

# An Efficient Reference Free Adaptive Learning Process for Speech Enhancement Applications

Girika Jyoshna<sup>1,\*</sup>, Md. Zia Ur Rahman<sup>1</sup> and L. Koteswararao<sup>2</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, KL University, Vaddeswaram, Guntur, Andhra Pradesh, India

<sup>2</sup>Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, KL University, Hyderabad, Telangana, India

\*Corresponding Author: Girika Jyoshna. Email: girikajyoshna1986@gmail.com

Received: 12 May 2021; Accepted: 21 June 2021

**Abstract:** In issues like hearing impairment, speech therapy and hearing aids play a major role in reducing the impairment. Removal of noise signals from speech signals is a key task in hearing aids as well as in speech therapy. During the transmission of speech signals, several noise components contaminate the actual speech components. This paper addresses a new adaptive speech enhancement (ASE) method based on a modified version of singular spectrum analysis (MSSA). The MSSA generates a reference signal for ASE and makes the ASE free from feeding reference component. The MSSA adopts three key steps for generating the reference from the contaminated speech only. These are decomposition, grouping and reconstruction. The generated reference is taken as a reference for variable size adaptive learning algorithms. In this work two categories of adaptive learning algorithms are used. They are step variable adaptive learning (SVAL) algorithm and time variable step size adaptive learning (TVAL). Further, sign regressor function is applied to adaptive learning algorithms to reduce the computational complexity of the proposed adaptive learning algorithms. The performance measures of the proposed schemes are calculated in terms of signal to noise ratio improvement (SNRI), excess mean square error (EMSE) and misadjustment (MSD). For cockpit noise these measures are found to be 29.2850, -27.6060 and 0.0758 dB respectively during the experiments using SVAL algorithm. By considering the reduced number of multiplications the sign regressor version of SVAL based ASE method is found to be better than the counter parts.

**Keywords:** Adaptive algorithm; speech enhancement; singular spectrum analysis; reference free noise canceller; variable step size

## 1 Introduction

Adaptive speech enhancement plays a key role in speech therapy as well as in hearing aids. In speech therapy the presence of noise degrades the quality of the speech and hence the therapy may not be faithful. Further, ambiguities due to noise leads improper identification of



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

audio signals and lead to unsatisfactory results in speech recognition. Hence, facilitating a high-resolution speech signal is highly desirable in speech therapy applications. In the literature several contributions are made in this aspect. In Laufer et al. [1], Bayesian hierarchical model approach is used for speech enhancement. This method relies on the Gaussian prior of the speech signal and gamma hyper. Spectra mapping [2] is used in throat microphone to avoid acoustic noises. Low band and high band spectral structures between acoustic microphone and throat microphone speech are considered in this analysis. A dynamic filter structure is proposed for speech enhancement, it performs instantaneously based on noisy speech signal. Enhancement of speech signal is done by using two gain functions, i.e., estimation of noise power spectrum and estimation of noisy speech spectrum signal. Variances of this power estimation signal degrades quality of speech signal, so the proposed method estimates noisy power spectrum signals based on adaptive time segmentation. Further demonstration of adaptive segmentation was done based on decision-based speech enhancement and maximum likelihood are proposed in [3,4]. Convolution neural networks (CNN) are used for speech enhancement by utilizing data from various modules. Audio visual deep CNN [5] network is proposed for speech enhancement by utilizing audio and visual stream frame networks. The reconstruction of audio and visual signals is done at the output of the system. Compressive sensing [6] is used for speech enhancement and is performed in frequency domain. A new adaptive beam-forming algorithm was suggested to avoid noisy car environments to improve speech recognition. It contains speech and noise signals in constrained sections as noise adaptive beam-former and speech adaptive beam-former. Later performance investigation was done using delay and sum beam-forming to decrease the word error rate in identification of speech signal. Also auto regression-based gaussian distribution and Laplacian distribution is used for enhancing speech signal are described in [7–12]. The proposed Laplacian prior estimators minimize unnecessary noise signals in desired speech signals. By deriving minimum mean square error and reducing distortion in speech signals with the use of linear bilateral Laplacian gain estimator and non-linear bilateral gain estimators. Main aim of speech enhancement algorithm [13] is to improve intelligibility and quality of noisy speech signal by using spectral or temporal modifications. Maximum speech signal enhancement is done using magnitude spectrum. In Mowlae et al. [14], magnitude and phase spectra are changed in order to enhance noisy speech signals using multi-level speech enhancement algorithm. Double spectrum consists of modulation transforms further pitch synchronization is used for enhancement of single channel speech signal [15]. Frame wise context modeling is also considered with adaptive filter coefficient frame wise tracking, so that robust performance is obtained in non-stationary noisy condition of speech signals [16]. For speech enhancement, semi supervised multi channels with non-negative matrix factorization [17–19] is used to reduce noisy signals along with constraint variants called as independent low matrix analysis. Unsupervised speech enhancement low power spectral densities are considered in Ming et al. [20]. Various adaptive learning algorithms are presented in [21–25]. Adaptive low rank matrix decomposition is often used for signal enhancement and enhanced efficiency in terms of speech quality. In these contributions mainly two aspects are not addressed. They are reference generation from the noisy speech signal and reduction of computational complexity of the speech enhancement algorithm.

In order to address this limitation, in this manuscript both the aspects of reference generation from the noisy speech and reduction of computational complexity are considered. These two are key entities for the development of system on chip realizations. A modified singular spectrum analysis (MSSA) with modified grouping step is used for reference generation from the noisy speech signal. This is fed to ASE module, which is driven by an adaptive learning algorithm. In our experiments we have used SVAL and TVAL methods. These algorithms are combined with

sign regressor function. This function minimizes the computational complexity of the algorithm by an amount equal to the tap length of the filter. The methodology of MSSA for reference generation and adaptive learning algorithms for speech enhancement process are discussed in Section 2. The experimental results are illustrated and presented in Section 3.

