Single-Channel Blind Source Separation using Adaptive Mode Separation- Based Wavelet Transform and Density-Based Clustering with Sparse Reconstruction

abdellah kacha
mina kemiha (kemihamina@yahoo.fr)

Research Article

Keywords:

Posted Date: August 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1937910/v1

Single-Channel Blind Source Separation using Adaptive Mode Separation-Based Wavelet Transform and Density-Based Clustering with Sparse Reconstruction

KEMIHA Mina* and KACHA Abdellah

Laboratoire de Physique de Rayonnement et Applications, University of Jijel, Algeria

* Corresponding author : Mina Kemiha, Laboratoire de Physique de Rayonnement et Applications, Université de Jijel, BP 98 Ouled Aissa, 18000 Jijel, Algeria

E-mail: kemihamina@yahoo.fr

Abstract

In this study, the signal-channel blind source separation (SCBSS) problem has been addressed using a novel approach. The approach is based on combining the adaptive mode separation-based wavelet transform with adaptive mode separation (AMSWT) and the density-based clustering with sparse reconstruction. The approach is performed in Time frequency domain and in reverberant environment. First, using the Fourier transform, the amplitude spectrum of the observed mixture signal is obtained. Then, using variational scaling and wavelet functions, the AMSWT is introduced to adaptively extract spectral intrinsic components (SIC). To obtain a better time-frequency distribution, the AMSWT is applied to each mode. Thus, the SCBSS problem is transformed into a non-underdetermined. Then, for each frequency bin; the density-based clustering, reformulated to eigenvector clustering problem, is performed to estimate the mixing matrix. Finally, the sparse reconstruction is introduced to reconstruct the estimated source. The proposed approach has been evaluated using an objective measure of separation quality. According to experimental results, the proposed approach presents a powerful method to solve the SCBSS problem, and provide better separation performances than the existing methods.

1 Introduction

The blind sources separation (BSS) aim to separates the sources signals from the mixed signals without any information. Applications for BSS include medical imaging and engineering [1, 2], astrophysics [3], image processing [4], geophysical data processing [5], speech processing [6,7], detection and radar localization [8], communication systems [9], automatic transcription of speech [10], musical instrument identification [11], mechanical flaw detection [12], multichannel telecommunications [13], multi-spectral astronomical imaging [14] and speech recognition [15].

The BSS categories are described in the literature as being linear and nonlinear, instantaneous and convolutive, over-complete and underdetermined. The convolutive BSS is demonstrated to be an effective way to represent the speech signal mixing mechanism in a reverberant environment [16, 17]. Either the time domain or the frequency domain can be used to formulate the BSS problem. The BSS can be also treated in time-frequency domain (TF) where the computational efficiency of BSS algorithms is higher.

In most situations and for many practical uses, only one-channel recording is available. This particular instance of the under-determined source separation problem called single channel source separation (SCSS), and has been the subject of many studies. In order to address the single channel audio source separation problem, numerous strategies have been introduced in the literature [18]. In [19] the authors attempt to combine the maximum-likelihood estimation and NMF based on Itakura-Saito divergence measurement. The Short Time Fourier Transform (STFT) representation of a single channel observed signal has been subjected to the nonnegative matrix factorization (NMF) approach in [20], although the method necessitates the use of extra training data. A combination of empirical mode decomposition (EMD) and ICA, as well as wavelet transformations, have been suggested in [21], although wavelet transforms need some specified basis functions to represent a signal, there is no rigorous mathematical theory underpinning the EMD or its improved algorithms [22]. The bark scale aligned wavelet packet decomposition has been introduced in [23], and the separation step has been performed using the Gaussian mixture model (GMM), which was employed before the Fourier transform. In [24] the authors propose the variational mode decomposition (VMD) method to solve the SCBSS problem. The separation process is performed using joint approximate diagonalization based on fourth-order cumulant matrices. In [25] authors present a novel method in noisy environment. The method is based on selecting the time-frequency (TF) units of signal presence and computing the mixture spectral amplitude, the separation process is performed based on TF masking. In [26] an adaptive signal separation operation (ASSO) has been proposed. The method is performed by introducing a time-varying parameter that adapts locally to Ifs, and using linear chirp (linear frequency modulation). The single-channel technique has been explored for muscle artifact removal from multichannel EEG in [27].

