisetty[1], U. Yedukondalu[2], I. Santiprabha[3]
[1]Department of Electronics and Communication Engineering, Aditya College of Engineering and Technology, India
[2]Electronics & Communication Engineering Department, Sri Vasavi Engineering College, India
[3]Department of Electronics & Communication Engineering, University College of Engineering, JNTUK, India

Article Info

Article history:

Received Mar 6, 2019
Revised Dec 12, 2019
Accepted Jan 12, 2020

Keywords:

A-priori-SNR
Gaussian distribution
Noise tracking
PSD
SPP

ABSTRACT

In the present-day communications speech signals get contaminated due to various sorts of noises that degrade the speech quality and adversely impacts speech recognition performance. To overcome these issues, a novel approach for speech enhancement using Modified Wiener filtering is developed and power spectrum computation is applied for degraded signal to obtain the noise characteristics from a noisy spectrum. In next phase, MMSE technique is applied where Gaussian distribution of each signal i.e. original and noisy signal is analyzed. The Gaussian distribution provides spectrum estimation and spectral coefficient parameters which can be used for probabilistic model formulation. Moreover, a-priori-SNR computation is also incorporated for coefficient updation and noise presence estimation which operates similar to the conventional VAD. However, conventional VAD scheme is based on the hard threshold which is not capable to derive satisfactory performance and a soft-decision based threshold is developed for improving the performance of speech enhancement. An extensive simulation study is carried out using MATLAB simulation tool on NOIZEUS speech database and a comparative study is presented where proposed approach is proved better in comparison with existing technique.

Corresponding Author:

V. Vijayasri Bolisetty,
Department of Electronics and Communication Engineering,
Aditya College of Engineering and Technology,
Aditya Nagar, ADB Rd, Surampalem, Andhra Pradesh, 533437, India.
Email: vasudha.tweety@gmail.com

1. INTRODUCTION

Now-a-days, speech communications have drawn demand with the advent of different broadband and multimedia applications in rapid usage at different environments. In such scenarios, preserving the information in speech signal plays an important role which has impact on the efficient communication. Moreover, due to low-cost digital signal processors and memory chips, speech processing applications, voice communication systems and voice recognition systems are widely utilized in various real-time application scenarios [1]. Various communication devices also have been manufactured i.e. mobile phones and mics etc. which can be used for indoor as well as outdoor applications. In these environments, background noise is present which gets accumulated with the original signal resulting degraded speech quality and it may lead to the performance degradation of voice recognition systems [2].

In outdoor environments like heavy traffic, bus terminals, malls, etc., presence of various types of background noises contaminate the original signal that leads to improper communication [3] in real-time applications like speaker recognition system, mobile communication, hearing aids etc. Hence noise is to be

estimated and reduced to improve the speech quality. During last decade, various enhancement techniques have been presented for speech signal enhancement. Substantially, these techniques are categorized single-channel and multi-channel speech enhancement. Usually, back ground noise features can be identified using single channel speech enhancement and the effects of reverberation can be reduced through multichannel speech enhancement. Though, Multichannel schemes show significant performance for speech enhancement when compared to single-channel speech enhancement, the enhancement is to be done through each microphone individually. Hence the work is concise to single channel speech enhancement.

The speech enhancement algorithm aims to rectify the damaged input or output signal and to increase the performance of the communication link. The damaged speech signal generates huge trouble mainly for speech recognition applications. The quality and intelligibility of speech are damaged by the noise involved in the speech signal. The term intelligibility denotes the understandability of final outcome of the speech signal. The accuracy of the exact content of speech signal is termed as quality. The different types of speech enhancement algorithm used to deduct noise are Adaptive or non-adaptive, frequency or time domain [4]. Over the last years, several techniques have been presented for speech signal enhancement i.e. Spectral Subtraction, Wavelet based methods, model-based techniques and filtering techniques [5, 6]. Furthermore, speech enhancement techniques can be categorized as spectral and temporal processing techniques. According to the spectral domain process, corrupted signal is processed through the transform domain technique whereas temporal processing method uses time-domain analysis for improving the quality of the speech signal.

The complete article is organized as follows: section 2 presents a brief literature survey and recent advancements in speech enhancement techniques. Section 3 presents proposed speech enhancement system modeling. Results are analyzed and discussion is elaborated in section 4, finally, concluding remarks are presented in the section 5.