## 2 Hybrid Adaptive Algorithm for Speech Enhancement

In real time applications, removal of noise from the desired signal is considered as inverse problem. It means artifacts are removed from contaminated signal. In this work, it is proposed to eliminate the noise signal from the contaminated speech waves. For this a modified SSA (MSSA) algorithm is used for the generation of reference signal. In an adaptive speech enhancer reference signal is a key element. The reference generated from MSSA is fed to the adaptive learning algorithm. The learning algorithm trains the weight coefficients and trains the weight coefficients, in such a way the reference and contamination in the actual speech signal are correlates with each other. Then both the correlated components get cancel with each other. The MSSA extracts an embedded feature matrix from the speech signal. This matrix is a delayed version and the matrix elements are grouped by using k-means algorithm. Then for each cluster group, Eigen values and eigen vectors are computed using singular value decomposition. For estimating noise components in speech signal, minimum description length concept is considered. It will give dimension length of each eigen vector for estimating speech signal. But to estimate dimension length of each eigen vector, magnitude difference between eigen values is maintained for representing enhanced speech signal and noise signal. As Magnitude of noisy speech signal is high, MSSA has better performance in removing noisy speech signal efficiently. These four steps are followed in MSSA [26]: they are Embedding, Decomposition, Grouping and Reconstruction. Let us consider contaminated speech signal and it is represented as

$$r(i) = s(i) + n(i) \quad (1)$$

where 's' is a desired signal, 'n' is noise component signal, 'i' is number of samples. In first step, sampled data vector of single channel 'i',  $r = [r(1), r(2), \dots, r(I)]$  maps to multivariate matrix R as

$$R = \begin{bmatrix} r(1) & r(2) \dots & r(T) \\ r(2) & r(3) \dots & r(T+1) \\ \vdots & \vdots & \vdots \\ r(K) & r(K+1) & r(I) \end{bmatrix} \quad (2)$$

where 'K' is window length  $T = I - K + 1$ , window length 'K' is selected based on criteria  $K > f_s/f$ , here f and  $f_s$  are signal frequency and sample frequency of a signal of interest respectively. Trajectory matrices of desired speech signal are  $s(i)$  and  $n(i)$  respectively, then measured signal trajectory matrix  $r(i) = s(i) + n(i)$  is represented as  $R = S + N$ , where matrix 'N' is estimated by using trajectory matrix 'R'. In next step, singular value decomposition is performed on trajectory matrix  $R = V \sum W$ , where V and W are left and right orthogonal matrices respectively and its column elements contains eigen vectors of matrix D,  $\sum$  is a rectangular diagonal matrix, its elements are eigen values with squared roots. Golub and Reinsh algorithms generally perform singular value decomposition on rectangular matrix with  $K \times T$  dimension and it involves  $o(K^2T + KT^2 + T^3)$ . By using eigen value decomposition and covariance matrix, singular value decomposition is performed on trajectory matrix as  $H = BB^T$ , then the covariance matrix 'H',

eigen values and eigen vectors are represented as  $\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_K$  and  $\sigma_1, \sigma_2, \dots, \sigma_K \geq 0$ , so that right orthogonal matrix  $W$  of  $t^{\text{th}}$  eigen vector is represented as

$$w_k = B^T \frac{\sigma_t}{\sqrt{\bar{\delta}_t}} \quad (3)$$

where,  $t = 1, 2, \dots, K$ . Measured signal  $R$  for  $t^{\text{th}}$  component trajectory matrix is expressed as

$$B_t = \sqrt{\bar{\delta}_t} \sigma_t \sigma_t^T \quad (4)$$

By substituting Eq. (3) in Eq. (4),  $t^{\text{th}}$  trajectory matrix  $B_t$  is represents as

$$B_t = \sigma_t \sigma_t^T B \quad (5)$$

$\sigma_t \sigma_t^T$  term in Eq. (5) forms a  $t^{\text{th}}$  component subspace element in single vector  $B$ , then Eq. (2) trajectory matrix is decomposing into  $K$  elements so that  $R = R_1 + R_2 + \dots + R_K$  was obtained. In grouping step, basic SSAs trajectory matrix  $B_t$  with elements  $t = \{1, 2, \dots, K\}$  splits into 'J' groups, here J value is considered as two because in measuring speech signal enhanced speech signal is obtained from combination of desired signal and artifact signal. But in general, SSA grouping step is performed based on eigen value magnitudes of trajectory matrix. For example, initially in grouping trajectory matrices high energy signal magnitudes with large own values are considered, then corresponding trajectory matrices with recognized arguments are added to get a trajectory matrix of high energy signal. For trajectory matrix automatic grouping and minimum description length is considered to get desired signal. In case of noisy speech signal reduction this grouping criterion does not work well as artifact signal varies with time signal. So, a new grouping criterion is considered, based on eigen vectors local mobility. For given signal, eigen vectors are represented with frequency components then in grouping step local mobility of every eigen vector is exploited. To define  $t^{\text{th}}$  eigen vector local mobility  $\sigma_t = [\sigma_t(1), \sigma_t(2), \dots, \sigma_t(K)]$ , defined difference signal  $d(i)$  as  $d(i) = \sigma_t(i) - \sigma_t(i-1), i = 1, 2, \dots, K$ . Then for eigen vector  $\sigma_t$ , local mobility  $m_k$  is represented as

$$m_k = \frac{\mathcal{F}_1}{\mathcal{F}_0} \quad (6)$$

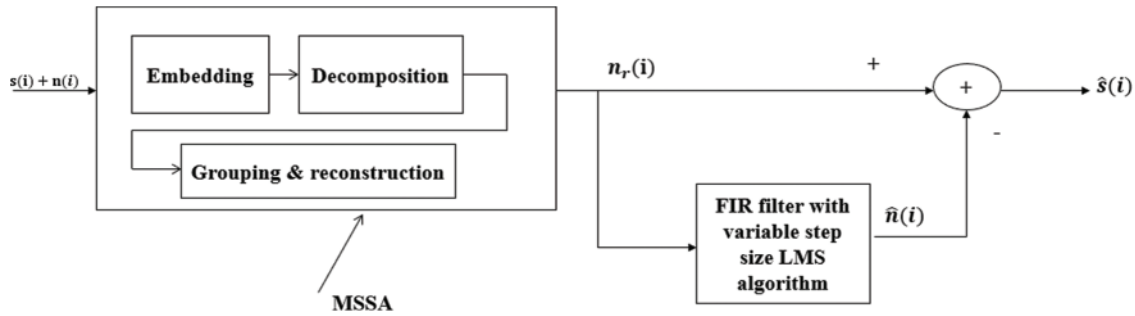
where  $\mathcal{F}_1 = \sqrt{\frac{\sum_{i=1}^{K-1} d(i)}{K-1}}$ ,  $\mathcal{F}_0 = \sqrt{\frac{\sum_{i=1}^K \sigma_t(i)}{K}}$ , here  $\mathcal{F}_1$  represents average of difference signal and it increases as  $\sigma_t$  frequency increases. To determine threshold value, investigated local mobility for sinusoidal frequency then its maximum frequency component and local mobility value is also fixed. By projecting data matrix  $R$ , estimated trajectory matrix with eigen vectors  $\sigma_t$  is obtained as  $\hat{N} = \sigma_1 \sigma_1^T R$ . Finally, reconstruction step for MSSA is obtained by estimating trajectory matrix with signal of interest as  $\hat{N}$  and it maps to single channel signal. Now for example, trajectory