The classic TF resolution is computed using the STFT transformation, this TF resolution does not reflect the time-varying information; in addition, the STFT yields a time-frequency resolution with only uniform frequency and time resolutions. A new Adaptive Mode Separation-Based Wavelet Transform (AMSWT) has been proposed in [29] based on [29, 30]. The AMSWT method involves solving a recursive optimization problem in order to adaptively extract spectral intrinsic components (SIC). The limited support of each spectral mode is implemented in order to

establish the spectral boundaries for wavelets bank configuration. Then, the created wavelets bank configuration using the obtained spectral boundaries to highlight the spectrum information. The AMSWT strategy is a fully adaptive one that doesn't require prior knowledge.

In [31] a new method to solve the under-determined BSS problem for convolutive mixture is proposed. The method is based on combining Density-based grouping and sparse source reconstruction. The method is performed in time-frequency domain. The density-based clustering is introduced to estimate the mixing matrix. The method is performed based on a certain local dominant assumption; the mixing matrix estimate is converted as an eigenvector clustering issue. The rank-one structure of the local covariance matrices of the mixture TF vectors is first used to extract the eigenvectors. By combining weight clustering and density-based clustering, these eigenvectors are subsequently grouped and tweaked to provide an approximated mixing matrix. The sparse reconstruction is performed for sources estimation by using the iterative Lagrange multiplier approach, the source reconstruction is converted into a ℓ_p-norm minimization.

In this paper a new method has been proposed to solve the SCBSS problem. The method is based on combining the AMSWT [28] and density-based clustering with sparse reconstruction method introduced in [31]. The method is performed in three stages. The amplitude spectrum of the observed mixture signal is obtained using STFT. The convolution in the time domain can be approximated by a multiplication in the STFT domain. Then, a better TF resolution is obtained using the variational scaling and wavelet functions, which are applied on the spectral intrinsic components (SIC), this one is adaptively extracted using the AMSWT. By creating virtual multi-channel signals of the TF resolution, the single channel has been changed into a non-underdetermined problem. Then, for each TF resolution and for each frequency bin, the density-based clustering which is converted to eigenvector clustering problem, combined with the sparse reconstruction, which is converted to a sparse reconstruction minimization problem, these approaches are respectively performed for each TF resolution to estimate the mixing matrix and estimated sources reconstruction. The BSSeval is introduced to evaluate the proposed method in terms of source-to-distortion ratio (SDR), source-to-artifact ratio (SAR), source-to-interference ratio (SIR), and compared to the BSS performance results obtained via VDM method [24], adaptive spectrum amplitude estimator and masking method [25] and the nonnegative tensor factorization of modulation spectrograms method [32].

The following sections make up the remaining content: the SCBSS problem formulation is presented in the second section, the adaptive mode separation-based wavelet transform is introduced in the third section. The fourth section shows Density-based Clustering method; the fifth

section presents the Source Reconstruction. The main steps of the proposed algorithm with the application of this algorithm in the simulation experiments and the comparison results with other algorithms in the sixth section; finally, conclusions and discussions are given in the seventh section.

2 Convolutive Mixing System Model

Let $\mathbf{xx}(t) = [xx_1(t), .., xx_M(t)]^T$ a vector of M observed sources abstained via the mixing of N independent sources $\mathbf{s}(t) = [s_1(t), .., s_N(t)]^T$. The BSS problem aim to estimate the N sources from the M mixtures. The convolutive mixture is occurs by the propagation of the sound through space and multiple paths which cause the reflections from different objects, especially in rooms and closed environments. The convolutive mixture is modeled as the flowing equation:

$$xx_j(t) = \sum_{i=1}^{N} \sum_{k=0}^{K-1} h_{ji}(k) s_i(t-k) \tag{1}$$

The matrix form is given as :

$$xx(t) = \boldsymbol{H} * s(t) = \sum_{k=0}^{K-1} \boldsymbol{H}_k s(t-k) \tag{2}$$

where h_{ji} denotes the impulse response from source i to sensor j, and $\boldsymbol{H}$ is an MxN matrix that contains the k^{th} filter coefficients.

The only one-channel recording is accessible in most cases and for many practical purposes. Numerous studies have examined this instance known as single channel source separation. In this case $= 1$. The convoltuvive SCBSS in time-frequency domain is described as the following equation:

$$X(f,t) = \sum_{i=0}^{N} xx_i(f,t) \tag{3}$$

The traditional source separation techniques are ineffective in this scenario. The SCSS study area in which the issue might be viewed as a single observation combined with numerous unidentified sources.