2. LITERATURE SURVEY

Conventional multiband speech enhancement acquires two operations: one is splitting the spectrum into frequency bands, and the other is executing speech enhancement in each band independently. The polo-interaction problem in the spectral domain leads to the suppression of few coefficients in the estimation of clean speech by the influence of the formants in the neighboring bands and thus grades in poor quality. To reduce the domination of stronger formants over the neighboring bands, the assessment of clean speech is done by, in the temporal domain. The unsuppressed speech is filtered into various equivalent rectangular bandwidth based subbands and followed by enhancement of spectral speech in each band based on Discrete Cosine Transform (DCT) using Spectral Subtraction/Minimum Mean Square Error (MMSE) [7].

Park et. al. [8] discussed about the use of speech enhancement technique in mobile communication systems. Authors developed an efficient scheme for noise reduction which can improve the performance of speech recognition. However, mobile devices have limited capacity which motivates to develop a low complex scheme for noise reduction. In this work, a speech coder is also utilized for packet data estimation. In general, this work uses pitch information for comb filtering. Adaptive filtering technique has a significant impact on the speech processing system. Adaptive filtering has been used widely in the various applications such as channel equalization, system identification and echo cancellation. Although, MMSE and DWT shows improvement in speech enhancement, their performance is poor at low SNR conditions [9, 10].

The Dual Tree Complex Wavelet Packet Transform (DTCWPT) employed in [11] provides a solution to avoid the aliasing and oscillations due to shift invariance in DWT. Even though compatibility exists between MMSE based SPP and DTCWPT, the loss of intelligibility in the reconstructed speech due to the finite wavelet filters. In this field, Choi. et. al. [12] improved the existing adaptive filtering and developed a novel sub-band adaptive filter. Further, this technique takes the advantage of norm-optimization and norm as the computation of cost function. Authors have claimed that this technique is capable to improve the system performance and robustness in impulsive noise scenarios.

The optimally modified MMSE based log Spectral Amplitude estimator (OM-LSA) improves the performance compared to the MMSE in case of stationary and some sort of non-stationary noises but shows significantly poor performance at low and high SNR conditions [13, 14]. SMPO based MMSE technique for speech enhancement improves the PESQ in all kinds of noisy environments. Binary masking results good speech quality but sometimes intelligibility of the speech may be lost due to over thresholds in masking noise coefficients. SMPO-Weiner is going to provide better solution in these issues [15]. Hence combination of Weiner filter with MMSE based SPP through soft decision may improve the speech intelligibility in addition to speech quality.

– Challenges in Speech Enhancement: Speech is considered as powerful mode of communication not only for humans but also for human machine interface. Sometimes the speech signal travels through the noisy

medium before reaching to the listener (recognition system). Now-a-days, automated speech processing systems have gained lot of attraction from researchers and have been adopted widely in the real-life scenarios. State-of-art models of automated speech processing systems works well with controlled environment but real-time systems suffer from various background noises, reverberation and speech from other speakers, which causes partial loss of information to complete loss especially at low SNR. Noise removal can be done only by identifying the characteristics of the noise through preprocessing. However, huge number of researches have been carried out in this field but due to computational complexities, speech enhancement still remains a challenging task as the process itself introduces complexity.

– Contribution of the work. Main contributions of this work are as follows:
  a. Initially, the features of the speech signal, like silence is estimated from the noisy signal spectrum through Speech Presence Probability (SPP) estimation.
  b. Extracting the noise characteristics from the silence.
  c. Finally, MMSE (Noise Tracker) based noise power estimation to suppress the noisy signal using noise power tracking scheme.

3. SINGLE-CHANNEL SPEECH ENHANCEMENT

Assuming that the majority of the transmission noise is additive, the speech enhancement for single-channel is represented in (1),

$$y(n) = x(n) + d(n) \quad (1)$$

where *x(n), d(n), y(n)* corresponds to speech, noise and noisy speech signals and *n* indicates discrete time index. Both *x(n)* and *d(n)* are independent with zero mean value. The block diagram of single channel speech enhancement is represented in Figure 1. To estimate the clean speech signal *x(n)*, the noisy signal *y(n)* is processed using speech enhancement algorithms.