matrix  $\hat{a}_{tg}$  with ‘t’ rows and ‘g’ columns then reconstruction step for single channel noisy signal  $\hat{n}(i)$  is mathematically expressed from estimated trajectory matrix  $\hat{N}$  as

$$\hat{n}(i) = \begin{cases} \frac{1}{i} \sum_{t=1}^i \hat{n}_{t,i-t+1} & \text{for } 1 \geq i < K \\ \frac{1}{K} \sum_{t=1}^K \hat{n}_{t,i-t+1} & \text{for } K \geq i \leq T \\ \frac{1}{I-i+1} \sum_{t=i-T+1}^{I-T+1} \hat{n}_{t,i-t+1} & \text{for } T < i \leq I \end{cases} \quad (7)$$

**Step Variable Size Adaptive Learning Algorithm with Modified Singular Spectrum Analysis Based Reference Generation**

Once the reference signal is generated by the MSSA based decomposition it is fed to the noise canceller. The learning algorithm trains the weight coefficients, such that the reference and noise component present in the actual speech correlates with each other and there by cancel with each other [27]. The block diagram of proposed MSSA based adaptive speech enhancer is shown in Fig. 1. Based on steepest descent algorithm, least mean square (LMS) algorithm weight update equation is estimated to get error free speech signal.



**Figure 1:** Block diagram of adaptive noise canceller using MSSA based reference generation

Output of FIR filter is

$$r(i) = a^T(i)c(i) \quad (8)$$

Error  $n(i)$  is defined as difference between desired response and actual response

$$n(i) = s(i) - r(i) \quad (9)$$

Then the weight update equation becomes

$$a(i+1) = a(i) + \omega n(i)c(i) \quad (10)$$

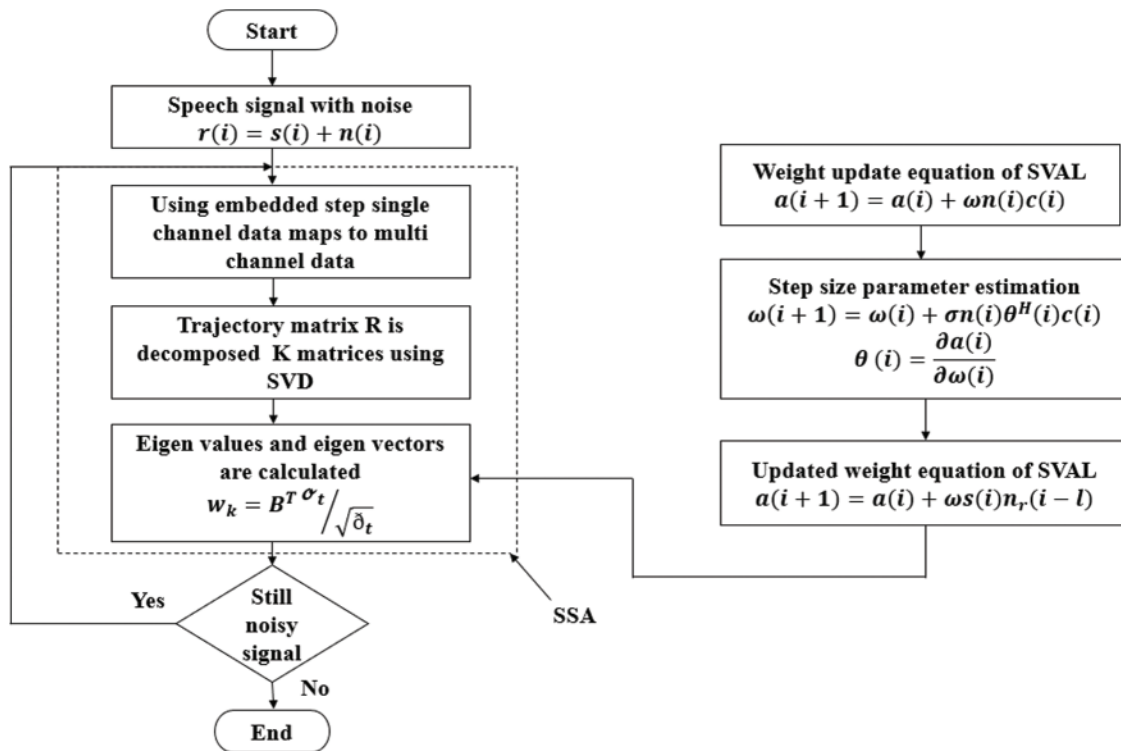
where  $n(i)$  is error in adaptive filter,  $s(i)$  is desired output,  $c(i)$  is filter input and  $\omega$  is step size for updating weight vector.

Normalized adaptive algorithm weight update equation is given as

$$a(i+1) = a(i) + \frac{\omega}{K + c^2(i)} \omega n(i) c(i) \quad (11)$$

Here 'K' is constant

In non-stationary environment, estimating variability and convergence time are main parameters, they are controlled by minimum cost function and vary with time and step size parameter. If the step size is smaller, a lot of iterations are required and there will be less residual noise. Hence, an optimized step value has to be chosen [27]. Similar to constant step size adaptive learning algorithm, in Ram et al. [28] an adaptive step size algorithm is proposed. Flowchart for speech enhancement process using SVAL algorithm with MSSA based reference generation is shown in Fig. 2.



