Article

An Efficient Multistage Approach for Blind Source Separation of Noisy Convolutive Speech Mixture

Junaid Bahadar Khan [1], Tariqullah Jan [1], Ruhul Amin Khalil [1], Nasir Saeed [2,*] and Muhannad Almutiry [3]

[1] J. B. K., T. J. and R. A. K. are with the Department of Electrical Engineering, University of Engineering and Technology, Peshawar 25120, Pakistan e-mail: jbkh08@gmail.com; tariqullahjan@uetpeshawar.edu.pk; ruhulamin@uetpeshawar.edu.pk.

[2] N. S. is with the Department of Electrical Engineering, King Abdullah University of Science and Technology, Thuwal, Makkah 23955, Saudi Arabia e-mail: mr.nasir.saeed@ieee.org.

[3] M. A. is with the Department of Electrical Engineering, Northern Border University, Arar 73222, Saudi Arabia e-mail: muhannad.almutiry@nbu.edu.sa.

[*] Correspondance: mr.nasir.saeed@ieee.org.

Abstract: This paper proposes a novel efficient multistage algorithm to extract source speech signals from a noisy convolutive mixture. The proposed approach comprises of two stages named Blind Source Separation (BSS) and De-noising. A hybrid source prior model separates the source signals from the noisy reverberant mixture in the BSS stage. Moreover, we model the low and high-energy components by generalized multivariate Gaussian and super-Gaussian models, respectively. We use Minimum Mean Square Error (MMSE) to reduce noise in the noisy convolutive mixture signal in the de-noising stage. Furthermore, two proposed models investigate the performance gain. In the first model, the speech signal is separated from the observed noisy convolutive mixture in the BSS stage, followed by suppression of noise in the estimated source signals in the de-noising module. In the second approach, the noise is reduced using the MMSE filtering technique in the received noisy convolutive mixture at the de-noising stage, followed by separation of source signals from the de-noised reverberant mixture at the BSS stage. We evaluate the performance of the proposed scheme in terms of signal-to-distortion ratio (SDR) with respect to other well-known multistage BSS methods. The results show the superior performance of the proposed algorithm over the other state-of-the-art methods.

Keywords: Blind Source Separation (BSS), Minimum Mean Square Error (MMSE), convolutive mixture, source Prior, generalized Gaussian distribution

1. Introduction

In a noisy real-time environment, the performance efficiency of Blind Source Separation (BSS) applications is degraded by background noise and interfering signals. The classical methods used for speech enhancement have reached their saturation level in terms of enhancement and performance. The estimation of the desired source signal from a mixture with noise, especially for non-stationary noisy conditions, is a bottleneck for these techniques. Therefore, the BSS applications require a solution that can suppress the noise according to the nature of the environment.

The speech signal enhancement problem is well studied in the past decades. Different solutions are provided to enhance the intelligibility and quality of the speech signals and improve the performance of the BSS systems. The classical techniques overcome this problem by using adaptive techniques like minimum means square error (MMSE) [1–7]. The MMSE adjusts itself according to the observed convolutive mixture. Another solution uses statistical models which accurately diagonalize the second-order statistical properties of the noisy reverberant mixture. This approach uses an auto-correlation covariance matrix and its one-sample delayed matrix, forming two positive definite symmetry matrices. Then, exploit the matrices' diagonalization accurately by computing generalized singular value decomposition (GSVD) using the tangent algorithm [8].

The BSS methods extract the desired source speech signal from the convolutive mixture in the presence of noise. The main advantage of BSS methods is to separate the targeted source speech signal from the reverberant mixture without prior knowledge of the mixing process nor the number of source signals. Some of the popular BSS approaches are Independent Component Analysis (ICA), and FASTICA which extract the source speech signals in a noisy reverberant environment. First, the ICA speech enhancement method de-noises the noisy reverberant mixture, followed by the FASTICA algorithm to separate the de-noised estimated speech signal from the observed convolutive mixture [9]. Additionally, in an undermined scenario, ICA is combined with a speaker recognition system (SRS) to extract the desired targeted speech signal [10].

The main problem encountered by BSS techniques is the permutation and scaling ambiguities after the speech separation process. Therefore, in [11], the authors proposed a solution that can easily recognize the desired source speech signal in a noisy environment by looking at the speaker's face. An audiovisual coherence is used to estimate the speech signals using statistical methods where statistical tools model the audio and visual information in the frequency domain (FD).

Furthermore, in multiple audio sources with multiple microphones scenarios, the performance of the BSS separation process is improved by using the BSS output to generate the wiener filter coefficients and apply them to the desired speech signals [12]. Moreover, adaptive filtering with BSS can also reduce the noise, leading to speech enhancement and noise reduction. Forward Blind Source Separation (FBSS) combined with Simplified Fast transversal filters (SFTF) method results in adoption gain from forwarding prediction [13]. Nevertheless, adaptive filtering methods face problems while canceling or suppressing the acoustic noise. This issue is tackled using the Modified Predator-prey particle swarm optimization (MPPPSO) approach. It also solves the problem of steady-state error of PPPSO for non-stationary inputs and large filter length [14].The acoustic noise can also be suppressed by introducing variable step size in a two-channel sub-band forward algorithm (2CSF) that improves the convergence speed and overcomes fixed step size problem in the traditional 2CSF method [15]. Another approach using variable step size is adaptive blind source separation through a two-channel forward-backward structure based on the normalized least-mean square (NLMS) method that uses variable step size for steady-state condition [16]. The estimated source signal enhancement in the presence of acoustic noise is performed by Threshold Wavelet-based Forward Blind Source Separation (TWFBSS). This approach reduces the computational complexity from the Wavelet-based Forward Blind Source Separation (WFBSS) method [17].

Kalman filters can also be used with BSS techniques to deal with the noisy convolutive mixture. First, the BSS approach extracts the estimated source speech signal from the non-stationary noisy reverberant mixture. Then, Kalman filtering suppresses the noise components in the estimated speech signal [18]. Recently, new evolving techniques such as deep learning are also applied with the BSS approach in the reverberant noisy environment [19]. In general, the BSS methods are tested under non-Gaussian noise modeled by the fourth-order cumulant, and singular value decomposition-total least square method [20]. Moreover, the speech signals are often corrupted by different types of noise produced in the surrounding environment that can be tackled by the Dual Recursive non-Quadratic (DRNQ) adaptive method combined with FBSS to enhance the speech quality [21].

1.1. Background

The BSS methods estimate the desired source speech signals from the observed convolutive mixture containing noise. However, accurate identification of the targeted speech signal in a noisy reverberant environment is the fundamental goal of the speech processing systems. The traditional BSS methods are limited to multiple speech signals and sensors, where the de-noising process is challenging. Nevertheless, various signal processing methods, such as Single-channel Blind Source Separation (SBSS), Sparse Component Analysis (SCA), Variation Mode Decomposition (VMD), can tackle this issue. The VMD method is applied to decompose a single channel into two channels, and then SCA separates the speech signals. This approach shows enhancement of speech signal in under-determined conditions [22].

