Journal of Engineering Science and Technology
Vol. 12, No. 4 (2017) 972 - 986

PERFORMANCE ANALYSIS OF SPEECH SIGNAL ENHANCEMENT TECHNIQUES FOR NOISY TAMIL SPEECH RECOGNITION

VIMALA C.*, RADHA V.

Department of Computer Science, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore – 641043, Tamil Nadu, India
*Corresponding Author: vimalac.au@gmail.com

Abstract

Speech signal enhancement techniques have reached a considerable research attention because of its significant need in several signal processing applications. Various techniques have been developed for improving the speech signals in adverse conditions. In order to apply a good speech signal enhancement technique, an extensive comparison of the algorithms has always been necessary. Therefore, the performance evaluations of eight speech signal enhancement techniques are implemented and assessed based on various speech signal quality measures. In this paper, the Geometric Spectral Subtraction (GSS), Recursive Least Squares (RLS) Adaptive Filtering, Wavelet Filtering, Kalman Filtering, Ideal Binary Mask (IBM), Phase Spectrum Compensation (PSC), Minimum Mean Square Error estimator Magnitude Squared Spectrum incorporating SNR Uncertainty (MSS-MMSE-SPZC), and MMSE-MSS using SNR Uncertainty (MSS-MMSE-SPZC-SNRU) algorithms are implemented. These techniques are evaluated based on six objective speech quality measures and one subjective quality measure. Based on the experimental outcomes, the optimal speech signal enhancement technique which is suitable for all types of noisy conditions is exposed.

Keywords: Geometric spectral subtraction (GSS), RLS adaptive filtering, Wavelet filtering, Kalman filtering, Ideal binary mask (IBM), Phase spectrum compensation (PSC).

1. Introduction

In a real time environment, the speech signals are corrupted by several types of noise such as competing speakers, background noise, channel distortion and room reverberation, etc. The intelligibility and quality of a signal is severely degraded by these distortions [1]. Researchers have found out that, an error rate will increase up to 40% when speech signal enhancement techniques are not employed. However, an error rate will decrease from 7 to 13% when speech enhancement techniques are applied (for SNR 10dB). But, only 1% error rate is maintained by the human listener in noisy surroundings. Therefore, the speech signal has to be enhanced with Digital Signal Processing (DSP) tools, before it is stored, transmitted or processed. Hence, the speech signal enhancement technique plays a vital role and it is useful in many applications like telecommunications, enhancing the quality of old records, pre-processor for speech and speaker recognition and audio based information retrieval, etc. There are various types of speech signal enhancement techniques available, to enhance the noisy speech signal. They are briefly explained in this paper and its performances are analyzed based on both subjective and objective speech quality measures.

The paper is organized as follows. Section 2 gives details about the adopted speech signal enhancement techniques for this work. Section 3 discusses the performance evaluation metrics used for speech signal enhancement. Experimental results and the performance evaluations are presented in Section 4. Conclusion and future works are given in Section 5.

2. Adopted Speech Signal Enhancement Techniques

Speech signal enhancement techniques are mainly used as a pre-processor for noisy speech recognition applications. They can improve the intelligibility and quality of a speech signal, but it also sounds less annoying. Several techniques have been developed for this purpose, namely, Spectral Subtraction, Adaptive Filtering, Extended and Iterative Wiener filtering, Kalman filtering, Fuzzy algorithms, HMM based algorithms, and Signal subspace methods. All these techniques have their own merits and demerits. Based on the ability of the above algorithms, eight types of speech signal enhancement techniques are adopted for this research work and they are briefly explained below.

- Geometric Spectral Subtraction (GSS),
- RLS Adaptive Filtering,
- Wavelet Filtering,
- Kalman Filtering,
- Ideal Binary Mask (IBM),
- Phase Spectrum Compensation (PSC),
- Minimum Mean Square Error estimator Magnitude Squared Spectrum incorporating SNR Uncertainty (MSS-MMSE-SPZC), and
- MMSE-MSS using SNR Uncertainty (MSS-MMSE-SPZC-SNRU).

2.1. Geometric spectral subtraction (GSS)

Spectral Subtraction (SS) is the conventional technique which was initially proposed for reducing additive background noise [2]. This technique was found to be simple and cost effective. But, it significantly suffers from musical noise, therefore it has gone through many modifications later [1, 3].The performance evaluations of the six types of SS algorithms were implemented by Vimala and

Radha [4]. They are basic SS algorithm by Boll [5] and Berouti et al. [6], Nonlinear Spectral Subtraction (NSS), Multi Band Spectral Subtraction (MBSS), MMSE and Log Spectral MMSE.

All these algorithms were analyzed for the speech signals corrupted by white and babble noise and they were evaluated based on Signal-to-Noise Ratio (SNR) and Mean Squared Error (MSE) values. It was proved from the experimental outcomes that, the NSS algorithm works better for noise reduction when compared with the other algorithms involved. It is also observed from the experiments, as well as from the overall studies carried out by many researchers that, the spectral subtraction algorithm improves speech quality but not speech intelligibility [2]. Consequently, in this research work, the most recent improvement with SS using a geometric approach is considered for performance evaluation and it is explained below.