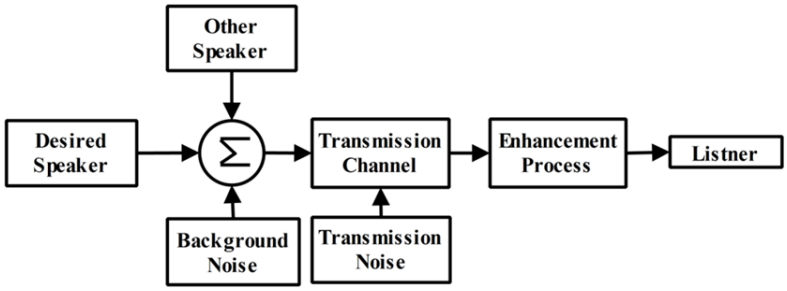

Figure 1. General Speech Enhancement Systems

An analysis of speech enhancement method for a noisy speech signal *y(n)* is illustrated in Figure 2. Initially, the noisy signal is segmented into overlapping frames, then transformed into frequency domain using DFT or STFT. To achieve good quality of speech signal, DFT technique is applied as it easily understands the spectral content of the signal.

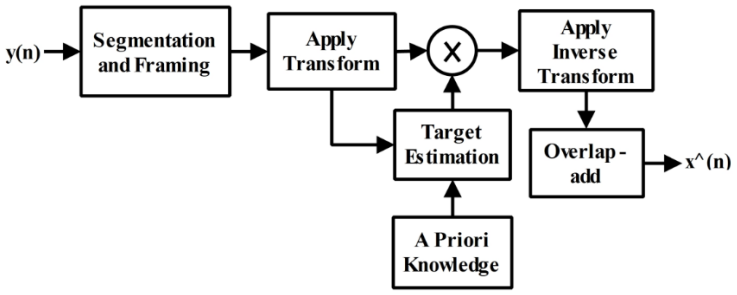

Figure 2. Analysis – synthesis of a noisy speech signal

Prior to the speech enhancement technique, the noise characteristics can be identified by the statistical model. The gain values are calculated based on the features extracted from the original input so that an upgraded speech signal is acquired. The target evaluation can be made through the power spectrum density of the noisy speech segments and noise. This helps to generate the gain coefficients of the signals.

Wiener filtering helps to reduce the Mean Square Error (MSE) among the actual signal and the estimated signal and to boost up the original signal from the noisy signal. Similarly, the noise can be minimized through hard or soft thresholding. Sometimes the important components of the original speech might be lost due to over thresholding.

3.1. Modified wiener filtering technique

The advantage of the modified adaptive wiener filter is that the speech signal is processed through filter by varying local statistics such as mean, variance. Here, the mean value of the additive noise *d(n)* is considered to be zero and holds a white nature with variance σ_d^2. Thus, the power spectrum $P_d(\omega)$ can be expressed in (2),

$$P_d(\omega) = \sigma_d^2 \tag{2}$$

Here the segmented speech signal $x_j(n)$ is treated as stationary and thus can be modeled by (3),

$$x_j(n) = m_{xj} + \sigma_{xj} w(n) \tag{3}$$

where m_{xj} is the local mean and σ_{xj} is the standard deviation of $x_j(n)$. $w(n)$ is a unit variance noise with a zero mean. For convenience x_j is represented as x, the mean of the original signal m_s is equal to average mean value of all the j frames, m_x. The Wiener filter transfer function can be estimated by (4),

$$H(\omega) = \frac{P_X(\omega)}{P_X(\omega)+P_d(\omega)} = \frac{\sigma_X^2}{\sigma_X^2+\sigma_d^2} \tag{4}$$

The impulse response of the wiener filter can be achieved by applying Inverse Transformation to $H(\omega)$ and is given in (5),

$$h(n) = \frac{\sigma_X^2}{\sigma_X^2+\sigma_d^2}\delta(n) \tag{5}$$

As the m_x and σ_x are updated at each sample, the speech signal can be estimated from (6), and is denoted as $\hat{x}$(n) .

$$\hat{x}(n) = m_x(n) + \frac{\sigma_x^2(n)}{\sigma_x^2(n)+\sigma_d^2}(x(n) - m_x(n)) \tag{6}$$

Now estimating and tracking the noise frames in the noisy signal is the crucial task in the process of speech enhancement. Usually, the voice activity detector is employed for finding the presence of noise. But in this Weiner filtering method, frame energy is synchronized with the minimum frame energy. Minimum energy variation is directly proportional to the signal conditions. Therefore, by using smoothed 32 points of the spectrum the spectral deviance is calculated. To gain noise spectrum these points are verified and upgraded additionally. Then the level of energy is fixed to 10dB. So by analyzing the energy level which is more than 10 dB and RMS level more than 8 dB, the presence of speech can be identified. When there is no availability of speech this noise spectrum $N(\omega,m)$ can be measured for varied time samples. The ultimate plan is to implement MMSE based noise power estimation technique. In this process, speech & noise spectral coefficients have complex Gaussian distribution which can be expressed in (7).