Another approach called AdaGrade is proposed in [23] for blind audio speech extraction that uses the gradient-based algorithm. The gradient learning rule is modified by pre-conditioning the input signal and using AdaGrade update. In this method, the natural gradient method with two-step pre-processing suppresses the noise in the receiving reverberant mixture. First, the bias removal method followed by least-square is applied to de-noise the noisy convolutive mixture. Then, a joint algorithm with a gradient method estimates the noisy signals and their mixing matrix [24]. Moreover, in [25], the BSS involves Eigen filtering, which receives the dominant frequencies of the signal, and then Wavelet de-noising is applied. It suppresses the noise components and retains the speech signal regardless of its frequency components. The authors in [26] propose an alternate method based on temporal predictability to get the individual independent noise signal where a non-negative matrix factorization algorithm enhances the speech signal. The performance is improved by adding time-correlation to the objective function, which restricts the time-varying gain of the noise [27]. Moreover, masking techniques can also be applied to separate the desired speech signal from the received mixture, where the time-frequency masking rule can define the BSS method [28]. In [29], the authors propose an EM algorithm to suppress the noise in the convolutive mixture for the complex-Gaussian signal model and the unknown deterministic model. The statistical model is defined for both models, and the EM algorithm is developed for these models to estimate the speech signal and its acoustic parameters.

Recently, unsupervised speech enhancement algorithms are gaining interest that uses a Real-Time (RT) two-channel BSS algorithm. In this method, a non-negative matrix factorization (NMF) dictionary is combined with generalized cross-correlation (GCC) spatial localization approach. The RT-GCC-NMF operates in a frame-by-frame manner, comparing individual dictionary atom with the desired speech signal or interfering noise based on the time-delay arrivals [30].

1.2. Contributions

The BSS approach separation gain depends on the selection of appropriate source prior function for extracting the desired speech signals [31,32]. For example, [33] proposes a mixed source prior model comprised of Super-Gaussian and Student's T to enhance the performance of the BSS. Consequently, in [34], the performance is improved by using a hybrid model, consisting of multivariate super-Gaussian and generalized Gaussian source priors. This approach models the higher amplitudes of the observed convolutive mixture by multivariate generalized Gaussian source prior, and the low amplitude are exploited by multivariate Gaussian source prior. Unlike these existing works, we propose an efficient multistage BSS method. In this method, a multivariate generalized Gaussian and Super-Gaussian source priors are combined as hybrid source prior model. The generalized Gaussian exploits higher-order statistical properties while other related information are modeled by multivariate Super-Gaussian. The contributions of this research work are as follows:

- We propose a novel efficient multistage approach for the BSS applications. This method concatenates the hybrid approach. Our proposed hybrid models combine multivariate generalized Gaussian and Super-Gaussian source priors.
- Based on the hybrid model, two different schemes are introduced, i.e., first BSS followed by de-noising and second de-noising in the first stage followed by BSS.
- The performance of proposed multistage hybrid model is evaluated with other multistage BSS methods having single source priors.
- The performance of the proposed models are investigated via extensive simulations in a noisy reverberant environment.

1.3. Organization

The article is organized into following sections. Section II describes the hybrid source prior signal model for the Independent Vector Analysis (IVA). Section III provides a detailed description of the proposed multistage approach for speech enhancement, followed by Results and discussion in Section IV. In Section V; we evaluate the performance of the multistage proposed model. Section VI presents the conclusion and future works.

2. Signal Model

Consider a clean source speech signal $x(t)$, noise signal $n(t)$, and received speech signal $y(t)$ contaminated by noise. It can be mathematically modeled as,

$$y(t) = x(t) + n(t). \tag{1}$$

The clean speech source signal, noise signal and the received noisy speech signal are transformed to FD domain and these parameters are denoted by $X(k)$, $N(k)$, and $Y(k)$ respectively. while k denotes the position index of the coefficient in the transformed domain. The design criteria of the estimator for the observation is to minimize the MSE is given by,

$$E\{X(k) - \hat{X}(k)\}, \tag{2}$$

where $E\{\cdot\}$ is the expectation operator and $\hat{X}(k)$ is the estimated source signal. Minimum Mean Square Error (MMSE) filter can be used to minimize the mean square error (MSE) in (2).

In a given noisy observation $\{y(t); \quad 0 \leq t \leq T\}$ with received signal $Y(K)$. The estimated $\hat{X}(k)$ can be obtained by [35,36],

$$\hat{X}(k) = E\{X(k)/Y(k)\}. \tag{3}$$

Equation (3) can be rewritten by Baye's theorem [35–37],

$$\hat{X}(k) = \frac{\int_{-\infty}^{\infty} a_k p(Y(k)/a_k) p(a_k)\, da_k}{\int_{-\infty}^{\infty} p(Y(k)/a_k) p(a_k)\, da_k} \tag{4}$$

where $p(.)$ is the probability density function (pdf) and a_k denotes the dummy variable representing all possible values of $X(k)$. Assuming a Gaussian distribution model, then $p(Y(k)/a_k)$ and $p(a_k)$ can mathematically written as,

$$p(Y(k)/a_k) = \frac{1}{\sqrt{2\pi\lambda_n(k))}} \exp\left(-\frac{(Y(k) - a_k)^2}{2\lambda_n(k)}\right) \tag{5}$$

and

$$p(a_k) = \frac{1}{\sqrt{2\pi\lambda_x(k)}} \exp\left(-\frac{a_k^2}{2\lambda_x(k)}\right) \tag{6}$$

where $\lambda_n(k) = E\{|N(k)|^2\}$ and $\lambda_x(k) = E\{|X(k)|^2\}$ are the variances of the noisy signal and clean signal respectively. Putting (5) and (6) in (4), then $\hat{X}(k)$ can be rewritten as [37],[36],

$$\hat{X}(k) = \frac{\xi(k)}{\xi(k) + 1} Y(k) \tag{7}$$

where $\xi(k)$ is the priori SNR and $\xi(k) = \frac{\lambda_x(k)}{\lambda_n(k)}$. The value λ_x and λ_n must be known. [38,39] shows the detail method for estimating λ_x. Decision directed estimated method develop to estimate λ_x [37]. The equation of estimating $\hat{\lambda}_x$ for λ_x is given by [36],

$$\hat{\lambda}_x = \alpha\hat{\lambda}_x(k)_p + (1 - \alpha) \max(Y(k)^2 - \lambda_x(k), 0) \tag{8}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 May 2021 doi:10.20944/preprints202105.0543.v1

where max(.) is the maximum function. It is used to obtain non-negative values. $\hat{\lambda}_x(k)_p$ is the estimated value of λ_x of the previous frame. α is the constant tuned for the best results. The parameter λ value is set to 0.98. If λ is set to 1. It deteriorates the speech signal and smaller values result in high musical noise.