**Figure 2:** Flowchart of adaptive speech enhancement using modified singular spectrum analysis-based reference generation

For this a Step size  $\omega$  is updated with the following equation and it is expressed as

$$\omega(i+1) = \omega(i) + \sigma n(i) \theta^H(i) c(i) \quad (12)$$

$\theta^H$  is a gradient vector and it is defined as ratio of partial derivative of weight update vector at each sample with respect to step size at each sample.

$$\theta^H(i) = \frac{\partial a(i)}{\partial \omega(i)} \quad (13)$$

Here  $\sigma$  is a small positive constant and it will control step size parameter updating. In practice, LMS is exponentially convergent, whereas sign algorithms are linearly convergence. Therefore, signed algorithms makes convergence slower when initial weight updating equation is far from optimum [29]. For minimizing computational complexity, the variable step algorithms are combined with signum function. The resultant mathematical recursions are written as,

$$a(i+1) = a(i) + \omega n(i) \text{sign}c(i) \quad (14)$$

$$a(i+1) = a(i) + \frac{\omega}{K + c^2(i)} \omega n(i) \text{sign}c(i) \quad (15)$$

Finally, exact speech signal  $\hat{s}(i)$  is extracted by subtracting estimated noisy signal from the calculated speech signal  $r(i)$  and it is expressed as

$$\hat{s}(i) = r(i) - \hat{n}(i) = s(i) + n(i) - \hat{n}(i) \quad (16)$$

$$a(i+1) = a(i) + \omega s(i) n_r(i-l) \quad (17)$$

Here  $n_r$  noise reference signal, it is obtained by passing through reference generator, 'J' is filter order with length  $l = 0, 1, 2, \dots, J$ .

### 3 Experimental Results and Analysis

In our experiments, noise cancellation from speech signals is done using modified singular spectrum analysis with adaptive step size algorithm. Singular spectrum analysis is a technique to extract the reference signal from the contaminated speech signal. As a result, the enhancement process does not require any prior knowledge on types of speech and noise signal. In order to circumvent these issues adaptive algorithm is recommended for speech enhancement along with singular spectrum analysis. It uses prior information about noise and speech types, then it will map to functions of clear speech and noise type features. For this entire process, the reference signal generated by the MSSA is taken in to consideration. By the proposed method, speech intelligibility is enhanced when it trains to specific scenarios. Performance of MSSA based adaptive algorithms is investigated in terms of intelligibility and objective measures. General SSA method shows improvement in objective measures, but it has failed in intelligibility test of speech signals across SNR regions. The MSSA with adaptive step size learning algorithm can overcome this drawback. Auto regressor coefficients of speech and noise signals parameters are considerably less when compared to the proposed algorithm. If weight parameters are trained smaller then it is possible to train these parameters for enhancement process. This results improvement in single channel intelligibility test signals. Diverse type of real noises is considered in our experiment's namely: cockpit noise, elevator noise, random noise and high voltage murmuring. As described in Section 2, the enhancement process is carried in two steps. First one is reference generation using MSSA and noise removal using adaptive learning algorithm.

The computational complexity is reduced in the proposed algorithm by applying sign regressor function to weight update equation. For the comparison of performance measures, adaptive noise cancellation due to LMS, TVAL, SRTVAL, SVAL and SRSVAL algorithms are considered. Performance measures are calculated in terms of signal to noise ratio improvement (SNRI), mis-adjustment (MSD) and excess mean square error (EMSE). The speech enhancement experiments are performed for ten times and average values are tabulated in [Tabs. 1–3](#). Simulations are carried using MATLAB tool, window size of adaptive filter is considered as ten, step size is 0.01. Elimination of the noise process initially done by adding additive Gaussian noise, then simulations of

polluted speech signals are performed by the proposed procedure, with zero mean and variance of 0.02 in white Gaussian noise. Real noises and synthetic noises are considered in the experiments. Random noise was shaped in accordance with threshold value estimations using spectral masking. The five speech samples labeled as Wave-1, Wave-2, Wave-3, Wave-4, Wave-5 and are contaminated with different types of noises. By using estimated thresholds and adaptive learning process, noise components are removed. The enhancement results of wave 1 contaminated with cockpit noise are shown in Fig. 3, due to space constraint only one signal is shown, the performance measures of all signals are tabulated. After achieving constant impulse response output, impulse response is considered for the next half of the samples and the results are observed. On observation, the before and after results in variations of impulse response feedback path exhibit faster convergence for proposed MSSA based SRSVAL algorithm as sign regressor function is used in comparison to MSSA-SVAL algorithm. Comparisons of various performance measures are also graphically shown in Figs. 4–6. The key benefit of the proposed implementation is that the reference produced from the noisy speech signal itself is the MSSA mechanism. The adaptive algorithm-related sign regressor operation minimizes the number of multiplications required for the filtering operation to be performed. The SVAL algorithm filters the speech signal with better convergence and filtering ability. It's been observed in our work that the proposed MSSA based sign regressor version of SVAL is found to be a better candidate for speech enhancement applications and suitable for immediate applications like mobile communications, speech, hearing aids and noise cancelers in defense, space applications.

**Table 1:** SNRI (dB) comparison of various algorithms in speech signal enhancement process

| Sl. No. | Noise type                   | Sample         | LMS           | TVAL           | SRTVAL         | SVAL           | SRSVAL         |
|---------|------------------------------|----------------|---------------|----------------|----------------|----------------|----------------|
| 1.      | Cockpit noise                | Wave-1         | <b>9.7763</b> | 25.0063        | 23.1543        | 29.0338        | 26.8487        |
|         |                              | Wave-2         | 9.8754        | 25.3824        | 23.5371        | 29.4270        | 26.3564        |
|         |                              | Wave-3         | 9.4367        | 25.3722        | 23.8944        | 29.1031        | 26.0918        |
|         |                              | Wave-4         | 9.8754        | 25.2598        | 23.4842        | 29.5335        | 26.9735        |
|         |                              | Wave-5         | 9.5431        | 25.7682        | 23.4735        | 29.3276        | 26.8373        |
|         |                              | <b>Average</b> | <b>9.7013</b> | <b>25.3577</b> | <b>23.5087</b> | <b>29.2850</b> | <b>26.6215</b> |
| 2.      | Crane noise                  | Wave-1         | <b>8.2598</b> | 24.8275        | 22.7436        | 28.6454        | 27.9953        |
|         |                              | Wave-2         | 8.6543        | 24.4124        | 22.4735        | 28.3345        | 27.3563        |
|         |                              | Wave-3         | 8.4321        | 24.3523        | 22.7391        | 28.4376        | 27.3720        |
|         |                              | Wave-4         | 8.2309        | 24.6841        | 22.3627        | 28.4957        | 27.8268        |
|         |                              | Wave-5         | 8.6520        | 24.7605        | 22.8635        | 28.4534        | 27.5472        |
|         |                              | <b>Average</b> | <b>8.4452</b> | <b>24.6073</b> | <b>22.6364</b> | <b>28.4733</b> | <b>27.6195</b> |
| 3.      | High voltage murmuring noise | Wave-1         | <b>7.7605</b> | 22.9652        | 21.9163        | 25.6786        | 23.1735        |
|         |                              | Wave-2         | 7.4532        | 22.4395        | 21.3854        | 25.4653        | 23.9334        |
|         |                              | Wave-3         | 7.6521        | 22.9008        | 21.5635        | 25.0345        | 23.5735        |
|         |                              | Wave-4         | 7.6532        | 22.2137        | 21.4893        | 25.7454        | 23.6562        |
|         |                              | Wave-5         | 7.6530        | 22.0559        | 21.4636        | 25.1554        | 23.4967        |
|         |                              | <b>Average</b> | <b>7.6344</b> | <b>22.5150</b> | <b>21.5636</b> | <b>25.4158</b> | <b>23.5666</b> |