GSS for speech signal enhancement

GSS for speech signal enhancement is proposed by Lu and Loizou [7]. It is largely a deterministic approach, and represents the noisy speech spectrum in a high level surface, as the summation of clean and noisy signals. GSS addresses the two major shortcomings of SS, namely, musical noise, and invalid assumptions about the cross terms being zero. GSS provides the difference between phase spectrum of noisy and clean signal and it does not make any assumptions about the cross terms, being zero. Hence, it works better than a conventional spectral subtraction algorithm. The noise magnitude calculations assume that the first five frames are noise or silence. Therefore, the accurate estimate of the magnitude spectrum of the clean signal is obtained by discarding the noise magnitude spectrum. By using this spectrum, whether the clean signal can be recovered in the given noisy speech spectrum can be determined. This representation provides important information to the SS approach for achieving better noise reduction. The above GSS technique has been compared with the adaptive filtering algorithms and it is explained below.

2.2. RLS adaptive filtering

Speech signals are non-stationary in nature therefore non-adaptive filtering techniques may not be suitable for speech signal related applications. As a result, the adaptive filter became popular with the ability to operate in an unknown and changing environment. The adaptive filter does not carry any prior knowledge about the signals and they do not have constant filter coefficients [8]. In contrast to other filtering techniques, it has the ability to update the filter coefficients with respect to the signal conditions and new environment [9]. Moreover, it can suppress the noise without changing the originality of the signal.

RLS adaptive algorithm is a recursive implementation of the Wiener filter, which is used to find the difference between the desired and the actual signals. In RLS, 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. The RLS adaptive filter recursively computes the RLS estimate of the FIR filter coefficients [10]. The filter tap weight vector is updated using Eq. (1).

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

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

RLS Adaptive Algorithm

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

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

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

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

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

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

Table 1. Variables used in RLS algorithm.

Variable	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

where, λ^{-1} denotes the reciprocal of the exponential weighting factor. RLS algorithm performs at each instant an exact minimization of the sum of the squares of the desired signal *d(n)* and the estimation error *e(n)*. Therefore, output from the adaptive filter matches closely the desired signal *d(n)*. 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 [11]. 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.

Vimala and Radha [12] 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. It is observed from the experiments that, RLS technique provides faster convergence and smaller error, but it increases the complexity when compared with LMS and NLMS techniques. Likewise, the LMS and NLMS algorithms are very effective and simple to implement, but they are slow in processing the noisy signals [13]. It is proved from the experimental outcomes that, the RLS adaptive algorithm was found to be an optimal speech enhancement technique for noisy Tamil ASR. In order to make a wider performance comparison between the speech signal enhancement techniques, the wavelet filters are implemented subsequently.

2.3. Wavelet filtering

Wavelet transforms are widely used in various signal processing related tasks, namely, speech or speaker recognition, speech coding and speech signal enhancement. By using only a few wavelet coefficients, it is possible to obtain a good approximation about an original speech signal and the corrupted noisy signal. The wavelet transform for speech signal enhancement is given in the following Eq. (2).

$$Y_{j,k} = X_{j,k} + N_{j,k} \tag{2}$$

where, $Y_{j,k}$ represents the kth set of wavelet coefficients across the selected scale j, X represents the original signal and N represents the noisy signal. In wavelet transform, the larger coefficients are used to represent the energy of the signal, whereas the smaller coefficients are used to represent the energy of the noisy signal. By using the threshold methods, the discrimination between the signal and noise energy is calculated, thereby the possibility of separating the noise from the signal has been improved [14]. In this research work, Coiflet5, Daubechies 8, 10, 15, Haar, Symlets 5, 10, 15 are implemented for performance comparison with different types of decomposition levels 2, 3, 4, 5, 8 and 10. Among them, Daubechies 5 wavelets with 3rd level decomposition have obtained moderate results for the experiments. Since it involves simple threshold method, it cannot make an efficient discrimination between the speech and noise. Therefore, further attempts are made on using Kalman filtering and it is explained below.

2.4. Kalman filter

Kalman filter is an unbiased, time domain linear MMSE estimator, where the enhanced speech is recursively estimated on a sample-by-sample basis. Hence, the Kalman filter can be assumed to be a joint estimator for both phase and magnitude spectrum of speech [15]. Kalman filtering involves a mathematical operation that works based on a prediction and correction mechanism using LPC estimation of clean speech. Additionally, it predicts a new state from its previous estimation by adding a correction term proportional to the predicted error. Therefore, the error is statistically minimized. Moreover, it does not require all previous data to be kept in storage and it can be reprocessed every time a new measurement is taken. The algorithm of Kalman filter is given below.

For the experiments, Kalman filter has shown moderate results. Apart from Kalman filtering, the most recent approach which is popularly utilized for signal separation is applying Ideal Binary Mask (IBM). Based on the significance of

IBM technique, it is also involved in the performance comparison, and it is explained in the subsequent section.