3. Proposed Multistage BSS Approach

This section presents the proposed multistage approach for BSS and speech enhancement in a noisy reverberant environment. The multistage method comprises of BSS stage and de-noising stage using MMSE filtering as shown in Figure 1 and Figure 2, respectively. The proposed scheme evaluates different combinations of the BSS hybrid model and the de-noising MMSE method. In the first model (Figure 1), the observed convolutive mixture speech signal is first processed by BSS stage with hybrid source prior model for the extraction of estimated speech signals from the reverberant mixture. The de-noising module processes the resultant noisy extracted speech signals where the noisy elements in the separated speech signals are suppressed to improve the quality of the estimated signals. In the second model (Figure 2), the received reverberant observed speech mixture is de-noised by the MMSE filtering method in the first stage. In the second stage, the enhanced convolutive speech mixture is processed by the BSS stage with a hybrid source prior model to extract the de-noised estimated source speech signal from the enhanced reverberant mixture.

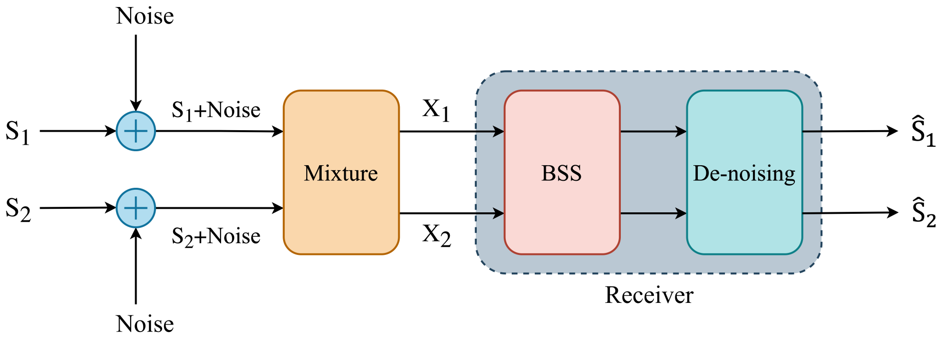

Figure 1. First Model

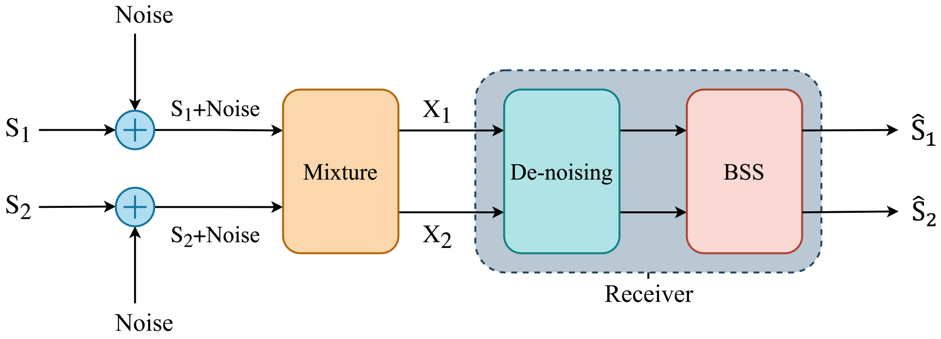

Figure 2. Second Model

A multivariate generalized Gaussian and Super-Gaussian source priors are combined as the hybrid source prior model in the BSS stage. The generalized Gaussian model exploits higher-order statistical properties while multivariate Super-Gaussian models other related information in the hybrid source prior model approach. The weights of the source priors in the hybrid model are adopted following the energy components of the received convolutive mixture [34]. In the de-noising stage, the MMSE filtering method is used to suppress the noisy component in the received convolutive mixture signal.

The hybrid source prior model provides a better separation performance and preserves the frequency dependencies between different frequency blocks for the IVA algorithm. Instead of using a single source prior distribution, a combination of multivariate generalized Gaussian and Super-Gaussian are used as source priors for the IVA to preserve the frequency dependencies. By using the KL divergence cost function to preserve the dependencies within the source speech signal while removing the dependencies among different source signals [40]. Mathematically the non-linear cost function for the hybrid model can be written as [31]

$$\begin{aligned} C &= KL(P(\hat{s}_1,\cdots,\hat{s}_N)||\prod_{i=1}^{N} q(\hat{s}_i)) \\ &= \text{const} - \sum_{k=1}^{K} \log|\det(W(k))| - \sum_{i=1}^{N} E\log q(\hat{s}_i), \end{aligned} \tag{9}$$

where $q(\hat{s}_i)$ is the i-th estimated source signal, $W(k)$ is the k-th separating matrix, and $q(\hat{s}_i)$is the source prior of i-th estimated source signal. The multivariate cost function in (9) is minimized by the Gradient Descent algorithm to remove the dependencies among different source signals and mathematically can be expressed as [31],

$$\begin{aligned} \Delta w_{ij}(k) &= -\frac{\partial C}{\partial w_{ij}(k)} \\ &= \sum_{l=1}^{N} (I_{il} - E\varphi^{k}(\hat{s}_i^{(1)},\cdots,\hat{s}_i^{(K)})\hat{s}_l^{(k)})w_{lj}^{(k)}, \end{aligned} \tag{10}$$

where I is the identity matrix and $\varphi^{(k)}(.)$ is the non-linear score function which can be mathematically expressed as [34],

$$\varphi^{k}\left(\hat{s}_i^{(1)},\cdots,\hat{s}_i^{(K)}\right) = -\frac{\partial \log q(\hat{s}_i^{(1)},\cdots,\hat{s}_i^{(K)})}{\partial \hat{s}_i^{k}} \tag{11}$$

The non-linear score function retains the dependency between different frequency bins, which is the main theme of the IVA algorithm and plays a vital role in the separation process. Fundamentally, the IVA method [31], it uses multivariate Super-Gaussian distribution source prior to model the different frequency bins inter-frequency dependencies which is expressed as,

$$q(s_i) \propto \exp\left(-\sqrt{\sum_{k=1}^{K}\left|\frac{\hat{s}_i(k)}{\sigma_i(k)}\right|^2}\right), \tag{12}$$

where $\sigma_i(k)$ represents the standard deviation of i-th source at k-th frequency block. Using equation (11) to determine the score function of equation (12), we get

$$\varphi^{(k)}\left(\hat{s}_i(1),\cdots,\hat{s}_i(K)\right) = \frac{\hat{s}_i(k)}{\sqrt{\sum_{k=1}^{K}|\hat{s}_i(k)|^2}}. \tag{13}$$

Equation (13) shows the non-linear score function of the fundamental IVA algorithm and is used for inter-frequency dependencies between source signals. However, the non-linear score function is not unique and is strongly dependent on the source prior. Therefore, we can use different source prior to exploit higher-order statistics. The generalized multivariate Gaussian can also be used as a source prior distribution to retain inter-frequency dependencies between different frequency blocks. Due to its heavy tails, it exploits higher-order statistical properties between the source signal and can be expressed as [32],

$$q(s_i) \propto \exp\left(-\sqrt[3]{(s_i-\mu_i)^{\dagger}\Sigma_i^{-1}(s_i-\mu_i)}\right). \tag{14}$$