(Continued)



**Table 1:** Continued

| Sl. No. | Noise type         | Sample         | LMS            | TVAL           | SRTVAL         | SVAL           | SRSVAL         |
|---------|--------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 4.      | Battle field noise | Wave-1         | <b>9.6511</b>  | 24.6511        | 22.1036        | 27.1634        | 25.2325        |
|         |                    | Wave-2         | 9.2873         | 24.3438        | 22.8452        | 27.5365        | 25.4734        |
|         |                    | Wave-3         | 9.5432         | 24.7617        | 22.8735        | 27.4628        | 25.0634        |
|         |                    | Wave-4         | 9.2098         | 24.3782        | 22.0332        | 27.3453        | 25.3469        |
|         |                    | Wave-5         | 9.3462         | 24.4083        | 22.6281        | 27.5259        | 25.1769        |
|         |                    | <b>Average</b> | <b>9.4075</b>  | <b>24.5086</b> | <b>22.4967</b> | <b>27.4067</b> | <b>25.2586</b> |
| 5.      | Random noise       | Wave-1         | <b>11.3722</b> | 23.7486        | 22.8767        | 26.5067        | 24.5186        |
|         |                    | Wave-2         | 11.7632        | 23.4492        | 22.4633        | 26.5454        | 24.3574        |
|         |                    | Wave-3         | 11.3298        | 24.4643        | 22.3878        | 26.4534        | 24.2443        |
|         |                    | Wave-4         | 11.3762        | 24.4673        | 22.4484        | 26.6353        | 24.0763        |
|         |                    | Wave-5         | 11.3872        | 24.5063        | 22.5836        | 26.1542        | 24.8635        |
|         |                    | <b>Average</b> | <b>11.4457</b> | <b>24.1271</b> | <b>22.5519</b> | <b>26.4590</b> | <b>24.4120</b> |

**Table 2:** EMSE (dB) comparison of various methods in speech signal enhancement process

| Sl. No. | Noise type                   | Sample         | LMS             | TVAL            | SRTVAL          | SVAL            | SRSVAL          |
|---------|------------------------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1.      | Cockpit noise                | Wave-1         | -12.829         | -24.9537        | -22.4732        | -27.8237        | -25.2841        |
|         |                              | Wave-2         | -12.098         | -24.1484        | -22.3782        | -27.2846        | -25.1879        |
|         |                              | Wave-3         | -12.876         | -24.9373        | -22.5365        | -27.6438        | -25.2937        |
|         |                              | Wave-4         | -12.459         | -24.4937        | -22.6036        | -27.8453        | -25.4739        |
|         |                              | Wave-5         | -12.986         | -24.8372        | -22.3395        | -27.4329        | -25.3367        |
|         |                              | <b>Average</b> | <b>-12.6496</b> | <b>-24.6740</b> | <b>-22.4662</b> | <b>-27.6060</b> | <b>-25.3152</b> |
| 2.      | Crane noise                  | Wave-1         | -10.988         | -22.8362        | -21.9465        | -26.8573        | -24.3692        |
|         |                              | Wave-2         | -10.543         | -22.0913        | -21.4201        | -26.0235        | -24.1052        |
|         |                              | Wave-3         | -10.763         | -22.2373        | -21.8452        | -26.7541        | -24.6383        |
|         |                              | Wave-4         | -10.963         | -22.1383        | -21.7733        | -26.3753        | -24.4273        |
|         |                              | Wave-5         | -10.724         | -22.3737        | -21.0387        | -26.8653        | -24.6172        |
|         |                              | <b>Average</b> | <b>-10.7962</b> | <b>-22.3353</b> | <b>-21.6047</b> | <b>-26.5751</b> | <b>-24.4314</b> |
| 3.      | High voltage murmuring noise | Wave-1         | -9.984          | -23.1839        | -25.7309        | -28.2479        | -27.3367        |
|         |                              | Wave-2         | -9.254          | -23.4844        | -25.4737        | -28.2254        | -27.5963        |
|         |                              | Wave-3         | -9.987          | -23.4635        | -25.4932        | -28.4394        | -27.1243        |
|         |                              | Wave-4         | -9.239          | -23.6338        | -25.3978        | -28.1302        | -27.4639        |
|         |                              | Wave-5         | -9.654          | -23.0358        | -25.4262        | -28.4274        | -27.4163        |
|         |                              | <b>Average</b> | <b>-9.6236</b>  | <b>-23.3602</b> | <b>-25.5043</b> | <b>-28.2940</b> | <b>-27.3875</b> |
| 4.      | Battle field noise           | Wave-1         | -10.8287        | -24.2832        | -21.6273        | -29.4926        | -27.9732        |
|         |                              | Wave-2         | -10.2371        | -24.4383        | -21.4214        | -29.3849        | -27.3074        |
|         |                              | Wave-3         | -10.3565        | -24.4603        | -21.0353        | -29.1947        | -27.4239        |
|         |                              | Wave-4         | -10.0989        | -24.3257        | -21.3426        | -29.1456        | -27.7547        |
|         |                              | Wave-5         | -10.9870        | -24.8173        | -21.6367        | -29.5748        | -27.4834        |
|         |                              | <b>Average</b> | <b>-10.5016</b> | <b>-24.4694</b> | <b>-21.4126</b> | <b>-29.3585</b> | <b>-27.5885</b> |