Algorithm of Kalman Filtering

Step 1: Estimate the mean of a random sample $Z_1, Z_2, Z_3, \ldots. Z_N$,
Step 2: Refine the estimate after every new measurement, and
Recursive Solution
Step 3:
a) First measurement - Compute the estimate as $m_1 = Z_1$ store m_1 and discard Z_1.
b) Second measurement -Compute the estimate as a weighted sum of previous estimate and current measurement Z_2,
$m_2 = (m_1)/2 + (Z_2)/2$
Store m_2 and discard Z_2 and m_1,
c) Third measurement - Compute the estimate as a weighted sum of m_2 and Z_3,
$m_3 = 2/3\ (m_2) + (Z_3)/3$
Store m_3 and discard Z_3 and m_2, and
d) Estimate the weighted sum at the n^{th} stage $m_n = (n-1)/n * (m_{n-1}) + (1/n)\ Z_n$.

2.5. Ideal binary mask (IBM)

Human listeners are able to understand speech even when it is masked by one or more competing voices or distortions. But, the computer system cannot understand speech affected by the distortions. In such cases, the IBM technique has been recently demonstrated and has a large potential to improve the speech intelligibility in difficult listening conditions [16]. It includes, modelling of the human auditory scene analysis for evaluating the overall perception of auditory mixtures [17], and for improving the accuracy of an ASR system [18]. It has the ability of improving the intelligibility of a speech signal corrupted by different types of maskers for both Normal Hearing(NH) and Hearing Impaired(HI) people.

The Ideal Binary Masking technique is commonly applied to Time-Frequency (T-F) representation for increasing the speech intelligibility of the corrupted signals. T-F representation of signals makes it possible to utilize both the temporal and spectral properties of the speech signal. The goal of IBM technique is to segregate only the target signal by assigning the values of 0 and 1 by comparing the local SNR with each T-F unit against a threshold value. The speech segment whose value is assigned to 0 is eliminated and the speech segment whose value is assigned to 1 is allowed for further processing. It is defined in Eq. (3).

$$IBM(\tau.k) = \begin{cases} 1, if\ \dfrac{T(\tau,k)}{M(\tau.k)} > LC, \\ 0, otherwise \end{cases} \tag{3}$$

where $T(\tau,k)$ is the power of the target signal, $M(\tau, k)$ is the power of the masker signal, LC is a local SNR criterion, τ is the time index and k the frequency index. The threshold value used for binary masking are -3 and -10 for negative SNR dB levels (-5 dB and -10 dB). For positive SNR dB levels, (0dB, 5dB and 10 dB) the threshold value has been assigned as 2. In this research study, next to RLS filtering, the IBM method has produced better results. Subsequent attempts are made to implement Phase Spectrum Compensation (PSC) method for performance comparison.

2.6. Phase spectrum compensation(PSC)

Most of the speech enhancement techniques are based on Short Time Fourier Transform (STFT), which does not concentrate more about the phase spectrum. However, in speech enhancement, both noise and phase spectrum is important to improve the perceptual property of a signal. Recently, the PSC method has been proposed by Wojcicki et al. [19], which utilizes both; phase and noise spectra. In PSC method, the noise magnitude spectrum is recombined with the phase spectrum to produce a new complex spectrum. The estimated phase spectrum is then used for reconstructing the enhanced speech signal.

By using the new complex spectrum, the noise estimates are used to compensate the phase spectrum. The noise reduction is mainly concentrated on the low energy components of the modified complex spectrum, instead of suppressing the high energy components [20]. The compensated short time phase spectrum is computed using Eq. (4).

$$\wedge(\mathrm{n},\mathrm{k}) = \lambda\psi(\mathrm{k})\left|\hat{\mathrm{D}}(\mathrm{n},\mathrm{k})\right|, \tag{4}$$

where, λ is a real-valued empirically determined constant, $\Psi(k)$ is the anti-symmetry function and $\left|\hat{D}(n,k)\right|$ is an estimate of the short-time magnitude spectrum of noise. The time invariant anti-symmetry function is given in Eq. (5).

$$\psi(k) = \begin{cases} 1, if\ 0 < k/N < 0.5 \\ -1, if\ 0.5 < k/N < 1 \\ 0, otherwise \end{cases} \tag{5}$$

Next, the complex spectrum of the noisy speech is compensated by the additive real-valued frequency-dependent using Eq. (6).

$$X_{\wedge}(n,k) = X(n,k) + \wedge(n,k) \tag{6}$$

Finally, the compensated phase spectrum is obtained through Eq. (7).

$$\angle X_{\wedge}(n,k) = ARG\left[X_{\wedge}(n,k)\right], \tag{7}$$

where, ARG is the complex angle function. Here, the compensated phase spectrum does not represent the property of a true phase spectrum, i.e., real valued signal [18]. Therefore, it is recombined with the noisy magnitude spectrum to produce a modified complex spectrum given in Eq. (8).

$$\hat{S}_{\wedge}(n,k) = \left|X(n,k)\right| e^{j\angle X_{\wedge}(n,k)} \tag{8}$$

The resultant signal $\hat{S}_{\wedge}(n,k)$ is then converted into a time-domain representation which involves overlapping. For the experiments, the frame duration has been set to 32 (ms), frame shift is assigned to 4 (ms) and the lamba value of 3.74 has been used as a scale of compensation. This method has been compared with the above mentioned signal enhancement techniques and have shown reasonable results.