Assuming mean $\mu_i = 0$ and covariance Σ_i equal to identity. Then, using equation (11) for equation (14), the score function will be,

$$\varphi^{(k)}\left(\hat{s}_i(1), \cdots, \hat{s}_i(K)\right) = \frac{2\hat{s}_i(k)}{3\sqrt[3]{(\sum_{k=1}^{K} |\hat{s}_i(k)|^2)^2}}. \tag{15}$$

In a noisy real-time environment, the non-stationary nature of the observed convolutive mixture contains high as well as low energy components. Hence, it is difficult for a single source prior to model the statistical properties of a non-stationary convolutive mixture. Therefore, a hybrid model is proposed containing multivariate generalized Gaussian and Super-Gaussian source priors. The hybrid source prior model can better model low and higher amplitudes [34]. The Super-Gaussian source prior function models low energy amplitude, and the high energy amplitude is modeled by multivariate generalized Gaussian source prior. The weights between these source priors in the hybrid source prior model are adopted based on the energy of a noisy convolutive mixture. The hybrid model can be expressed as

$$q(s_i) = \begin{cases} f_{GGD} & \text{if } \phi \geq 0.5 \\ f_{SGD} & \text{if } \phi < 0.5 \end{cases}. \tag{16}$$

f_{GGD} is the multivariate generalized Gaussian source prior distribution and f_{SGD} is the Super-Gaussian source prior distribution. The non-linear hybrid score function is mathematically written as,

$$\varphi^{(k)}\left(\hat{s}_i(1), \cdots, \hat{s}_i(K)\right) = \begin{cases} \left(\frac{2\hat{s}_i(k)}{3\sqrt[3]{(\sum_{k=1}^{K} |\hat{s}_i(k)|^2)^2}}\right) & ; \phi \geq 0.5 \\ \\ \left(\frac{\hat{s}_i(k)}{\sqrt{\sum_{k=1}^{K} |\hat{s}_i(k)|^2}}\right) & ; \phi < 0.5 \end{cases} \tag{17}$$

where $\phi = [0, 1]$ is the weighting parameter, which depends on the normalized energy of the received noisy convolutive mixture. The weights of the non-linear score functions and ϕ are adjusted by the normalized energy of the mixture at every frequency block.

4. Results and Discussion

This section provides the performance evaluation of the proposed work using the Matlab simulation tool. We generate artificially noisy convolutive mixed signals using a simulated room model and then apply the proposed multistage algorithm.

4.1. Experimental Setup

We consider 10 source speech signals comprised of 5 female and 5 male speakers from the TIMIT database [41]. All the source speech signals have the same loudness and a sampling rate of 8 kHz. The Hamming window of having a 75% overlapping factor is used. A noisy reverberant environment is used to evaluate the separation performance of the proposed multistage approach. For the fair comparison the methods used in [31], [42], and[32] are extended such that they are composed of their respective BSS technique and MMSE filtering mechanism for de-noising as presented in Figure 1 and 2, respectively. The proposed models are investigated for different parameters such as signal-to-distortion ratio (SDR) and RT. ΔSDR is defined as the difference between the desired SDR of the estimated speech signal and SDR of the speech mixtures, i.e., Δ SDR=SDR$_{\text{desired}}$-SDR$_{\text{mixture}}$.

4.2. Objective Evaluation

For the first proposed model shown in Figure 1, the SNR values are varied from −2 dB to 10 dB. The NFFT, window size, and RT are considered 1024, 512, and 100 msec, respectively. The obtained

doi:10.20944/preprints202105.0543.v1

results of different input speech mixtures are averaged and the results are provided in Table 1. Table 1 shows that the proposed model shows performance improvement as compared to BSS methods with multivariate Super-Gaussian [31], multivariate Student's T [42], and generalized Gaussian [32] source priors for estimated speech signals S_1 and S_2. Next, the RT parameter is varied from 40 to 200 msec. The window size and NFFT are set to 512 and 1024 respectively. In Table 1, the proposed model shows better results on SNR = 4 dB in comparison to rest of SNR values. Therefore, SNR = 4 dB is considered for the RT experiments. From the Table 2, it is also concluded that the proposed model shows improvement in comparison to [32], [31], and [42]. The first proposed model shows better performance due to its adaptability according to the non-stationary nature of the observed convolutive mixture as it contains low and high energy components compared to the single source prior BSS models.

Table 1. Average SNR results for the First Proposed Model Shown in Figure 1 with Variable SNR for Multistage BSS Models having Different Source Priors.

SNR	**Multivariate Gaussian Source Prior [31]**		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
(dB)	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2
-2	10.22	6.17	8.05	4.57	10.30	6.24	10.44	6.35
0	9.58	5.36	7.49	4.18	9.73	5.39	9.83	5.41
2	9.30	5.08	5.98	2.02	9.51	5.31	9.66	5.33
4	8.80	3.49	5.82	1.86	8.85	5.05	9.31	5.12
6	8.75	3.38	5.57	1.14	8.81	3.48	8.84	3.57
8	8.62	2.37	5.33	1.00	8.71	3.36	8.81	3.42
10	8.31	2.01	5.21	0.26	8.39	2.33	8.53	2.41

Table 2. Average RT results for the First Proposed Model Shown in Figure 1 with Variable RT for Multistage BSS Models having Different Source Priors.

RT	**Multivariate Gaussian Source Prior [31]**		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
(ms)	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2
40	16.10	7.47	9.64	2.27	16.29	7.56	16.37	7.68
80	12.45	5.79	6.94	2.00	12.65	5.72	12.81	5.87
120	7.06	3.02	5.92	1.65	7.26	3.22	7.41	3.39
160	4.49	2.11	2.88	1.04	4.63	2.63	4.83	2.85
200	3.92	1.62	2.25	0.28	4.01	1.81	4.23	1.97

For the second proposed model reflected in (Figure 2), the same procedure is followed to generate different convolutive speech mixtures having two speech signals and White Gaussian noise by simulated room model [43]. The SNR values are varied from -2 to 10 dBs. The obtained results of different multistage BSS approaches are averaged and presented in Table 3. From Table 3, the proposed model shows performance gain in comparison to [31], [42], and [32] for the $\hat{S}_1$ and $\hat{S}_2$. Similarly, the parameter RT is varied to evaluate the proposed model robustness. The values for window size, NFFT, and SNR are 512, 1024, and 4 dBs, respectively. The results provided in Table 4 shows performance improvement of the proposed model in comparison to the methods in [31], [42], and [32]. The results of the second model conclude that the switching between the two source priors according to the low and high energy components in the received convolutive mixture improve the performance from the single source prior BSS models [31], [42], and [32].