3 Adaptive Mode Separation-Based Wavelet Transform

The STFT is used to calculate the classic TF resolution, which has an even bandwidth distribution across all frequency channels, and suffers from the TF resolution limitation due to the fixed window size. The speech signal is described as being substantially non-periodic and non-

stationary. Therefore, using the STFT transform will result in mistakes, particularly when complex transitory phenomena like voice mixing occur in the signal under study.

To each mode, the AMSWT performs a time-frequency analysis using variational scaling and wavelet functions. The method is built on the ADMM solver [33], which then defines a bank of variational scaling functions and wavelets depending on the spectral boundaries that have been defined. As a result, the approximate coefficients are derived by multiplying the analyzed signal xx by the variational scaling function inner product. However, the inner product of the analyzed signal xx with variational wavelets yields the detailed coefficients, which are given by the following formulae respectively:

$$W_{xx}(0,t) = \langle xx, \emptyset_1 \rangle = \int xx(\tau)\overline{\emptyset}_1(\tau - t)d\tau \tag{4}$$

and

$$W_{xx}(k,t) = \langle xx, \psi_k \rangle = \int xx(\tau)\bar{\psi}_k(\tau - t)d\tau \tag{5}$$

where xx is the input signal.

In [28], under the amplitude-modulated frequency-modulated (AM-FM) assumption, the intrinsic modes $u(t)$ have distinguishable features in the frequency domain. Using the alternate direction method of multiplier (ADMM) solver, the spectral modes can be adaptively obtained, similar to intrinsic mode functions (IMF) extraction, to estimate compact modes:

$$\min_{u_k,\omega_k} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_K u_k = xx(t) \tag{6}$$

Where $xx(t)$ is the signal to be decomposed under the constraint that over all modes should be the input signal. $\delta(.)$ is a Dirac impulse. $\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$ denotes the original data and its Hilbert transform. u_k, ω_k and k denote the modes and their central frequencies and the mode number respectively.

The spectral segmentation boundary number can be determined empirically using the equation below.

$$\widetilde{K} = min\{n \in \mathbb{Z}^+ | n \geq 2\rho \ln N\} \tag{7}$$

where N presents the signal length and ρ is the scaling exponent determined by the detrended fluctuation analysis (DFA).

According to [28], the equation is solved using a quadratic penalty term, the parameter λ design the Lagrangian multiplier to render the problem unconstrained,.

$$L(u_k, \omega_k, \lambda) = \eta \sum_k \left\| \delta_t [\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)] e^{j\omega_k t} \right\|_2^2 + \langle \lambda, xx - \sum_K u_k \rangle + \left\| xx - \sum_K u_k \right\|_2^2 \quad (8)$$

therefore u_k is determined recursively as

$$\hat{u}_k^{n+1}(\omega) = \frac{XX(\omega) - \sum_{i \neq j} \hat{u}_i^{n+1}(\omega) + \frac{\hat{\lambda}^n}{2}}{1 + 2\eta(\omega - \omega_k^n)^2} \quad (9)$$

where $XX(\omega)$, $\widehat{u_\iota}(\omega)$ and $\hat{\lambda}(\omega)$ denote the Fourier transform of the input signal $xx(t)$, the mode function $u_i(t)$ and $\lambda(t)$ respectively. η denotes the balancing parameter of the data-fidelity constraint. The center frequencies ω_k^{n+1} are updated as the center of gravity of the corresponding mode's power spectrum using the following equation

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega} \quad (10)$$

As a result, rather than using a predefined wavelet bank, we create adaptive wavelet banks based on spectral modes and their corresponding center frequencies, which represent the intrinsic components.

Authors in [28] used the mode bandwidth and central frequencies to define the boundaries between each mode, however in the literature, some authors just used the average of the two central frequencies as the spectral boundary, which ignores the spectrum distribution.

We consider the kth mode with the mean frequency ω_k and a spectral bandwidth β_k, then the boundary $\mathbf{\Omega}_k$ between kth the and the $k+1$ mode is given by the following equation

$$\mathbf{\Omega}_k = \frac{\omega_k + \frac{\beta_k}{2} + \omega_{k+1} - \frac{\beta_{k+1}}{2}}{2} \quad (11)$$

where $\mathbf{\Omega}_k = 0$ and $\mathbf{\Omega}_k = \pi$.