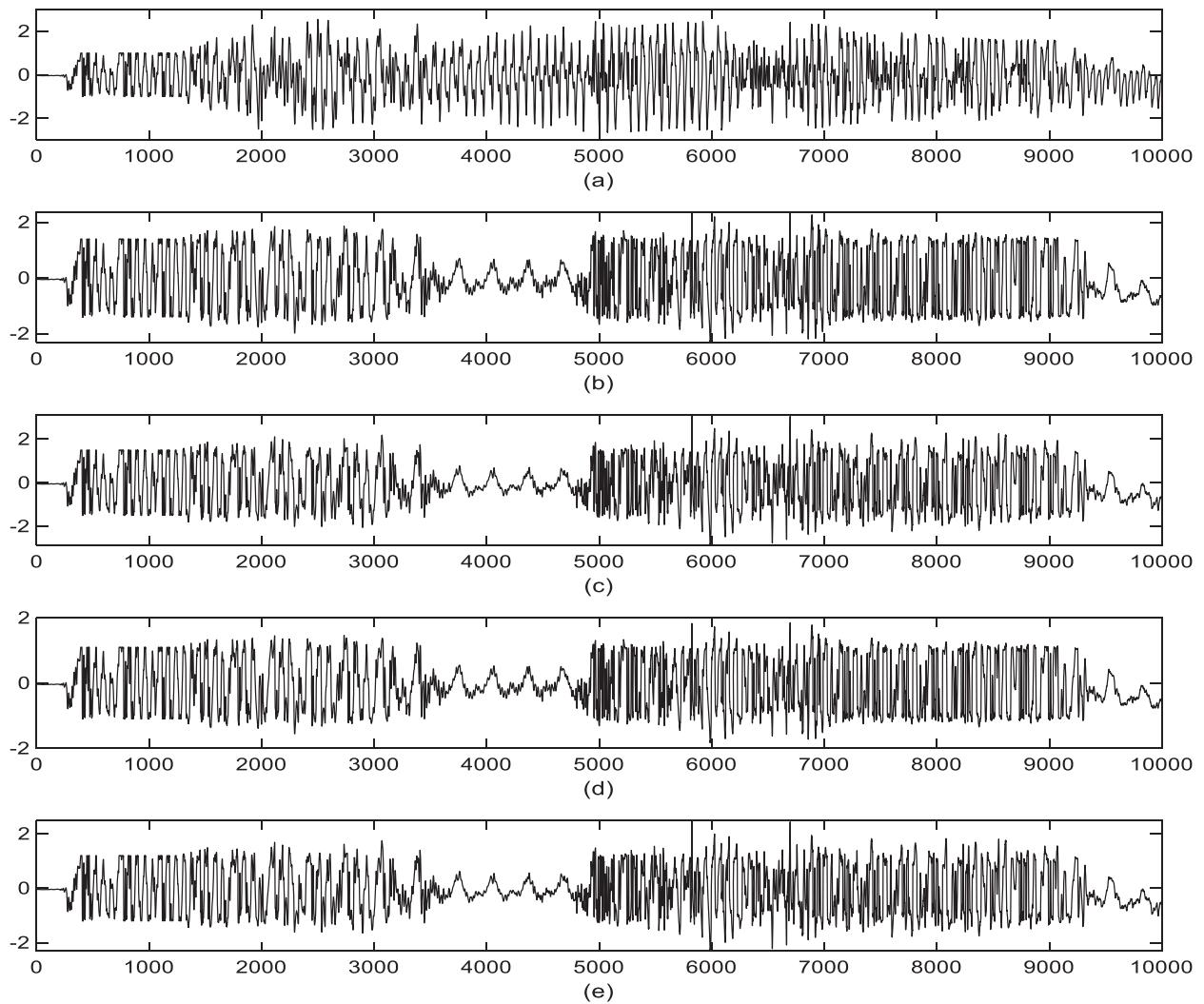
(Continued)

**Table 2:** Continued

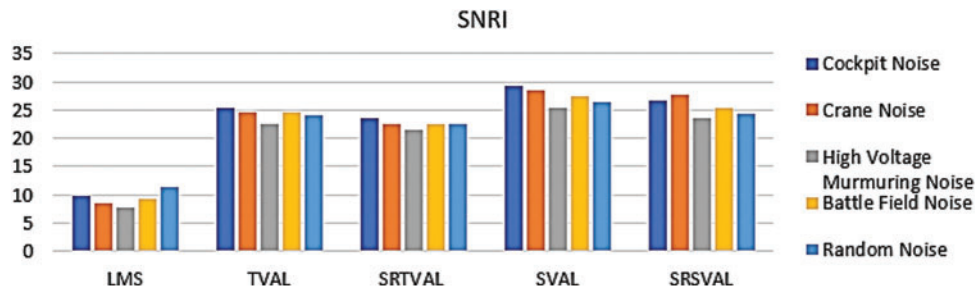
| Sl. No. | Noise type   | Sample         | LMS             | TVAL            | SRTVAL          | SVAL            | SRSVAL          |
|---------|--------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 5.      | Random noise | Wave-1         | -12.2407        | -23.9363        | -22.6285        | -29.9362        | -24.8834        |
|         |              | Wave-2         | -12.9870        | -23.7603        | -22.4516        | -29.6529        | -24.6539        |
|         |              | Wave-3         | -12.8757        | -23.6037        | -22.4683        | -29.2074        | -24.3964        |
|         |              | Wave-4         | -12.9854        | -23.1249        | -22.7538        | -29.0902        | -24.3035        |
|         |              | Wave-5         | -12.7643        | -23.7085        | -22.0668        | -29.4912        | -24.9804        |
|         |              | <b>Average</b> | <b>-12.7706</b> | <b>-23.6267</b> | <b>-22.4738</b> | <b>-29.4755</b> | <b>-24.6435</b> |