Table 3. Average SNR results for the Second Proposed Model Shown in Figure 2 with Variable SNR for Multistage BSS Models having Different Source Priors.

	Multivariate Gaussian Source Prior [31]		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
SNR (dB)	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2
-2	5.76	4.48	3.79	5.39	5.81	4.67	5.87	4.51
0	4.37	3.74	3.44	2.86	4.39	3.38	4.53	3.87
2	3.86	3.34	3.37	2.56	3.90	2.68	3.96	3.52
4	3.45	2.83	2.80	2.13	3.57	2.41	3.61	2.85
6	2.34	2.18	2.37	1.02	2.40	1.95	2.43	2.47
8	1.79	1.49	1.40	0.35	1.90	1.36	2.05	1.70
10	0.70	1.19	0.63	0.11	0.81	1.22	1.02	1.45

Table 4. Average RT results for the Second Proposed Model Shown in Figure 2 with Variable RT for Multistage BSS Models having Different Source Priors.

	Multivariate Gaussian Source Prior [31]		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
RT (ms)	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2	ΔSDR S_1	ΔSDR S_2
40	10.86	6.32	6.30	2.50	10.85	6.35	10.97	6.43
80	2.38	2.52	4.34	2.07	4.55	2.62	4.87	3.14
120	2.11	1.81	3.23	1.97	4.09	2.12	4.29	2.53
160	1.72	1.56	1.19	1.16	3.82	2.06	3.91	2.39
200	1.40	1.19	1.05	0.38	2.46	1.24	2.76	1.74

From the results of the two proposed models, it is clear that the first model (Figure 1) performs better than the second model (Figure 2). In the first model, the estimated source speech signals are first extracted from the observed noisy convolutive mixture. Then the noise in the separated speech signal is suppressed individually, leading to better performance. On the other hand, the second model performs de-noising first, suppressing the noise in the estimated source signal mixed as considering it received noisy convolutive mixture, resulting in performance degradation.

4.3. Subjective Evaluation

In the case of subjective evaluation, listening tests are performed to verify the simulation results obtained Table 5 to Table 8. Five participants conduct the subjective evaluation experiments (2 female, 3 male) where all the listening participants have normal hearing ability. Every listener is guided to mark a score from integer value 1 (estimated speech signals not audible) to integer 5 (estimated speech signals audible) of the extracted source speech signals from the noisy convolutive mixture. The listener listens to the original signal and the enhanced speech signals separated from the noisy reverberant mixtures using the two proposed models.

Table 5. Average MOS results of the subjective evaluation for the First Model Shown in Figure 1 with Variable SNR for Multistage BSS Models having Different Source Prior.

SNR	**Multivariate Gaussian Source Prior [31]**		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
(dB)	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2
-2	1.57	1.71	1.45	1.55	1.72	1.81	2.01	1.96
0	2.13	2.37	1.98	2.15	2.45	2.62	2.83	2.77
2	2.57	2.62	2.34	2.44	2.88	2.76	3.21	2.99
4	3.12	2.87	2.73	2.67	3.56	3.22	3.87	3.58
6	3.95	3.25	3.46	3.11	4.17	3.49	4.21	3.67
8	4.37	3.63	3.88	3.34	4.42	3.86	4.53	3.93
10	4.46	4.13	4.13	3.96	4.61	4.58	4.69	4.26

Table 6. Average MOS results of the subjective evaluation for the First Model shown in Figure 1 with Variable RT for Multistage BSS models having Different Source Prior.

RT	**Multivariate Gaussian Source Prior [31]**		**Student's T Distribution Source Prior [42]**		**Generalized Gaussian Source Prior [32]**		**Proposed Model**	
(ms)	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2	MOS for S_1	MOS for S_2
40	3.94	4.01	3.86	3.70	4.10	4.23	4.34	4.67
80	3.72	3.79	3.52	3.39	3.94	3.86	4.21	4.18
120	3.18	2.75	2.63	2.57	3.49	3.18	3.68	3.44
160	2.81	2.58	2.46	2.43	2.95	2.87	3.03	2.96
200	2.34	2.25	2.17	2.08	2.46	2.37	2.58	2.47

The same speech signals are chosen for both objective evaluation and subjective listening analysis in these experiments. In the multistage model presented in Figure 1, the values of the parameters for window size, NFFT, and RT are set to 512, 1024, and 100 msec. The SNR value is varied from -2 to 10 dBs. The score marked by the participants is based on the cleanness of the extracted signals from the convolutive mixture containing White Gaussian noise. The clean estimated speech signals

doi:10.20944/preprints202105.0543.v1

Table 7. Average MOS results of the subjective evaluation for the Second Model shown in Figure 2 with variable SNR for Multistage BSS models having Different Source Prior.

SNR (dB)	**Multivariate Gaussian Source Prior [31]** MOS for S_1	MOS for S_2	**Student's T Distribution Source Prior [42]** MOS for S_1	MOS for S_2	**Generalized Gaussian Source Prior [32]** MOS for S_1	MOS for S_2	**Proposed Model** MOS for S_1	MOS for S_2
-2	1.23	1.27	1.07	1.24	1.37	1.30	1.55	1.48
0	1.91	2.08	1.74	1.63	2.21	2.26	2.43	2.39
2	2.24	2.31	1.98	1.78	2.53	2.49	2.72	2.58
4	2.81	2.67	2.46	2.27	2.98	2.81	3.15	3.04
6	3.22	2.91	2.83	2.71	3.45	3.22	3.59	3.45
8	3.39	3.21	2.96	2.88	3.62	3.47	3.82	3.51
10	3.54	3.45	3.20	3.13	3.79	3.55	3.92	3.68

Table 8. Average MOS results of the subjective evaluation for the Second Model shown in Figure 2 with Variable RT for Multistage BSS models having Different Source Prior.

RT (ms)	**Multivariate Gaussian Source Prior [31]** MOS for S_1	MOS for S_2	**Student's T Distribution Source Prior [42]** MOS for S_1	MOS for S_2	**Generalized Gaussian Source Prior [32]** MOS for S_1	MOS for S_2	**Proposed Model** MOS for S_1	MOS for S_2
40	3.39	3.52	3.43	3.51	3.61	3.55	3.89	3.63
80	3.19	3.35	3.05	3.16	3.35	3.48	3.55	3.52
120	2.79	2.58	2.38	2.18	2.91	2.67	3.02	2.88
160	2.51	2.35	2.23	1.92	2.73	2.46	2.95	2.54
200	2.36	2.14	1.89	1.67	2.51	2.33	2.68	2.39

are marked with a higher mean opinion score (MOS) and vice versa. The results obtained from the listening participants for each extracted signal are averaged and presented in Table 5. In the Table 5, the proposed model shows improvement in its MOS in comparison with the multistage BSS having source priors [31], [42], and [32]. Next, the parameter RT is varied from 40 to 200 msec with fixed window length = 512, NFFT = 1024, and SNR = 4 dB. The averaged results obtained are provided in Table 6. From Table 6, the proposed model shows improvement over the methods in [31], [42], and [32].