2.7. Minimum mean square error estimator magnitude squared spectrum

MMSE estimator is proposed by Ephraim and Malah [21], to produce an optimal Magnitude Squared Spectrum (MSS). It assumes the probability distributions of speech and noise Discrete Fourier Transform (DFT) coefficients without using a linear model. The MMSE estimators of the magnitude spectrum can perform well for various noisy conditions and provides better speech quality. Recent attempts have been made by Lu and Loizou [22] for providing the MSS for incorporating SNR Uncertainty. The authors have derived a gain function of the MAP estimator of the MSS, which works similar to Ideal Binary Masking technique. From the study, two important algorithms are implemented in this research work.

1) MSS-MMSE-SPZC: MMSE estimator MSS of incorporating SNR Uncertainty, and
2) MSS-MMSE-SPZC-SNRU: MMSE estimator of MSS using SNR Uncertainty.

These two algorithms can help to reduce the residual noise without altering the original speech signals. The above algorithms are implemented and their performances are evaluated with the adopted speech signal enhancement techniques. The subsequent section explains the metrics used for performance evaluation.

3. Performance Evaluation Metrics used for Speech Signal Enhancement

The performance evaluation of the adopted speech signal enhancement techniques are evaluated with four types of noise and five types of SNR dB levels, using various speech quality measures. The perception of a speech signal is usually measured in terms of its quality and intelligibility [23]. The quality is a subjective measure, which gives an individual opinion from the listeners about the enhanced speech signal. The intelligibility is an objective measure, which predicts the percentage of words that can be correctly identified by the listeners. In this research work, both objective and subjective speech quality measures are used to evaluate the adopted techniques. Six types of objective quality measures and one subjective quality measure are involved in this research work. They are briefly explained below.

3.1. Objective speech quality measures

Objective metrics are evaluated, based on the mathematical measures. It represents the signal quality, by comparing the original speech signals with the enhanced speech signals. The objective speech quality measures used in this work are listed below:

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

3.2. Subjective speech quality measure

Subjective quality evaluations are performed by involving a group of listeners to measure the quality of an enhanced signal. 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 research work, instead of using a regular MOS, the composite objective measure introduced by Yang Lu and Philipos, C. Loizou are implemented. The authors have derived new accurate measure 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 are used for both signal and background distortions (1= bad, 2=poor, 3= fair, 4= good and 5= excellent).

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. 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. The experimental results obtained by the adopted techniques and the performance evaluations are presented in the next section.

4. Experimental Results

In this paper, the experiments are carried out by using Tamil speech signals. Since, the noisy dataset is not available for Tamil language it is created artificially by adding noise from NOIZEUS database. 10 Tamil spoken words that are uttered in 10 different ways are used to create a database. The signals are corrupted by four types of noise (White, Babble, Mall and Car) and five types of SNR dB levels (-10dB, -5dB, 0dB, 5dB and 10dB). So, the total dataset size is 10*10*4*5=2000. Tables 2 to 5 illustrate the performance evaluation of the existing speech signal enhancement techniques for white, babble, mall and car noise respectively (corrupted by -10 dB,-5 dB, 0dB, 5 dB and 10 dB SNR) .

Based on the experimental results, it is observed that the ***RLS adaptive algorithm has performed exceptionally well for all type of noise types and SNR dB levels***. RLS has produced ***maximum PESQ, MOS, SegSNR and output SNR*** values when compared with the other algorithms involved. Moreover, it has provided significant performance in ***minimizing WSS and MSE values***. Also, the RLS technique has produced extensive results for negative SNR dB levels which are an added advantage. It was found to be an optimal speech signal enhancement technique for Tamil speech recognition. Next to RLS technique, the IBM and PSC methods have obtained reasonable results. The wavelet filtering, Kalman filtering, GSS and MMSE techniques have not provided comparable results based on the experimental results.

Table 2. Performance evaluations of speech signal enhancement techniques for white noise.