**Table 3:** MSD comparison of various methods in speech signal enhancement process

| Sl. No. | Noise type                   | Sample         | LMS           | TVAL          | SRTVAL        | SVAL          | SRSVAL        |
|---------|------------------------------|----------------|---------------|---------------|---------------|---------------|---------------|
| 1.      | Cockpit noise                | Wave-1         | <b>0.1352</b> | 0.0821        | 0.0920        | 0.0742        | 0.0641        |
|         |                              | Wave-2         | 0.1398        | 0.0857        | 0.0935        | 0.0783        | 0.0611        |
|         |                              | Wave-3         | 0.1340        | 0.0824        | 0.0948        | 0.0781        | 0.0629        |
|         |                              | Wave-4         | 0.1362        | 0.0807        | 0.0982        | 0.0750        | 0.0678        |
|         |                              | Wave-5         | 0.1387        | 0.0896        | 0.0923        | 0.0735        | 0.0646        |
|         |                              | <b>Average</b> | <b>0.1367</b> | <b>0.0841</b> | <b>0.0941</b> | <b>0.0758</b> | <b>0.0641</b> |
| 2.      | Crane noise                  | Wave-1         | <b>0.2264</b> | 0.2703        | 0.2503        | 0.2251        | 0.2021        |
|         |                              | Wave-2         | 0.2237        | 0.2700        | 0.2559        | 0.2287        | 0.2027        |
|         |                              | Wave-3         | 0.2209        | 0.2761        | 0.2558        | 0.2248        | 0.2009        |
|         |                              | Wave-4         | 0.2298        | 0.2756        | 0.2529        | 0.2264        | 0.2079        |
|         |                              | Wave-5         | 0.2283        | 0.2787        | 0.2539        | 0.2461        | 0.2027        |
|         |                              | <b>Average</b> | <b>0.2258</b> | <b>0.2741</b> | <b>0.2537</b> | <b>0.2302</b> | <b>0.2032</b> |
| 3.      | High voltage murmuring noise | Wave-1         | <b>0.2930</b> | 0.2361        | 0.2417        | 0.2110        | 0.2247        |
|         |                              | Wave-2         | 0.2987        | 0.2358        | 0.2465        | 0.2120        | 0.2257        |
|         |                              | Wave-3         | 0.2974        | 0.2396        | 0.2476        | 0.2107        | 0.2202        |
|         |                              | Wave-4         | 0.2983        | 0.2322        | 0.2470        | 0.2135        | 0.2236        |
|         |                              | Wave-5         | 0.2973        | 0.2343        | 0.2464        | 0.2160        | 0.2291        |
|         |                              | <b>Average</b> | <b>0.2969</b> | <b>0.2356</b> | <b>0.2458</b> | <b>0.2126</b> | <b>0.2246</b> |
| 4.      | Battle field noise           | Wave-1         | <b>0.1590</b> | 0.0851        | 0.1478        | 0.0695        | 0.0967        |
|         |                              | Wave-2         | 0.1563        | 0.0814        | 0.1456        | 0.0688        | 0.0978        |
|         |                              | Wave-3         | 0.1582        | 0.0830        | 0.1433        | 0.0697        | 0.0994        |
|         |                              | Wave-4         | 0.1598        | 0.0881        | 0.1430        | 0.0674        | 0.0972        |
|         |                              | Wave-5         | 0.1509        | 0.0866        | 0.1423        | 0.0624        | 0.0902        |
|         |                              | <b>Average</b> | <b>0.1568</b> | <b>0.0848</b> | <b>0.1444</b> | <b>0.0675</b> | <b>0.0962</b> |
| 5.      | Random noise                 | Wave-1         | <b>0.1092</b> | 0.0750        | 0.0947        | 0.0684        | 0.0813        |
|         |                              | Wave-2         | 0.1063        | 0.0749        | 0.0997        | 0.0686        | 0.0898        |
|         |                              | Wave-3         | 0.1084        | 0.0797        | 0.0994        | 0.0665        | 0.0877        |
|         |                              | Wave-4         | 0.1093        | 0.0718        | 0.0906        | 0.0606        | 0.0877        |
|         |                              | Wave-5         | 0.1072        | 0.0739        | 0.0980        | 0.0663        | 0.0835        |
|         |                              | <b>Average</b> | <b>0.1080</b> | <b>0.0750</b> | <b>0.0964</b> | <b>0.0660</b> | <b>0.0860</b> |



**Figure 3:** Speech enhancement using modified SSA based reference free ANC, (a). Speech signal contaminated with cockpit noise corresponds to the wave “adaptive noise cancellation”, (b). Enhanced signal with TVLAL ANC, (c). Enhanced signal with SRTVAL ANC, (d). Enhanced signal with SVAL ANC, (e). Enhanced signal with SRSVAL ANC. (X-axis number of samples, Y-axis amplitude of the signal)



**Figure 4:** Performance measure comparison of SNRI (dB) for various adaptive algorithms

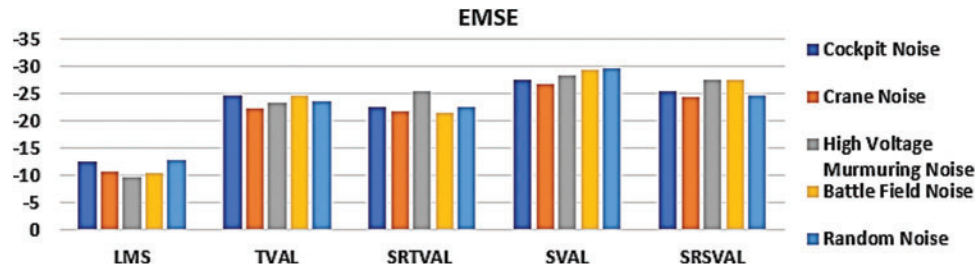


Figure 5: Performance measure comparison of EMSE (dB) for various adaptive algorithm

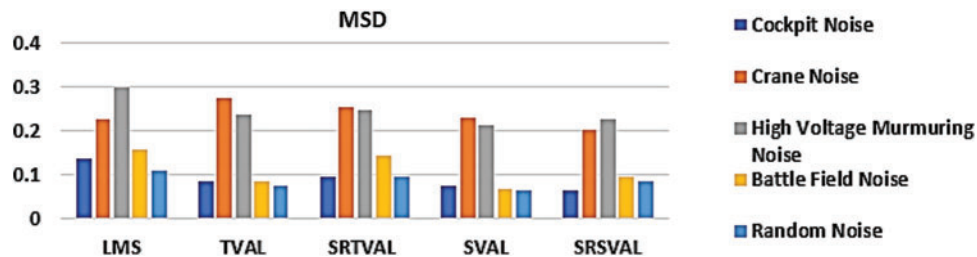


Figure 6: Performance measure comparison of MSD (no units) for various adaptive algorithms

#### 4 Conclusion

In general, speech signals are contaminated with the background noise and results ambiguities in speech recognition. In order to avoid those noises, a modified SSA based variable step size driven adaptive learning algorithms are proposed. The new grouping technique improves the performance of MSSA in the process of reference generation. The step variable adaptive learning process eliminates the noise components very effectively. The combination of sign regressor function reduces the number of multiplications involved in the process of noise cancellation. The performance measures are calculated, averaged for ten experiments and are tabulated in Tabs. 1–3. Several real time noises like cockpit noise, crane noise, high voltage murmuring noise, battle field noise and random noise are considered in the experiments. Between TVAL and SVAL, the performance of SVAL is found to be better than the counterpart. By considering the performance measures like SNRI, EMSE, MSD and computational complexity, it is found that sign regressor version of SVAL is better than the other learning methods. Hence, SRSVAL based adaptive speech enhancement unit is well suited for real time realization as system on chip or lab on chip.