The authors apply the same notion used in the production of both Littlewood–Paley and Meyer's wavelets [34] for variational scaling functions and wavelets based on spectral boundaries. $\widehat{\emptyset}_k$ and $\hat{\psi}_k$ are respectively defined by the following equation, with γ is the parameter that ensures no overlap between the two consecutive transitions.

$$\widehat{\emptyset}_k = \begin{cases} 1, & \omega \leq (1-\gamma)\mathbf{\Omega}_k \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \mathbf{\Omega}_k)\right), & (1-\gamma)\mathbf{\Omega}_k \leq \omega \leq (1+\gamma)\mathbf{\Omega}_k \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

and

$$\hat{\psi}_k = \begin{cases} 1, & (1+\gamma)\boldsymbol{\Omega}_k \le \omega \le (1-\gamma)\boldsymbol{\Omega}_{k+1} \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \boldsymbol{\Omega}_{k+1})\right), & (1-\lambda)\boldsymbol{\Omega}_{k+1} \le \omega \le (1+\lambda)\boldsymbol{\Omega}_{k+1} \\ \sin\left(\frac{\pi}{2}\alpha(\gamma, \boldsymbol{\Omega}_k)\right), & (1-\lambda)\boldsymbol{\Omega}_k \le \omega \le (1+\lambda)\boldsymbol{\Omega}_k \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Where $\alpha(\gamma, \boldsymbol{\Omega}_k) = \beta\{\left(\frac{1}{2\gamma\boldsymbol{\Omega}_k}\right)[|\omega| - (1-\gamma)\boldsymbol{\Omega}_k]\}]$ and $\beta(x)$ is an arbitrary function defined as follow:

$$\beta(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 1 \\ \beta(x) + \beta(1-x) = 1, & 0 < x < 1 \end{cases} \tag{14}$$

The algorithm adaptive mode separation-based wavelet transform is summarized as following:

Step1 : Time frequency presentation
Input : Observed mixture.

- Using the Fourier transform, obtain the amplitude spectrum signal.
- Obtain the appropriate spectrum spectral modes (segments). Execute the first inner loop and the second inner loop to update u_k according to equation (9); and update ω_k according to equation (10); respectively
- Compute proper spectral boundaries using equation (11). Then, using equation (12) and (13), the bank of variational scaling functions and wavelets based on the spectral boundaries is defined.
- Finally, using equations (4) and (5) respectively, apply variational scaling and wavelet functions to each mode to obtain the time-frequency distribution.

Output: time frequency distribution (TF) of observed mixture.

4 Density-based Clustering

In [31] the authors introduce the eigenvector clustering as an alternative to estimate the mixing matrix. The eigenvector clustering is based on two factors, such as the local density ρ_q, and the minimum distance δ_q that may be taken between point q and any additional points with a higher density, are given respectively by the following equations

$$\rho_q \triangleq \sum_{k \ne q} e^{-\frac{v_{qk}^2}{\tau_c^2}} \tag{15}$$

and

$$\delta_q = \min_{k:\rho_k > \rho_q} (v_{qk}) \tag{16}$$

where the region for each data point is defined by a cutoff distance τ_c, and $\boldsymbol{V}$ denote the similarity matrix whose elements v_{qk}.

From the eigenvectors $\boldsymbol{A}$ whose elements $\boldsymbol{a}_q$, the similarity matrix $\boldsymbol{V}$ is generated as follow:

$$\boldsymbol{V} \triangleq \begin{bmatrix} \upsilon_{11} & \cdots & \upsilon_{1Q} \\ \vdots & \ddots & \vdots \\ \upsilon_{Q1} & \cdots & \upsilon_{QQ} \end{bmatrix} \tag{17}$$

where $\upsilon_{qk} = \left\| \boldsymbol{a}_q - (\boldsymbol{a}_q^H \boldsymbol{a}_k) \right\|_F^2$ and $q, k = 1, .., Q$

The eigenvector extraction is based on the local covariance matrix $\boldsymbol{R}_q^{\mathrm{X}}$ where $\boldsymbol{R}_q^{\mathrm{X}} = \sum_{i=1}^{N} \sigma_{i,q}^2 h_i h_i^H$ where h_i is called as steering vector representing each direction of mixing matrix.