SNR dB Types	Metrics	RLS	GSS	MSS-MMSE SPZC	MMSE MSS-SPZC–SNRU	Kalman	Wavelet	IBM	PSC
-10 dB	**PESQ**	**3.48**	1.00	1.17	1.13	1.02	1.07	2.32	1.17
	LLR	**0.86**	2.34	2.25	2.11	2.08	2.26	0.42	2.41
	WSS	**9.41**	136.94	133.91	133.50	157.43	202.57	63.8	143.8
	SegSNR	**6.54**	-1.28	-3.18	-1.99	0.14	-1.27	-1.85	-0.11
	Output SNR	**4.73**	-1.58	-3.04	-1.98	0.15	-1.01	-1.98	0.16
	MSE	**0.14**	0.29	0.35	0.31	0.24	0.27	0.31	0.24
	MOS	**3.89**	0.25	0.45	0.49	0.14	-0.37	2.92	-0.43
-5 dB	**PESQ**	**3.71**	1.28	1.42	1.48	1.33	1.40	2.50	1.56
	LLR	**0.52**	1.83	1.79	1.76	1.64	2.23	0.46	1.92
	WSS	**4.92**	100.02	101.17	107.97	129.87	159.62	46.7	102.5
	SegSNR	**11.31**	-1.05	-2.68	-1.96	0.72	1.40	-1.82	1.66
	Output SNR	**8.83**	-1.11	-2.81	-2.23	0.78	1.55	-2.03	1.89
	MSE	**0.09**	0.28	0.34	0.31	0.22	0.20	0.31	0.20
	MOS	**4.28**	0.99	1.11	1.13	0.85	0.22	3.17	0.74
0 dB	**PESQ**	**3.86**	1.76	1.80	1.95	1.62	1.81	2.62	1.98
	LLR	**0.29**	1.46	1.46	1.49	1.43	2.22	0.45	1.51
	WSS	**1.85**	78.29	76.70	91.05	94.12	123.77	42.7	66.22
	SegSNR	**16.02**	-1.40	-2.28	-1.84	1.77	3.99	-1.86	3.25
	Output SNR	**13.54**	-1.56	-2.60	-2.19	1.80	3.80	-2.08	3.42
	MSE	**0.05**	0.29	0.33	0.31	0.20	0.16	0.31	0.16
	MOS	**4.54**	1.72	1.76	1.76	1.46	0.81	3.30	1.82
5 dB	**PESQ**	**4.00**	2.09	2.25	2.30	2.08	2.12	1.55	2.49
	LLR	**0.12**	1.16	1.18	1.27	1.22	2.17	0.77	1.19
	WSS	**0.64**	71.56	60.22	74.83	64.91	99.23	111.4	42.9
	SegSNR	**20.03**	-1.49	-2.09	-1.70	2.94	5.72	-0.92	4.17
	Output SNR	**17.84**	-1.67	-2.46	-2.04	2.81	5.06	-0.93	4.25
	MSE	**0.03**	0.29	0.32	0.31	0.18	0.14	0.27	0.15
	MOS	**4.75**	2.18	2.38	2.27	2.17	1.27	1.64	2.79
10 dB	**PESQ**	**4.28**	2.04	2.48	2.36	2.25	2.24	0.71	2.98
	LLR	**0.04**	0.93	0.97	1.10	1.03	2.13	1.87	0.93
	WSS	**0.27**	67.48	53.60	68.77	47.85	80.20	226.6	29.8
	SegSNR	**23.07**	-1.21	-2.01	-1.63	3.89	6.76	-0.06	4.61
	Output SNR	**21.19**	-1.31	-2.37	-1.94	3.58	5.85	-0.06	4.63
	MSE	**0.02**	0.28	0.32	0.30	0.16	0.12	0.24	0.14
	MOS	**4.82**	2.29	2.72	2.45	2.54	1.52	-0.86	3.60

Table 3. Performance evaluations of speech signal enhancement techniques for babble noise.