$$P_D(n) = \frac{1}{\sigma_D^2 \Pi} exp\left(-\frac{|n|^2}{\sigma_x^2}\right) \tag{7}$$

Hence, the noisy power spectral coefficients can be expressed as given in (8).

$$P_y(x) = \frac{1}{\sigma_D^2(1+\xi)} exp\left(-\frac{|x|^2}{\sigma_D^2(1+\xi)}\right) \tag{8}$$

3.2. Noise PSD estimation and tracking

The noise is represented as $n = De^{j\Delta}$and $x = Re^{j\theta}$. Spectral coefficients can be transformed into polar coordinates using (9).

$$p_{D,\Delta}(d,\delta) = \frac{1}{\sigma_D^2\pi} exp\left(-\frac{d^2}{\sigma_D^2}\right) \quad (9)$$

Probability distribution $p_{x|D,\Delta(x|(d,\delta))}$ of signal is given in (10).

$$p_{x|D,\Delta(x|d,\delta)} = \frac{1}{\pi\sigma_s^2} exp\left(\frac{2dr\,cos(\delta-\theta)-r^2-d^2}{\sigma_s^2}\right) \quad (10)$$

In the MMSE method there are several noise power estimators present in the noisy signal periodogram estimation. Let us consider a priori SNR ξ and estimated noise power is σ_d^2. The noise periodogram can be obtained in (11).

$$|\hat{n}|^2 = E(|n|^2 x) = \left(\frac{1}{1+\hat{\xi}}\right)|x|^2 + \frac{\hat{\xi}}{1+\hat{\xi}}\hat{\sigma}_d^2 \quad (11)$$

From the (11) it is identified that the noise periodogram can be gained. But here the signal gets altered from time to time, therefore the spectral density has to be revised frequently. To update the spectral density parameters the recursive smoothing is applied as shown in (12).

$$\hat{\sigma}_d^2(l) = \alpha\hat{\sigma}_d^2(l-1) + (1-\alpha)|D(l)|^2 \quad (12)$$

Where α = 0.0. Whenever $\xi < 1$, the noise periodogram is updated. In the same way by verifying a priori SNR factors the spectral noise power is improved. Hence the sum of the observed noisy signal and previous estimation of spectral noise power $\hat{\sigma}_{N^2}$ is represented by MMSE. From this estimation the priori SNR value is obtained which is in between 0 and 1. In A priori SNR can be obtained by following (13).

$$\hat{\xi} = max(0, \hat{\xi}^{ml}) = max(0, \hat{\gamma} - 1) \quad (13)$$

where *ml* denotes maximum likelihood computation and$\hat{\gamma}(l) = \frac{|x(l)|^2}{\sigma_n^{2(l-1)}}$. The (13) is unbias in nature. Hence, a priori SNR information is not known. So, the estimation of this unknown SNR can generate abias factor as expressed in (14).

$$B^{-1}(\xi) = \left((1+\xi)\gamma\left(2, \frac{1}{1+\xi}\right) + e^{-\frac{1}{1+\xi}}\right) \quad (14)$$

where γ represents incomplete gamma function. According to the nature of a priori SNR, estimator *E(|d|²x)* value is unbiased whereas for low SNR values it is under-biased.

3.3. Soft–decision technique for noise presence estimation

Speech presence or absence is to be identified, to identify the noise characteristics. Generally, noise characteristics can be obtained when the speech is absent. The speech or silence part classification can be done by (15).

$$\left|\hat{d}(l)\right|^2 = E\left(\left|\hat{d}(l)\right|^2 x(l)\right) = \begin{cases} \hat{\sigma}_d^2(l-1), & if\hat{\gamma}(l) \geq 1 \\ |\gamma(l)|^2, & if\hat{\gamma}(l) < 1 \end{cases} \quad (15)$$

However, proposed model utilizes soft decision-based framework for noise estimation in the presence or absence of speech. The probability of speech presence or absence can be estimated from (16) and (17).

$$p(H_1|\zeta)|_{P_{Z|(H_1|\zeta)} >> P_{Z|(H_0|\zeta)}} = 1 \quad (16)$$

$$p(H_1|\zeta)|_{P_{Z|(H_1|\zeta)} << P_{Z|(H_0|\zeta)}} = 0 \quad (17)$$

where,

$$p_{z|H_0}(\zeta) = \frac{1}{\Gamma(\upsilon)}\upsilon^{\upsilon}\zeta^{\upsilon-1}\, exp(-\upsilon\zeta) \quad (18)$$

$$p_{z|H_1}(\zeta) = \frac{1}{\Gamma(\upsilon)}(\frac{\upsilon}{1+\xi H_1})^{\upsilon}\zeta^{\upsilon-1}\, exp(-\upsilon\frac{\zeta}{1+\xi H_1}) \quad (19)$$