In the second multistage model shown in Figure 2, previous parameter values are used for RT, window length, and FFT frame length. The SNR parameter value is varied from -2 to 10 dBs. The averaged MOS results are presented in Table 7, showing the performance improvement of the proposed model compared to Super-Gaussian [31], Student's T [42], and generalized Gaussian [32] source priors. Next, the RT parameter is varied from 40 to 200 msec with window length = 512, FFT = 1024, and SNR = 4 dB. The proposed model-averaged MOS results in Table 8 reflects an improvement from [42], [31], and [32].

5. Performance Evaluation

In this section, the separation performance of the proposed model is compared with various multistage BSS approaches with different source priors such as multivariate Super-Gaussian [31], Student's T [42], and generalized Gaussian distribution [32]. The two proposed multistage models are composed of the BSS approach to separate the estimated speech signals from the noisy convolutive mixture followed by the MMSE filtering technique to de-noise the signals.

For the performance evaluation of first model, we generate 20 different noisy convolutive speech mixtures with the help of a simulated room model by randomly selecting speech signals from a pool of 10 source speech signals (5 male and 5 female). We vary SNR and RT to obtain the average results where the SNR varies between -2 to 10 dB with window length = 512, NFFT = 1024, and RT = 100 msec. The average results are presented in Table 1 in terms of SDR, showing that the proposed model gains an enhancement of 0.3 dB for $\hat{S}_1$ and 0.5 dB for $\hat{S}_2$. Moreover, the proposed model is compared with the literature, showing its effectiveness with an optimum gain of 0.2 dB and 1 dB for both estimated speech signal $\hat{S}_1$ and $\hat{S}_2$, respectively.

Also, RT is varied from 40 to 200 msec with window length = 512, FFT frame length = 1024, and SNR = 4 dB. The noisy reverberant mixtures are fed to the proposed model and the other multistage BSS methodologies [31], [42], [32]. The average objective analysis results are presented in Table 2, which shows performance gain for the proposed model of 0.3 dB and 0.4 dB for $\hat{S}_1$ and $\hat{S}_2$ in comparison with other BSS approachs . The proposed approach also shows significant performance improvement of 3.81 dB and 2.9 dB for the estimated source signals $\hat{S}_1$ and $\hat{S}_2$ respectively from the multistage BSS having source prior [42]. The objective analysis is also compared with [32] in which the proposed method shows optimum improvement of 0.16 dB and 0.2 dB for $\hat{S}_1$ and $\hat{S}_2$, respectively.

For the performance evaluation of second model, the noisy reverberant mixtures are de-noised by using MMSE filtering technique in the first stage. In the second stage, the estimated speech signals are separated from the de-noised mixture using BSS method. The average results based on objective analysis by varying SNR are shown in Table 3. It is reflected from the Table 3 that the proposed model shows performance improvement of 0.2 dB for both $\hat{S}_1$ and $\hat{S}_2$ in comparison with other multistage BSS approachs. The results in Table 3 shows that the proposed approach achieved significant performance improvement of 0.7 dB and 0.8 dB in comparison with Student's T method [42] for estimated source signals $\hat{S}_1$ and $\hat{S}_2$, respectively. Moreover, Table 3 demonstrates that the proposed approach shows 0.1 dB and 0.4 dB gain from [32] for $\hat{S}_1$ and $\hat{S}_2$, respectively. The objective evaluation with variable RT having window size = 512, FFT = 1024, and SNR = 4 dB are provided in Table 4. From Table 4, the proposed model shows performance gain of 1.7 dB, 2.14 dB, 0.21 dB for $\hat{S}_1$ and 0.6 dB, 1.63 dB, 0.4 dB for $\hat{S}_2$ in comparison with [31], [42], [32], respectively.

Experiments are also performed to cross verify the simulations where the window length, NFFT, RT parameters are set to 512, 1024, and 100 msec, respectively. The 5 participants were asked to mark the MOS of the estimated speech signals extracted from the multistage BSS model with source prior [31], [42], [32], and the two proposed methods. The average MOS results from the first model are presented in Table 5 and 6 with variable SNR and RT, respectively. In the Table 5 with variable SNR, it is observed that the proposed approach achieved the performance gain of 0.5, 0.8, 0.2 in terms of MOS for estimated source $\hat{S_1}$ and 0.4, 0.6,0.12 for estimated source $\hat{S_2}$ in comparison with other multistage BSS methods respectively. For varying RT parameter having fixed window length = 512, FFT = 1024, and SNR = 4 dB, the average MOS are shown in Table 6 with gain of 0.4, 0.6, 0.2 for $\hat{S}_1$ and 0.5, 0.7, 0.2 for $\hat{S}_2$. Same procedure is followed to verify the second proposed model and the results are displayed in Table 7 and in Table 8. For varying SNR, it is observed from Table 7 that the proposed model achieves MOS gain of 0.4,0.7, 0.2 for $\hat{S}_1$ and 0.3, 0.7, 0.2 for $\hat{S}_2$. Similarly, in Table 8 results are provided by varying RT that shows MOS gain of the proposed model i.e., 0.4, 0.6, 0.2 for $\hat{S_1}$ and 0.2, 0.5, 0.1 for $\hat{S_2}$.

5.1. Comparative analysis of the proposed models

A comparative analysis of the two proposed models are presented in Figure3 to Figure6 for estimated speech signals $\hat{S}_1$ and $\hat{S}_2$. The results of these figures are deduced from objective evaluation Tables 1 to 4 for variable SNR and RT. From Figure3 and Figure4, it is clear that the first model provides significant improvement in comparison to the second model for estimated source signals $\hat{S}_1$ and $\hat{S}_2$ with variable SNR. Similarly, for varying RT, it is observed from Figure5 and Figure6 that the first model shows considerable performance gain in comparison to the second model for $\hat{S}_1$ and $\hat{S}_2$. The performance of the first model (Figure 1) is better than the second model (Figure 2) for both RT and SNR because the first model suppresses the noise in the estimated source signal extracted from the noisy convolutive mixture while in the second model, the de-noising technique suppresses noise and estimated source signals mixed in the noisy convolutive mixture. The de-noising module considers the other estimated signals in the noisy convolutive mixture as noise resulting in performance degradation.

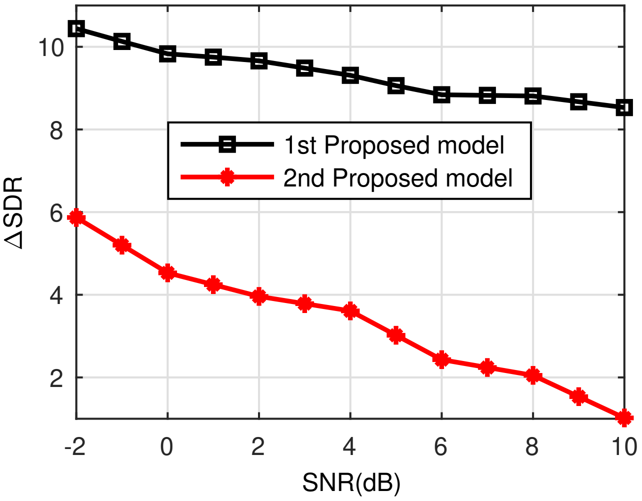

Figure 3. Comparison of the two proposed models for $\hat{S}_1$ with Variable SNR (dB).