SNR dB Types	Metrics	RLS	GSS	MSS-MMSE SPZC	MMSE MSS-SPZC–SNRU	Kalman	Wavelet	IBM	PSC
-10 dB	PESQ	**3.50**	0.43	0.69	0.63	0.40	0.93	2.14	0.71
	LLR	**0.29**	1.62	1.62	1.68	2.08	2.17	0.37	1.81
	WSS	**12.09**	144.3	150.07	157.22	173.8	182.7	71.04	168.6
	SegSNR	**8.81**	-1.49	-3.53	-2.72	-0.54	-4.67	-1.66	-0.99
	Output SNR	**5.35**	-1.76	-3.53	-2.72	-0.34	-4.65	-1.78	-0.99
	MSE	**0.13**	0.30	0.37	0.33	0.25	0.42	0.30	0.27
	MOS	**4.18**	0.10	0.27	0.14	-0.48	-0.28	2.76	-0.52
-5 dB	PESQ	**3.71**	0.55	0.92	0.94	1.03	1.10	2.39	0.86
	LLR	**0.10**	1.33	1.35	1.40	1.70	2.17	0.40	1.47
	WSS	**5.89**	121.2	118.79	125.33	145.76	161.24	61.8	143.7
	SegSNR	**13.01**	-1.29	-2.82	-2.25	0.55	-1.71	-1.52	0.73
	Output SNR	**9.10**	-1.30	-2.88	-2.33	0.65	-1.74	-1.69	0.88
	MSE	**0.09**	0.28	0.34	0.32	0.23	0.30	0.30	0.22
	MOS	**4.49**	0.51	0.81	0.76	0.47	0.00	3.01	0.15
0 dB	PESQ	**3.96**	0.87	1.21	1.21	1.38	1.45	2.51	1.22
	LLR	**0.05**	1.02	1.11	1.20	1.39	2.21	0.40	1.19
	WSS	**2.85**	101.4	94.61	105.79	105.81	135.31	56.7	105.8
	SegSNR	**17.19**	-1.18	-2.24	-1.78	2.07	1.74	-1.52	2.15
	Output SNR	**13.51**	-1.16	-2.42	-1.91	2.12	1.58	-1.70	2.44
	MSE	**0.05**	0.28	0.32	0.30	0.19	0.20	0.30	0.18
	MOS	**4.74**	1.07	1.34	1.21	1.21	0.45	3.14	1.07
5 dB	PESQ	**4.18**	1.47	1.63	1.62	1.83	1.84	1.57	1.83
	LLR	**0.03**	0.79	0.90	1.04	1.13	2.17	0.81	0.92
	WSS	**1.22**	80.27	71.37	83.93	73.86	111.47	123.4	67.26
	SegSNR	**21.14**	-1.20	-2.04	-1.62	3.43	4.34	-1.05	3.42
	Output SNR	**17.82**	-1.24	-2.33	-1.84	3.34	3.92	-1.15	3.65
	MSE	**0.03**	0.28	0.32	0.30	0.17	0.15	0.28	0.16
	MOS	**4.93**	1.81	1.94	1.78	1.96	0.96	1.52	2.21
10 dB	PESQ	**4.33**	2.05	2.21	2.08	2.09	2.16	0.81	2.34
	LLR	**0.02**	0.73	0.78	0.95	0.97	2.17	1.58	0.78
	WSS	**0.44**	64.48	56.66	71.73	52.24	94.24	239.9	43.56
	SegSNR	**24.07**	-1.43	-2.02	-1.64	4.30	5.91	-0.05	4.29
	Output SNR	**21.23**	-1.49	-2.35	-1.91	4.02	5.13	-0.03	4.40
	MSE	**0.02**	0.29	0.32	0.30	0.15	0.13	0.24	0.15
	MOS	**4.96**	2.42	2.58	2.28	2.42	1.34	-0.66	3.02

Table 4. Performance evaluations of speech signal enhancement techniques for mall noise.

SNR dB Types	Metrics	RLS	GSS	MSS-MMSE SPZC	MMSE MSS-SPZC–SNRU	Kalman	Wavelet	IBM	PSC
-10 dB	PESQ	**3.26**	0.91	0.98	0.92	1.39	1.19	2.26	0.89
	LLR	**0.37**	1.97	1.89	1.95	2.08	2.18	0.36	2.13
	WSS	**14.23**	148.53	155.79	168.24	176.39	196.6	66.5	183.1
	SegSNR	**4.47**	-1.48	-3.67	-2.88	-0.59	-3.61	-2.04	-2.09
	Output SNR	**3.03**	-1.52	-3.80	-3.04	-0.60	-3.33	-2.15	-2.15
	MSE	**0.17**	0.29	0.38	0.34	0.26	0.36	0.31	0.31
	MOS	**3.93**	0.27	0.32	0.16	0.31	-0.17	2.90	-0.78
-5 dB	PESQ	**3.59**	1.18	1.16	1.08	1.23	1.32	2.61	1.13
	LLR	**0.22**	1.44	1.47	1.49	1.70	2.19	0.40	1.73
	WSS	**6.62**	126.87	121.63	129.64	137.01	176.0	51.74	124.2
	SegSNR	**11.34**	-1.17	-2.65	-2.05	0.62	0.25	-1.59	1.29
	Output SNR	**7.22**	-1.24	-2.73	-2.22	0.73	0.00	-1.76	1.27
	MSE	**0.11**	0.28	0.33	0.31	0.22	0.24	0.30	0.21
	MOS	**4.33**	0.92	0.92	0.80	0.69	0.07	3.27	0.33
0 dB	PESQ	**4.18**	1.36	1.34	1.28	1.53	1.58	2.63	1.28
	LLR	**0.09**	1.13	1.24	1.32	1.46	2.20	0.39	1.43
	WSS	**2.31**	89.49	89.35	99.84	101.57	129.6	43.2	104.7
	SegSNR	**17.89**	-1.57	-2.45	-2.06	1.57	2.43	-1.79	2.29
	Output SNR	**15.30**	-1.68	-2.79	-2.39	1.62	2.56	-2.03	2.42
	MSE	**0.04**	0.29	0.34	0.32	0.20	0.18	0.31	0.18
	MOS	**4.89**	1.48	1.41	1.25	1.33	0.60	3.36	0.91
5 dB	PESQ	**4.14**	1.90	1.95	1.98	1.91	2.00	1.55	2.07
	LLR	**0.03**	0.91	0.97	1.09	1.22	2.15	0.85	1.01
	WSS	**1.51**	82.90	68.70	84.41	69.11	108.7	125.5	56.96
	SegSNR	**20.21**	-1.39	-2.03	-1.64	3.39	5.21	-1.01	3.67
	Output SNR	**16.14**	-1.45	-2.34	-1.92	3.20	4.55	-1.08	3.69
	MSE	**0.04**	0.29	0.32	0.30	0.17	0.14	0.28	0.16
	MOS	**4.90**	2.08	2.19	2.04	2.00	1.12	1.45	2.45
10 dB	PESQ	**4.33**	2.09	2.25	2.27	2.22	2.22	0.70	2.57
	LLR	**0.02**	0.77	0.81	0.96	1.01	2.16	1.86	0.79
	WSS	**0.74**	70.11	56.58	74.44	49.24	90.2	231.5	40.37
	SegSNR	**23.35**	-1.17	-2.02	-1.66	4.31	6.36	-0.12	4.28
	Output SNR	**20.10**	-1.23	-2.36	-1.95	3.95	5.42	-0.07	4.27
	MSE	**0.02**	0.28	0.32	0.30	0.15	0.13	0.25	0.15
	MOS	**4.97**	2.39	2.59	2.41	2.53	1.42	-0.90	3.25