According to [31], there is at least one sub-block indexed as q_i for which the associated local covariance $\boldsymbol{R}_{q_i}^{\mathrm{X}}$ where the local covariance matrix has roughly a rank-one structure. In [29], this conditions is exploited, the authors applies eigenvalue decomposition (EVD) to the local covariance matrix of $\boldsymbol{R}_q^{\mathrm{X}}$ resulting in the following equation:

$$\boldsymbol{R}_q^{\mathrm{X}} = \boldsymbol{U}_q \boldsymbol{\Sigma}_q \boldsymbol{U}_q^H \tag{18}$$

where $\boldsymbol{U}_q$ and $\boldsymbol{\Sigma}_q$ denote the eigenvector matrix and eigenvalue matrix respectively.

The extracted vector denoted $\mathbf{a}_q$ corresponds to the largest eigenvalue of $\boldsymbol{\Sigma}_q$, and also presents the first eigenvector in $\boldsymbol{U}_q$. To obtain an eigenvector matrix described by $\mathbf{A} \triangleq [\mathbf{a}_1, \ldots, \mathbf{a}_Q]$, the eigenvector extraction is done sub-block wisely.

According to [31], the global maximum in the density indexed as q^* has a minimum distance δ_{q^*} defined as follows:

$$\delta_{q^*} = \max_{q,k=1,\ldots,Q} (\upsilon_{qk}) \text{ if } \rho_{q^*} = \max_{q=1,\ldots,Q} (\rho_q) \tag{19}$$

The two components are multiplied together to provide the following score:

$$\gamma_q = \rho_q \times \delta_q \tag{20}$$

To get $\{\gamma_q\}_{q=1}^{Q}$, the scores from the equation (20) are applied to all of the sub-blocks. The obtained scores are then rated in order of decreasing order, as a result, the eigenvectors with the greatest N scores are retrieved as clusters, which are denoted by $\mathbf{C} \triangleq [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N]$.

As mentioned in [31], it would be difficult to cluster eigenvectors using solely the density-based strategy described above. To address this issue, a weight clustering approach to further tune the projected clusters introduced in [35] is used. The procedures of weighted eigenvector clustering can be concluded in three steps.

First, the eigenvector is weighted by a kernel function defined as follow:

$$\boldsymbol{b}_{qk} \triangleq e^{\omega_{qk}^2 / \tau_0^2} \boldsymbol{a}_q \tag{21}$$

where $k = 1,..,N$ and $\omega_{qk} = \left\| \boldsymbol{a}_q - \left(\boldsymbol{a}_q{}^H \, \boldsymbol{c}_k \right) \boldsymbol{c}_k \right\|_F^2$.

Then, the weighted covariance matrix is created an given as :

$$\boldsymbol{R}_k^{\mathrm{b}} = \sum_{q=1}^{Q} \boldsymbol{b}_{qk} \; \boldsymbol{b}_{qk}{}^H \tag{22}$$

Finally, the EVD is applied to the weighted covariance matrix $\boldsymbol{R}_k^b$ given as follow

$$\boldsymbol{R}_k^{\mathrm{b}} = \boldsymbol{U}_{qk} \boldsymbol{\Sigma}_{qk} \boldsymbol{U}_{qk}{}^H \tag{23}$$

As an updated of cluster $\boldsymbol{c}_k$ where $k = 1,\ldots,N$, the eigenvector that corresponds to the largest eigenvalue from the equation (23) is extracted.

The mixing matrix estimation algorithm is summarized as the following steps:

Step2 : Mixing Matrix Estimation
Input : X which present the TF resolution of observed signal whose element $\mathbf{x}_d$.

For each blocks q ∈ Q do

- Calculate the local covariance matrix of $\mathbf{R}_{\mathrm{q}}^{\mathrm{X}}$ using $\widehat{\mathbf{R}}_{f,q}^{\mathbf{X}} = \frac{1}{p} \sum_{d=q(P-1)+1}^{qP} \mathbf{x}_{f,d} \, \mathbf{x}_{f,d}^H$
- Construct the eigenvector matrix **A** from the equation (18).