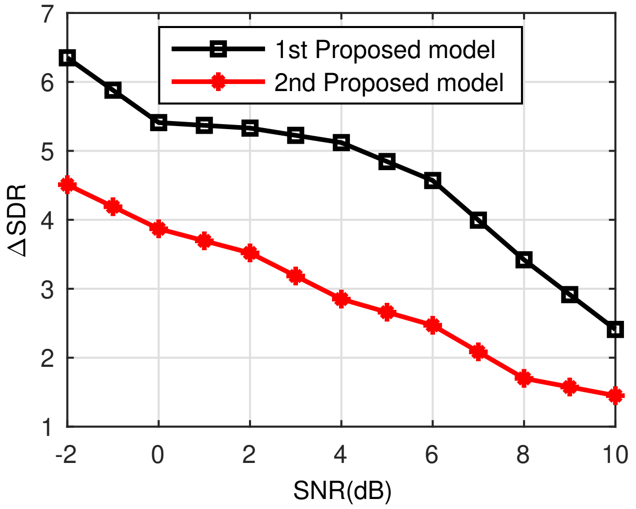

Figure 4. Comparison of the two proposed models for $\hat{S}_2$ with Variable SNR (dB).

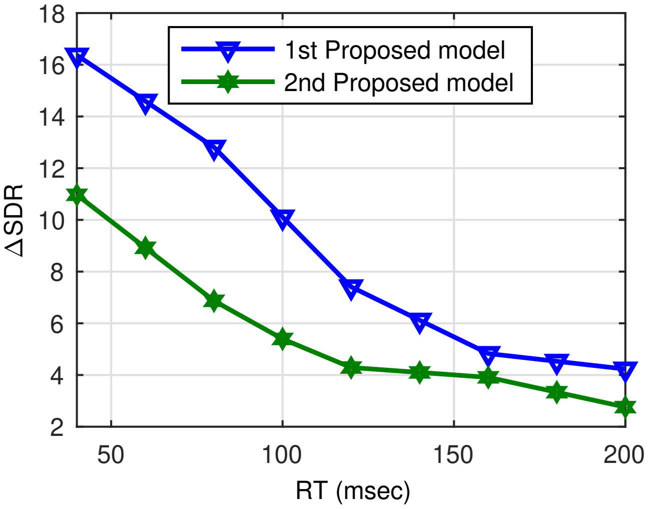

Figure 5. Comparison of the two proposed models for $\hat{S}_1$ with Variable RT (msec).

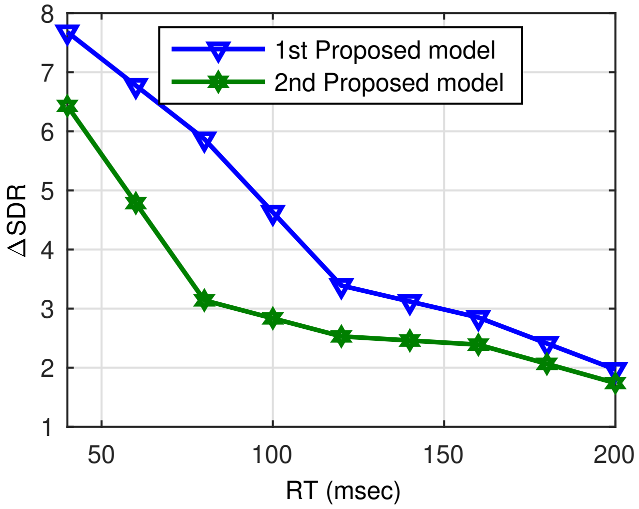

Figure 6. Comparison of the two proposed models for $\hat{S}_2$ with Variable RT (msec).

6. Conclusion and Future Work

This paper proposes an efficient hybrid multistage approach for blind source separation (BSS) of noisy convolutive speech mixture. In the BSS stage, a hybrid source prior model consisting of multivariate super-Gaussian and generalized Gaussian distribution model the source signals in the observed noisy reverberant mixture. The weights are assigned between the source priors following the energy of the observed convolutive mixture. In the de-noising stage, the noise is suppressed by the MMSE filtering technique using two different proposed models. In the first model, the BSS module is followed by the de-noising stage. In the second model, the de-noising module is followed by the BSS stage. Both proposed models are compared with the literature, where the results clearly show the performance improvement of the proposed schemes. Furthermore, it is observed from the results that the proposed model with BSS module followed by de-noising stage shows a significant gain in comparison with the model with first de-noising followed by BSS stage.

Author Contributions: The work was developed as a collaboration among all authors. J.B. K., T. J., R.A.K. and N.S. designed the study and system development. R. A. K., N. S and M. A. directed the research and collaborated in discussion on the proposed system model. The manuscript was mainly drafted by J.B. K., T. J., R.A.K. and N.S and was revised and corrected by all co-authors. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gupta, V.; Bhowmick, A.; Chandra, M.; Sharan, S. Speech enhancement using MMSE estimation and spectral subtraction methods. 2011 International Conference on Devices and Communications (ICDeCom). IEEE, 2011, pp. 1–5.
2. Souden, M.; Araki, S.; Kinoshita, K.; Nakatani, T.; Sawada, H. A multichannel MMSE-based framework for speech source separation and noise reduction. *IEEE Transactions on Audio, Speech, and Language Processing* **2013**, *21*, 1913–1928.
3. Enzner, G.; Thüne, P. Robust MMSE filtering for single-microphone speech enhancement. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 4009–4013.
4. Fenghua, Z.; Le, Y.; Jian, W.; Qiang, S. Speech signal enhancement through wavelet domain MMSE filtering. 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering. IEEE, 2010, Vol. 5, pp. 118–121.
5. Kirubagari, B.; Palanivel, S.; Subathra, N. Speech enhancement using minimum mean square error filter and spectral subtraction filter. International Conference on Information Communication and Embedded Systems (ICICES2014). IEEE, 2014, pp. 1–7.
6. Khalil, R.A.; Jones, E.; Babar, M.I.; Jan, T.; Zafar, M.H.; Alhussain, T. Speech emotion recognition using deep learning techniques: A review. *IEEE Access* **2019**, *7*, 117327–117345.
7. KHALIL, R.; ASHRAF, S.; JAN, T.; JEHANGIR, A.; KHAN, J. Enhancement of Speech Signals Using Multiple Statistical Models. *Sindh University Research Journal-SURJ (Science Series)* **2015**, *47*.
8. Yang, J.; Wang, Z. Blind separation algorithm for speech and noise based on diagonalizing second-order statistics accurately. 2010 2nd IEEE International Conference on Information Management and Engineering. IEEE, 2010, pp. 370–373.
9. Hongyan, L.; Guanglong, R. Blind separation of noisy mixed speech signals based Independent Component Analysis. 2010 First International Conference on Pervasive Computing, Signal Processing and Applications. IEEE, 2010, pp. 586–589.
10. Yin, J.; Liu, Z.; Jin, Y.; Peng, D.; Kang, J. Blind Source Separation and Identification for Speech Signals. 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). IEEE, 2017, pp. 398–402.
11. Rivet, B.; Girin, L.; Jutten, C. Mixing audiovisual speech processing and blind source separation for the extraction of speech signals from convolutive mixtures. *IEEE transactions on audio, speech, and language processing* **2006**, *15*, 96–108.
12. Parikh, D.N.; Anderson, D.V. Blind source separation with perceptual post processing. 2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE). IEEE, 2011, pp. 321–325.