Table 5. Performance evaluations of speech signal enhancement techniques for car noise.

SNR dB Types	Metrics	RLS	GSS	MSS-MMSE SPZC	MMSE MSS-SPZC–SNRU	Kalman	Wavelet	IBM	PSC
-10 dB	PESQ	**3.61**	1.47	1.52	1.68	0.90	1.38	2.17	1.63
	LLR	**0.05**	0.74	0.75	0.83	1.59	1.97	0.38	0.72
	WSS	**8.08**	147.5	146.95	141.86	180.7	172.64	68.6	153.9
	SegSNR	**7.87**	-0.82	-2.99	-1.29	-5.95	-8.05	-1.59	-0.46
	Output SNR	**4.05**	-1.41	-2.85	-1.27	-6.03	-8.57	-1.75	-0.31
	MSE	**0.15**	0.29	0.34	0.28	0.49	0.65	0.30	0.25
	MOS	**4.41**	1.37	1.40	1.53	0.18	0.30	2.79	1.43
-5 dB	PESQ	**3.84**	1.88	2.09	2.03	1.44	1.81	2.46	2.33
	LLR	**0.04**	0.64	0.73	0.89	1.39	2.04	0.35	0.70
	WSS	**4.26**	100.8	106.31	101.38	135.4	148.10	51.1	92.25
	SegSNR	**15.02**	-0.97	-2.32	-1.38	-3.31	-4.87	-1.61	1.75
	Output SNR	**8.52**	-1.15	-2.44	-1.58	-3.26	-5.04	-1.78	1.92
	MSE	**0.09**	0.28	0.32	0.29	0.35	0.43	0.30	0.19
	MOS	**4.63**	2.07	2.15	2.06	1.06	0.77	3.19	2.65
0 dB	PESQ	**4.04**	2.04	2.34	2.24	2.10	2.11	2.57	2.80
	LLR	**0.02**	0.65	0.70	0.87	1.05	2.06	0.34	0.71
	WSS	**2.07**	75.29	75.72	77.38	90.59	123.33	46.6	63.56
	SegSNR	**20.28**	-1.05	-2.13	-1.68	0.12	-0.83	-1.61	3.37
	Output SNR	**13.29**	-1.15	-2.45	-2.05	0.20	-0.83	-1.78	3.55
	MSE	**0.05**	0.28	0.32	0.31	0.24	0.27	0.30	0.16
	MOS	**4.82**	2.38	2.59	2.41	2.11	1.17	3.33	3.32
5 dB	PESQ	**4.18**	2.06	2.48	2.30	2.54	2.33	1.29	3.00
	LLR	**0.02**	0.67	0.69	0.87	0.86	2.04	1.41	0.72
	WSS	**0.91**	59.90	62.07	73.81	60.81	102.83	130.8	45.69
	SegSNR	**23.29**	-1.21	-2.02	-1.72	2.92	2.73	-1.29	4.24
	Output SNR	**17.61**	-1.30	-2.40	-2.09	2.91	2.72	-1.46	4.36
	MSE	**0.03**	0.28	0.32	0.31	0.17	0.18	0.29	0.15
	MOS	**4.94**	2.49	2.80	2.48	2.79	1.51	0.75	3.65
10 dB	PESQ	**4.31**	2.17	2.53	2.35	2.51	2.43	0.38	3.14
	LLR	**0.02**	0.69	0.69	0.89	0.81	2.03	2.13	0.73
	WSS	**0.42**	55.65	53.68	68.36	44.88	87.37	221.1	34.31
	SegSNR	**24.99**	-1.53	-1.99	-1.72	4.26	5.19	-0.05	4.64
	Output SNR	**21.01**	-1.66	-2.38	-2.07	3.89	4.95	-0.02	4.70
	MSE	**0.02**	0.29	0.32	0.31	0.16	0.14	0.24	0.14
	MOS	**4.96**	2.60	2.90	2.55	2.91	1.70	-1.30	3.87