The (18) and (19) describes speech presence and speech absence respectively. Noise periodogram under speech presence conditions can be expressed in (20). For further improvement chi-square distribution-based hypothesis analysis can be applied.

$$E(|\hat{n}(l)|^2|x) = P(H_0|Y)E(|\hat{n}(l)|^2|x, H_0) + P(H_1|x)E(|n|^2|x, H_1) \quad (20)$$

There are two important components to be considered in estimating speech presence probability using chi-square approximation: observation sequence and estimated noise variance as shown in (22). Spectral component computation is applied resulting in observation sequence and estimated noise variance generation where total N number of frequency bins is considered given in (21).

$$o = [o_1, o_2, \ldots, o_k, \ldots, o_N] \quad (21)$$
$$\varepsilon = [e_1, e_2, \ldots, e_k, \ldots, e_N]$$

In next phase, chi-square calculation is applied for these frequency bins and the chi-square statistic is given by (22).

$$NS^2 = \sum_{k=1}^{N}\frac{(o_k - e_k)^2}{e_k} \quad (22)$$

Through the chi-square calculation the obtained value is compared with the threshold parameter where (N-1) total degrees of frequency bins are available.

Condition:

If *(Ns^2>Th)//* threshold value is denoted by *Th*
I(λ,k)=1 // accept condition H_1
Else
I(λ,k)=0 // accept condition H_0
End

4. RESULTS AND DISCUSSION

The noise characteristics are estimated from the noisy speech signal during the absence of speech and the speech clearness is determined from Performance Evaluation of Speech Quality (PESQ) and segmented SNR (segSNR). In the aspect of mobile application, a sample-rate of 16 KHz, 8KHz TIMIT data base signals are used as input to measure the quality of the proposed algorithm. Initially, the signals are TF-decomposed and Hann-windowed frames with a length of 256 samples. But to calculate the objective intelligibility, each frame is zero-padded up to 512 samples. As the PESQ is highly correlated with the subjective measures, it is recommended by ITU-T for measuring the performance of the processing technique. The simulation has been done in MATLAB 2013b tool.

The noisy signals such as babble, airport, car, train noise are chosen at different SNR 0dB, 5 dB, 10 dB with 30 samples of male and female speakers from NOIZEUS database are chosen for performance evaluation of the proposed technique. By using the overlap-and-add procedure the signal is synthesized. General simulation parameters like down-sampling, Number of FFT points, window length and window overlap are described in Table 1.

Table 1. simulation parameters considered

Simulation Parmeter Name	Simulation Parameter Vlaue
Signal Downsampling	8 kHz
NFFT	2^14
Window Length(ms)	32 ms
Window overlap	16 ms
Number of averaging window	10
PSD (Power Spectral density) estimation parameter	0.8