End

- Using the eigenvector matrix **A**, compute the similarity matrix defined by equation (17)

For each blocks q ∈ Q do

- Calculate the local density ρ_{q} and the minimum distance δ_{q} and the score γ_{q} using equations (15), (16), and (20) respectively

End

- Calculate δ_{q^*} using equation (19), then, obtain de score sequence $\Upsilon = [\gamma_1, \ldots, \gamma_Q]$.
- To get the score sequence of Υ, reorder the eigenvector matrix with the same permutation of decreasing alignment. So, To get the estimated clusters $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$, truncate the first *N* reordered eigenvectors.

For $k = 1 \; to \; N \; do$
For each sub-blocks q ∈ Q do

- Calculate the weighted eigenvector $\mathbf{b}_{\mathrm{qk}}$ using $\mathbf{a}_{\mathrm{q}}$ and $\mathbf{c}_{\mathrm{k}}$, then calculate $\mathbf{R}_{\mathrm{qk}}^{\mathrm{b}}$ using respectively equations (21) and (22)
- calculate $\tilde{\mathbf{h}}_{\mathrm{k}}$ using equation (23)

end
end
Output: Estimated mixing matrix $\widehat{\mathbf{H}}$.

5 Source Reconstruction

In [31] the sparsity-based method is introduced as an alternative to reconstruct the estimated source signal. Using a ℓ_p-norm based-minimization measurement (the convergence is guarantee for $0 < p < 1$), the method consists to convert the source reconstruction problem to a sparse

reconstruction minimization problem. A designed iterative Lagrange multiplier approach with an appropriate initialization procedure is used to solve this minimization problem.

The source reconstruction is performed to find the sparsest term of s_d. For this, the maximum posterior likelihood of s_d is given as the following equation

$$\max_{s_d} \prod_{i=1}^{N} P(|s_{i,d}|) \quad s.t. \mathbf{x}_d = \widehat{\mathbf{H}} s_d \tag{24}$$

where the complex-valued super-Gaussian distribution $P(|s_{i,d}|)$ is given by the following equation:

$$P(|s_{i,d}|) = p \frac{\gamma^{1/p}}{\Gamma(\frac{1}{p})} e^{-|s_{i,d}|^p} \tag{25}$$

where p and γ control shape and variance of the probability function. Γ denoted the gamma function. $\mathbf{H}$ design the estimated mixing matrix. The problem returns to solve the equivalent optimization problem given as follow:

$$\min_{s_d} \sum_{i=1}^{N} |s_{i,d}|^p \quad s.t. \mathbf{x}_d = \widehat{\mathbf{H}} s_d \tag{26}$$

The Lagrange multiplier method is introduced to solve the optimization problem. Hence, the problem is reformulated to an unconstrained optimization problem as follows:

$$\min_{s_d,\alpha} \mathcal{F}(s_d, \alpha) \triangleq \sum_{i=1}^{N} |s_{i,d}|^p + \alpha^H (\mathbf{x}_d - \widehat{\mathbf{H}} s_d) \tag{27}$$

where α design the Lagrange multiplier. The problem implicit solution is given as follow:

$$s_d = \Psi^{-1}(s_d)\, \widehat{\mathbf{H}}^H (\widehat{\mathbf{H}}\, \Psi^{-1}(s_d)\, \widehat{\mathbf{H}}^H)^{-1}\, \mathbf{x}_d \tag{28}$$

where

$$\Psi^{-1}(s_d) \triangleq \begin{bmatrix} |s_{1,d}|^{2-p} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & |s_{N,d}|^{2-p} \end{bmatrix}$$

The iterative scheme to obtain the solution s_d is given as follow:

$$\hat{s}_d^{(iter+1)} = \begin{cases} \Psi^{-1}(\hat{s}_d{}^{(iter)}) \widehat{\mathbf{H}}^H \left(\widehat{\mathbf{H}}\, \Psi^{-1}(\hat{s}_d{}^{(iter)}) \widehat{\mathbf{H}}^H\right)^{-1} \mathbf{x}_d & if \left\|\hat{s}_d{}^{(iter)}\right\|_0 \geq M \\ \Psi^{-1}(\hat{s}_d{}^{(iter)}) \widehat{\mathbf{H}}^H \left(\widehat{\mathbf{H}}\, (\Psi^{-1}(\hat{s}_d{}^{(iter)}) + \epsilon \mathbf{I})^{-1}\, \widehat{\mathbf{H}}^H\right)^{-1} \mathbf{x}_d & elseif \left\|\hat{s}_d{}^{(iter)}\right\| < M \end{cases} \tag{29}$$