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

#### References

- [1] Y. Laufer and S. Gannot, "A Bayesian hierarchical model for speech enhancement with time-varying audio channel," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 225–239, 2019.
- [2] C. Zheng, J. Yang, X. Zhang, M. Sun and T. Cao, "A spectra-based equalization-generation combined framework for throat microphone speech enhancement," *IEEE Access*, vol. 6, pp. 71455–71463, 2018.

- [3] M. P. A. Jeeva, T. Nagarajan and P. Vijayalakshmi, "Adaptive multi-band filter structure-based far-end speech enhancement," *IET Signal Processing*, vol. 14, no. 5, pp. 288–299, 2020.
- [4] R. C. Hendriks, R. Heusdens and J. Jensen, "Adaptive time segmentation for improved speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2064–2074, 2006.
- [5] J. Hou, S. Wang, Y. Lai, Y. Tsao, H. Chang *et al.*, "Audio-visual speech enhancement using multimodal deep convolutional neural networks," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 117–128, 2018.
- [6] J. Wang, Y. Lee, C. Lin, S. Wang, C. Shih *et al.*, "Compressive sensing-based speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2122–2131, 2016.
- [7] R. Martinek, R. Jaros, J. Baros, L. Danys, A. Kawala-Sterniuk *et al.*, "Noise reduction in industry based on virtual instrumentation," *Computers, Materials & Continua*, vol. 69, no. 1, pp. 1073–1096, 2021.
- [8] O. M. A. Zaid, M. A. Tawfeek and S. Alanazi, "Applying and comparison of chaotic-based permutation algorithms for audio encryption," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3161–3176, 2021.
- [9] Y. Li and S. Kang, "Deep neural network-based linear predictive parameter estimations for speech enhancement," *IET Signal Processing*, vol. 11, no. 4, pp. 469–476, 2017.
- [10] A. Aroudi, H. Veisi and H. Sameti, "Hidden Markov model-based speech enhancement using multivariate Laplace and Gaussian distributions," *IET Signal Processing*, vol. 9, no. 2, pp. 177–185, 2015.
- [11] H. Taherian, W. Zhong-Qiu, J. Chang and W. DeLiang, "Robust speaker recognition based on single-channel and multi-channel speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1293–1302, 2020.
- [12] B. M. Mahmmod, A. R. Ramli, S. H. Abdulhussian, S. A. R. Al-Haddad and W. A. Jassim, "Low-distortion MMSE speech enhancement estimator based on Laplacian prior," *IEEE Access*, vol. 5, pp. 9866–9881, 2017.
- [13] T. Lavanya, T. Nagarajan and P. Vijayalakshmi, "Multi-level single-channel speech enhancement using a unified framework for estimating magnitude and phase spectra," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1315–1327, 2020.
- [14] P. Mowlae, M. Blass and W. B. Kleijn, "New results in modulation-domain single-channel speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 11, pp. 2125–2137, 2017.
- [15] P. C. Bhattacharya, N. Tansangiumvisai and A. Worapishet, "Performance improvement of adaptive wavelet thresholding for speech enhancement using generalized Gaussian priors and frame-wise context modeling," *IEEE Access*, vol. 8, pp. 168361–168380, 2020.
- [16] G. Shi, P. Aarabi and H. Jiang, "Phase-based dual-microphone speech enhancement using a prior speech model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 109–118, 2007.
- [17] K. Sekiguchi, Y. Bando, A. A. Nugraha, K. Yoshii and T. Kawahara, "Semi-supervised multichannel speech enhancement with a deep speech prior," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2197–2212, 2019.
- [18] C. Li, T. Jiang, S. Wu and J. Xie, "Single-channel speech enhancement based on adaptive low-rank matrix decomposition," *IEEE Access*, vol. 8, pp. 37066–37076, 2020.
- [19] Y. Bando, K. Itoyama, M. Konyo, S. Tadokoro, K. Nakadai *et al.*, "Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 215–230, 2018.
- [20] J. Ming and D. Crookes, "Speech enhancement based on full-sentence correlation and clean speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 3, pp. 531–543, 2017.

- [21] M. N. Salman, P. Trinatha Rao and M. Z. Ur Rahman, "Adaptive noise cancellers for cardiac signal enhancement for IOT based health care systems," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 10, pp. 2206–2213, 2017.
- [22] A. Sulthana and M. Z. Ur Rahman, "Efficient adaptive noise cancellation techniques in an IOT enabled telecardiology system," *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 2, pp. 74–78, 2018.
- [23] P. Srinivasareddy and M. Z. Ur Rahman, "Novel simplified logarithmic adaptive exon prediction for DNA analysis," vol. 10, pp. 1422–1432, 2018.
- [24] M. N. Salman, P. Trinatha Rao and M. Z. U. Rahman, "Novel logarithmic reference free adaptive signal enhancers for ECG analysis of wireless cardiac care monitoring systems," *IEEE Access*, vol. 6, pp. 46382–46395, 2018.
- [25] P. Srinivasareddy, M. Z. Ur Rahman and S. Y. Fathima, "Cloud-based adaptive exon prediction for DNA analysis," *Healthcare Technology Letters*, vol. 5, no. 1, pp. 25–30, 2018.
- [26] M. Ajay Kumar and S. Rafi Ahamed, "Motion artifact removal from single channel electroencephalogram signals using singular spectrum analysis," *Biomedical Signal Processing and Control*, vol. 30, no. 6, pp. 79–85, 2016.
- [27] S. S. Haykin, *Adaptive Filter Theory*. Eaglewood Cliffs, NJ, USA: Prentice-Hall, 2007.
- [28] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla and K. A. Reddy, "A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1445–1457, 2012.
- [29] A. Sulthana, M. Z. U. Rahman and S. S. Mirza, "An efficient Kalman noise canceller for cardiac signal analysis in modern telecardiology systems," *IEEE Access*, vol. 6, pp. 34616–34630, 2018.