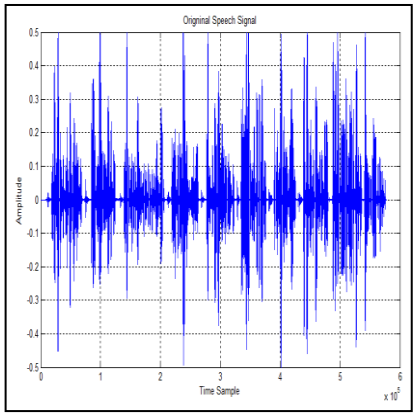

Figure 3. Original speech signal

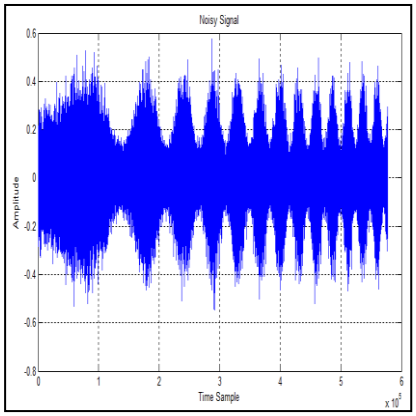

Figure 4. Noisy signal

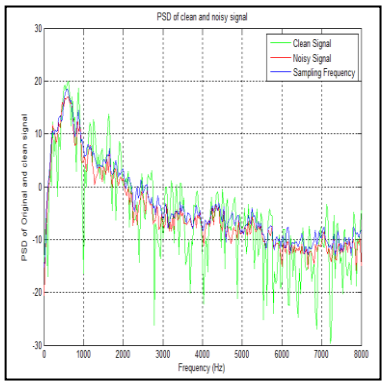

Figure 5. Power spectra

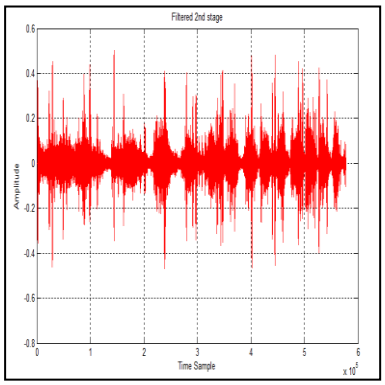

Figure 6. Filtered Signal and noise suppressed signal

Table 2. Comparative performance for babble noise

Input SNR	Parameter Name	OM-LSA [13]	SMPO [11]	MMSE-SPP [10]	BWT [16]	Wavelet Packet [11]	Proposed
0 dB	PESQ	2.06	2.01	2.08	2.25	2.11	2.37
	segSNR	3.62	4.00	2.22	0.96	5.25	5.51
5 dB	PESQ	2.42	2.38	2.47	2.29	2.44	2.52
	segSNR	3.59	3.62	1.50	2.55	4.65	4.73
10 dB	PESQ	2.86	2.83	2.87	2.50	2.87	2.99
	segSNR	3.10	4.11	0.76	3.27	4.23	4.7

The original speech, noisy signals are considered for simulation in .wav form and is represented in Figure 3 and Figure 4 respectively. Through these diagrammatical presentations the presence and absence can be identified. The probability distribution utilizes the phase analysis method and Figure 5. shows the Power spectra of original, noisy signals. Finally, filtered and noise suppressed signal analysis is represented in Figure 6.

The proposed model is compared with the other existing state of art (Pengfei Sun, Jun Qin [11]). PESQ and segSNR are the characteristics features which are used to evaluate the processing approach. By varying the noisy input SNR values from 0dB to 10dB the average values of PESQ and segSNR values are represented in the table for all samples. The Noizeus database with 30 speakers in each set contains babble noise, airport noise, car noise and train noise which is also treated as important characteristics to identify the performance of the method. The Table 2, Table 3, Table 4, and Table 5 describe the performance analysis for babble, airport, car and train noises respectively.

The calculated PESQ and segmental SNR are related with benchmark algorithms to estimate the efficiency of the system. In terms of babble and train noises of 0dB, 5dB and 10 dB SNR signals of existing state-of-art method the proposed method achieves enhanced results by means of babble and train noises 0dB, 5dB and 10 dB SNR signals. Also, better results are obtained in presence of airport and car noises for all

signals which are shown in Table 3 and 4. From the results it is clearly identified that noise detection is more efficient than the other existing approaches. The stationary noise is recognized easily whereas in some case the non-stationary noise like babble noise and train noise tracked with the help of soft decision based estimation.

Table 3. Comparative performance for airport noise

Input SNR	Parameter Name	OM-LSA [13]	SMPO [11]	MMSE-SPP [10]	BWT [16]	Wavelet Packet [11]	Proposed
0 dB	PESQ	2.13	1.98	2.13	2.10	2.22	2.35
	segSNR	3.34	4.71	4.45	1.07	5.15	5.36
5 dB	PESQ	2.78	2.58	2.72	2.41	2.72	2.7
	segSNR	2.10	4.12	3.67	2.71	4.91	4.86
10 dB	PESQ	2.88	2.80	2.79	2.65	2.86	2.98
	segSNR	1.10	3.71	2.96	2.82	4.30	4.36

Table 4. Comparative performance for car noise

Input SNR	Parameter Name	OM-LSA [13]	SMPO [11]	MMSE-SPP [10]	BWT [16]	Wavelet Packet [11]	Proposed
0 dB	PESQ	2.39	2.19	2.25	2.22	2.27	2.36
	segSNR	4.78	5.43	5.5	2.86	6.03	5.5
5 dB	PESQ	2.59	2.54	2.64	2.26	2.71	2.69
	segSNR	3.48	5.12	4.96	3.69	5.52	4.89
10 dB	PESQ	2.90	2.89	2.92	2.70	2.88	2.95
	segSNR	1.45	4.12	4.64	4.78	4.90	4.77