13. Rahima, H.; Djebari, M.; Mohamed, D. Blind speech enhancement and acoustic noise reduction by SFTF adaptive algorithm. 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B). IEEE, 2017, pp. 1–4.
14. Fisli, S.; Djendi, M.; Guessoum, A. Modified predator-prey particle swarm optimization based two-channel speech quality enhancement by forward blind source separation. 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP). IEEE, 2018, pp. 1–6.
15. Bendoumia, R.; Djendi, M.; Guessoum, A. New symmetric subband forward algorithm based on simple variable step-sizes for speech enhancement. 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B). IEEE, 2017, pp. 1–6.
16. Bendoumia, R.; Djendi, M. Speech enhancement using backward adaptive filtering algorithm: Variable step-sizes approaches. 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT). IEEE, 2015, pp. 1–5.
17. Ghribi, K.; Djendi, M.; Berkani, D. Thresholding wavelet-based forward BSS algorithm for speech enhancement and complexity reduction. 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP). IEEE, 2018, pp. 1–6.
18. Beack, S.K.; Lee, B.; Hahn, M.; Nam, S.H. Blind source separation and Kalman filter-based speech enhancement in a car environment. Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. IEEE, 2004, pp. 520–523.
19. Wang, Z.Q.; Wang, D. Combining spectral and spatial features for deep learning based blind speaker separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **2018**, *27*, 457–468.
20. Wang, H.; Bi, A.; Xu, P.; Gao, C. Convolutive Blind Source Separation Algorithm Based on Higher Order Statistics. 2013 Third International Conference on Intelligent System Design and Engineering Applications. IEEE, 2013, pp. 487–490.
21. Abdessamed, B.; Yahia, B.; Mohamed, D. Hands Free Communication Improvement in Airplane by a New Dual RNQ Adaptive Algorithm. 2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM). IEEE, 2018, pp. 1–4.
22. Wang, C.; Zhu, X.; Li, X. Interference Suppression Based on Single-channel Blind Source Separation in Weather Radar. 2019 International Conference on Meteorology Observations (ICMO). IEEE, 2019, pp. 1–4.
23. Cmejla, J.; Koldovsky, Z. Multi-channel speech enhancement based on independent vector extraction. 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE, 2018, pp. 525–529.
24. Tang, H.; Wang, S. Noisy blind source separation based on adaptive noise removal. Proceedings of the 10th World Congress on Intelligent Control and Automation. IEEE, 2012, pp. 4255–4257.
25. Routray, A.; Das, N.; Dash, P. Robust preprocessing: Denoising and whitening in the context of blind source separation of instantaneous mixtures. 2007 5th IEEE International Conference on Industrial Informatics. IEEE, 2007, Vol. 1, pp. 377–380.
26. Yang, Y.; Li, Z.; Wang, X.; Zhang, D. Noise source separation based on the blind source separation. 2011 Chinese Control and Decision Conference (CCDC). IEEE, 2011, pp. 2236–2240.
27. Chen, Y. Single channel blind source separation based on nmf and its application to speech enhancement. 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN). IEEE, 2017, pp. 1066–1069.
28. Yatabe, K.; Kitamura, D. Time-frequency-masking-based Determined BSS with Application to Sparse IVA. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 715–719.
29. Schwartz, B.; Gannot, S.; Habets, E.A. Two model-based EM algorithms for blind source separation in noisy environments. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **2017**, *25*, 2209–2222.
30. Wood, S.U.; Rouat, J. Unsupervised low latency speech enhancement with RT-GCC-NMF. *IEEE Journal of Selected Topics in Signal Processing* **2019**, *13*, 332–346.
31. Kim, T.; Attias, H.T.; Lee, S.Y.; Lee, T.W. Blind source separation exploiting higher-order frequency dependencies. *IEEE transactions on audio, speech, and language processing* **2006**, *15*, 70–79.
32. Liang, Y.; Naqvi, S.M.; Wang, W.; Chambers, J.A. Frequency domain blind source separation based on independent vector analysis with a multivariate generalized Gaussian source prior. In *Blind Source Separation*; Springer, 2014; pp. 131–150.

doi:10.20944/preprints202105.0543.v1

33. Rafique, W.; Erateb, S.; Naqvi, S.M.; Dlay, S.S.; Chambers, J.A. Independent vector analysis for source separation using an energy driven mixed Student's t and super Gaussian source prior. 2016 24th European Signal Processing Conference (EUSIPCO). IEEE, 2016, pp. 858–862.
34. Khan, J.B.; Jan, T.; Khalil, R.A.; Altalbe, A. Hybrid Source Prior Based Independent Vector Analysis for Blind Separation of Speech Signals. *IEEE Access* **2020**, *8*, 132871–132881.
35. Ephraim, Y.; Malah, D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE transactions on acoustics, speech, and signal processing* **1985**, *33*, 443–445.
36. Soon, Y.; Koh, S.N.; Yeo, C.K. Noisy speech enhancement using discrete cosine transform. *Speech communication* **1998**, *24*, 249–257.
37. Ephraim, Y.; Malah, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on acoustics, speech, and signal processing* **1984**, *32*, 1109–1121.
38. Boll, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing* **1979**, *27*, 113–120.
39. McAulay, R.; Malpass, M. Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **1980**, *28*, 137–145.
40. Kim, T.; Lee, I.; Lee, T.W. Independent vector analysis: definition and algorithms. 2006 Fortieth Asilomar Conference on Signals, Systems and Computers. IEEE, 2006, pp. 1393–1396.
41. Garofolo, J.S. TIMIT acoustic phonetic continuous speech corpus. *Linguistic Data Consortium, 1993* **1993**.
42. Rafique, W.; Naqvi, S.M.; Jackson, P.J.; Chambers, J.A. IVA algorithms using a multivariate student's t source prior for speech source separation in real room environments. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015, pp. 474–478.
43. Allen, J.B.; Berkley, D.A. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America* **1979**, *65*, 943–950.