5. Conclusion and Future Work

In this paper, an extensive comparative study on various speech signal enhancement techniques have been analyzed. Eight types of techniques that are widely used for speech signal enhancement are involved in the work. These techniques were evaluated using six types of objective speech quality measures and one subjective speech quality measure. Based on the outcomes, it is proved that the RLS adaptive filtering was found to be better when compared with all the other adopted speech enhancement techniques in terms of both objective and subjective quality measures. One drawback of the RLS technique is different square root matrix and forgetting factor values need to be assigned for positive and negative SNR dB levels. Therefore, in future work, optimal parameters will be identified by fine tuning these values in order to provide better performance. The performance of the noisy speech recognition for Tamil language will also be addressed in future.

References

1. Jashvir Chhikara and Jagbir Singh (2012). Noise cancellation using adaptive algorithms. *International Journal of Modern Engineering Research (IJMER)*, 2(3), 792-795.
2. Fukane, A.R.; and Sahare, S.L. (2011). Different approaches of spectral subtraction method for enhancing the speech signal in noisy environments. *International Journal of Scientific & Engineering Research*, 2(5), 1-6.
3. Krishnamoorthy, P.; and Mahadeva Prasanna, S.R. (2008). Temporal and spectral processing methods for processing of degraded speech: A Review. *IETE Technical Review*, 26(2), 137-148.
4. Vimala, C.; and Radha, V. (2012). A family of spectral subtraction algorithms for Tamil speech enhancement. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(1), 397-401.
5. Boll, S.F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2), 113-120.
6. Berouti, M.; Schwartz, M.; and Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 208-211.
7. Lu, Y.; and Loizou, P.C. (2008). A geometric approach to spectral subtraction. *Speech Communication*, 50(6), 453-466.
8. Srinivasan, A. (2009). Adaptive echo noise elimination for speech enhancement of Tamil letter Zha. *International Journal of Engineering and Technology*, 1(3), 91-97.
9. Stasionis, L.; and Serackis, A. (2011). Selection of an optimal adaptive filter for speech signal noise cancellation using C6455 DSP. *Journal of Electronics and Electrical Engineering*, 9(115), 101-104.
10. Haykin, S. (1996). *Adaptive filter theory* (3rd ed). Englewood Cliffs, NJ: Prentice Hall.
11. Computer exercise 5 (2010). Adaptive Signal Processing. Retrieved August 15, 2015, from http://www.eit.lth.se/fileadmin/eit/courses/ett042/CE/CE5e.pdf.

12. Vimala, C.; and Radha, V. (2012). Optimal adaptive filtering technique for Tamil speech enhancement. *International Journal of Computer Applications (IJCA)*, 41(17), 23-29
13. Borisagar, K.R.; and Kulkarni, G.R. (2010). Simulation and comparative analysis of LMS and RLS algorithms using real time speech input signal. *Global Journal of Researches in Engineering*, 10(5), 44-47.
14. Johnson, M.T.; Yuan, X.; and Ren, Y. (2007). Speech signal enhancement through adaptive wavelet thresholding. *Speech Communication*, 49(2), 123-133.
15. Li, C.J. (2006). *Non-Gaussian, non-stationary, and nonlinear signal processing methods – with applications to speech processing and channel estimation*. Ph.D. Thesis, Aarlborg University, Denmark.
16. Wang, D.L.; and Brown, G.J. (2006). Computational auditory scene analysis: principles, algorithms, and applications. *Wiley- IEEE Press*, Hoboken, NJ.
17. Brungart, D.S.; Chang, P.S.; Simpson, P.D.; and Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of Acoustical Society of America*, 120(6), 4007-4018.
18. Cooke, M.; Green, P.; Josifovski, L.; and Vizinho, A. (2001). Robust automatic speech recognition with missing and uncertain acoustic data. *Speech Communication*, 34(3), 267-285.
19. Wojcicki, K.; Milacic, M.; Stark, A.; Lyons, J.G.; and Paliwal, K. (2008). Exploiting conjugate symmetry of the short−time Fourier spectrum for speech enhancement. *IEEE Signal Processing Letters*, 15, 461-464.
20. Stark, A.P.; Wojcicki, K.K.; Lyons, J.G.; and Paliwal, K.K. (2008). Noise driven short time phase spectrum compensation procedure for speech enhancement. *Proceedings of INTERSPEECH2008*, Brisbane, Australia, 549-552.
21. Ephraim, Y.; and Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustic, Speech, and Signal Processing*, 32(6), 1109-1121.
22. Lu, Y.; and Loizou, P.C. (2011). Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5), 1123-1137.
23. Ma, J.; Hu, Y.; and Loizou, P.C. (2009). Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of Acoustic Society of America*, 125(5), 3387-3405.