Table 5. Comparative performance for train noise

Input SNR	Parameter Name	OM-LSA [13]	SMPO [11]	MMSE-SPP [10]	BWT [16]	Wavelet Packet [11]	Proposed
0 dB	PESQ	2.15	2.17	2.07	2.08	2.26	2.29
	segSNR	4.85	3.84	4.77	3.45	4.90	5.50
5 dB	PESQ	2.51	2.22	2.48	2.28	2.53	2.59
	segSNR	4.5	2.43	4.28	3.89	4.06	4.79
10 dB	PESQ	2.75	2.67	2.76	2.59	2.89	2.94
	segSNR	4.06	1.07	3.71	3.5	3.64	4.47

The Figure 7, Figure 8, Figure 9, Figure 10 shows that comparative analysis representation of proposed approach in terms of PESQ for babble, airport, car and train noises for 0 dB, 5 dB and 10 dB signals, considering the five states of art techniques. The PESQ measurement and improvement in the Table 2, 3, 4 and 5 shows that proposed method develops the quality of the enhanced signal at a rate of 7.76% average improvement for 0 dB, 2.65% improvement for 5 dB and finally 3.78% improvement is marked at 10 dB SNR signals.

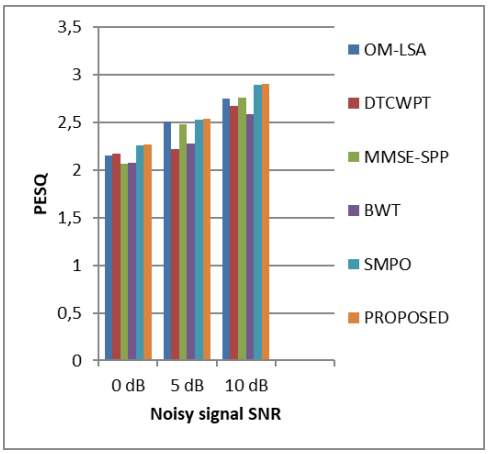

Figure 7. PESQ performance analysis for babble noise

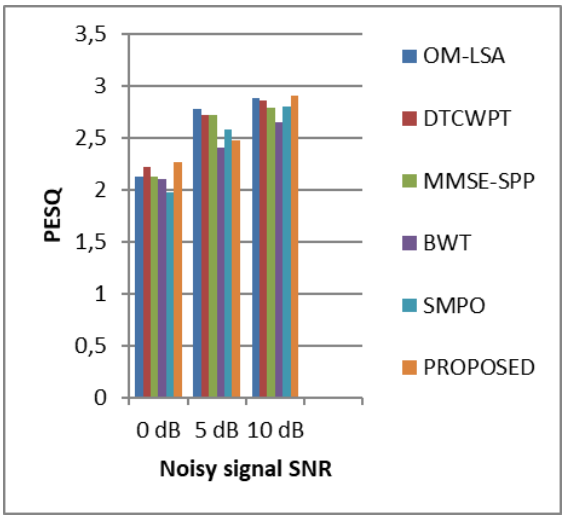

Figure 8. PESQ performance analysis for airport noise

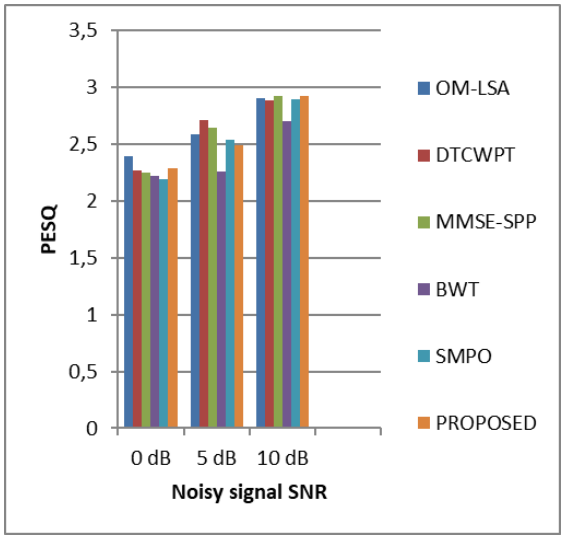

Figure 9. PESQ performance analysis for car noise

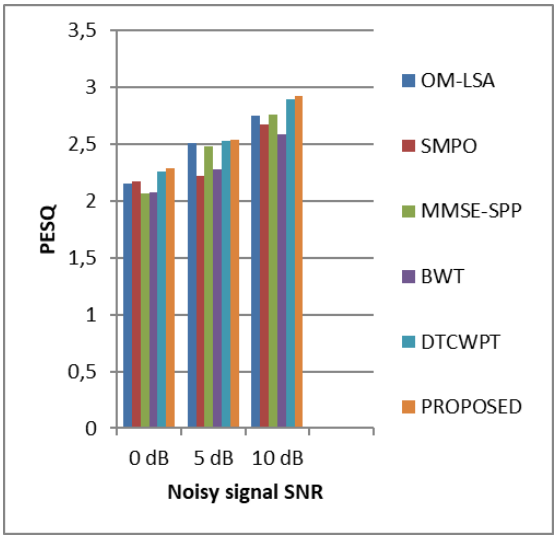

Figure 10. PESQ performance analysis for train noise

5. CONCLUSION

The objective intelligibility measured for various voiced and unvoiced signals is found to be better in the proposed method. Wiener filtering is implemented along with spectral coefficient, and probability distribution models, to improve the performance of speech quality in terms of PESQ and segSNR. While extracting the noise characteristics phase coefficients are preserved and reused after the reconstruction. It performs better in case of low SNR where noise power is calculated locally. Furthermore, SPP is also used to make a soft-decision in detecting the noise statistics in the absence of the speech and MMSE based noise tracking is used in the speech presence. This gives better noise estimation and improves the performance at high SNR. Finally, recursive smoothing is applied resulting in the efficient single-channel speech enhancement. The speech enhancement using adaptive filtering may reduce the cepstral smoothing, also echo cancelation and improve the results further. The results indicate that the proposed work is improving the speech quality at 0 dB, 5 dB and 10 dB and stands mostly good, and comparative in some particular cases. Speech intelligibility is also good according to the Mean Opinion Score (MOS) of the subjects. Also, it improves the quality of the lower energy concentrated signals efficiently when compared to the other state of art techniques.

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol.82, no.1, pp. 35-45, 1960.

[2] Boll, S. F, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics, Speech, and Signal Processing, pp. 113–120, 1979.

[3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in the *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*., pp. 208–211, 1979.

[4] K. T. Talale, and S.T. Gandhe, "Speech Compression using ADPCM," *IJCA - International Journal of Computer Applications In the Proceedings on International Conference in Computational Intelligence*, no. 8, 2012.

[5] L. R. Rabiner, and R. W. Schafer, "Digital Processing of Speech Signal," *The Journal of the Acoustical Society of America,* vol. 67, no. 4, pp.1406-1407, 1980.

[6] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *Journal of the Acoustical Society of America,* vol. 67, no. 1, pp. 318–326, 1980.

[7] T. Gerkmann and R. C. Hendriks, "Noise Power Estimation Based on The Probability of Speech Presence," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011.

[8] Park, Jeong-Sik, Gil-Jin Jang, and Ji-Hwan Kim. "Noise Reduction Scheme for Speech Recognition in Mobile Devices," *Information Technology Convergence*. Springer, Dordrecht, pp. 949-955, 2013.

[9] Zhao, Huan, et al. "A new soft masking method for speech enhancement in the frequency domain." *Elektronika ir Elektrotechnika*, vol. 20, no. 2, pp. 58-63, 2014.

[10] J. Jensen, R. C. Hendriks, and T. Gerkman, "DFT domain based single-microphone noise reduction for speech enhancement," *A survey of the state-of-the-art,* 2013.

[11] P. Sun, Jun Qin, "Speech enhancement via two-stage dual tree complex wavelet packet transform with a speech presence probability estimator", *The Journal of Acoustic Society of America*, vol. 141, no. 2, pp. 808-817, Jan 2017.

[12] Young-Seok Choi, "A New Subband Adaptive Filtering Algorithm for Sparse System Identification with Impulsive Noise", *Journal of Applied Mathematics*, vol. 2014.

[13] Tran, Tien Dung, Quoc Cuong Nguyen, and Dang Khoa Nguyen. "Speech enhancement using modified IMCRA and OMLSA methods," *International Conference on Communications and Electronics 2010*., IEEE, 2010.

[14] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transaction in Speech Audio Processing*, vol. 11, pp. 466-475, 2003.

[15] Zhao, Huan, et al. "A new soft masking method for speech enhancement in the frequency domain." *Elektronika ir Elektrotechnika*, vol. 20. no. 2, pp. 58-63, 2014.

[16] Sonu Bala, and Mohammad Arif, "Performance comparison of discrete transforms on speech compressed sensing," *In the Proceedings of IEEE International Conference on Computational Intelligence & Communication Technology*, pp.632-637, 2015.