The source reconstruction algorithm is summarized as the following steps

Input : Time frequency presentation of observed signal denoted $\mathbf{X}$ whose element $\mathbf{x}_d$ and Estimated mixing matrix $\widehat{\mathbf{H}}$

- For each frequency bin d
- Initialize the sources a $\hat{s}_d^{(0)} = \sum_{j=1}^{C_N^M} \omega_j \mathrm{y}_{j,d}$
- Repeat
- Update $\hat{s}_d^{(iter)}$ using equation 29
- $iter = iter + 1$
- Until $\left\|\hat{s}_d{}^{(iter)}\right\|_p^p - \left\|\hat{s}_d{}^{(iter+1)}\right\|_p^p$ is less than a given threshold.
- End

Output: time frequency presentation of estimated sources.

6 Result and discussion

In order to investigate the effectiveness of the proposed method, numerical simulations have been performed in reverberant environment. The TIMIT database [36] and NOIZEUS database [37] were used to build the speech dataset, which was chosen at random (available online). The sampling rate for the voice signals is $f_s = 16\,kHz$, and the speakers might be either female or male. Using the technique outlined in [38], the propagation environment is simulated as a reverberant room shown by figure 1.

The room impulse response from source i to sensor is illustrated by figure 2. By adjusting the reverberant time, a variety of convolutive mixed signals can be produced. It is crucial to evaluating the transmission duration of signal decay to 60 dB in order to reflect the room reverberation.

The spectrum of the observed signal is obtained by the STFT transformation; where the convolution in the time domain is transformed into multiplication in the STFT domain. AMSWT approach is introduced to obtain an optimal spectral mode separation, by applying wavelet and variational scaling to each mode, the TF output with high time frequency resolution is produced, the following steps are given by algorithm 1. Thus, the SCBSS problem is transformed into a non-underdetermined problem by establishing virtual multi-channel signals of the TF resolution of the observed signals. Then, the M time-frequency presentation of the mixture is divided into Q non-overlapping blocks.

As a pre-processing step at the mixing matrix estimation stage, the TF resolution of the observed signal, for each frequency bin $\mathbf{x}_d$ is whitened. The whitening process is performed using the eigenvector matrix $\mathbf{U_x}$, and the eigenvalue matrix $\mathbf{\Sigma_x}$ of $E(\mathbf{x}_d{\mathbf{x}_d}^H)$, and expressed by the following equation $\mathbf{x}_d{}^w = \mathbf{\Sigma}_\mathbf{x}^{-1/2}\mathbf{U}_\mathbf{x}^H\mathbf{x}_d$.

The estimation of the mixing matrix is reformulated into an eigenvector clustering issue. First, the local covariance matrices of mixture signal's rank-one structure was used to extract the

eigenvectors; Secondly, a density-based clustering technique was used to create clusters from these eigenvectors; Third, the clusters were modified using a lightweight clustering approach to produce the estimated mixing matrix, the steps are summarized by algorithm 2.

The ambiguity of scaling is solved by rescaling the estimated mixing matrix by restricting the first row. The order of the reconstructed sources is aligned, by grouping the nearby source TF vectors based on their correlation, in terms of power ratio, in order to resolve the permutation ambiguity [31].

The post-processing stage involves de-whitening the predicted mixing matrix by $\widehat{\mathbf{H}} = \mathbf{U}_x\boldsymbol{\Sigma}_{\mathbf{x}}^{1/2}\widetilde{\mathbf{H}}$.

Then, the source reconstruction is reformulated into a sparse minimization problem, whose solution was achieved using an initialization-corrected iterative Lagrange multiplier approach as summarized in algorithm 3. The algorithm outputs are the TF resolution of the N estimated sources

Finally, the estimated sources are obtained in TF resolution, which are transformed into time domain using the modified method proposed in [39]. The proposed method is summarized by the flowchart shown in figure 3.

The BSSeval toolbox [40] is used to analyze the performance of the proposed approach. The estimated sources are expressed as $\hat{s} = s_{target} + e_{interf} + e_{noise} + e_{artif}$ for the objective performance criteria measurement, where s_{target} refers to the source signals, e_{interf} stands for interference from other sources, e_{noise} stands for distortion brought on by noise, and e_{artif} includes all other artifacts introduced by the separation algorithm.

In [31] the parameter p plays a significant impact in source reconstruction performance. Many tests have been performed using different p value to assess the effect of SDRs using the given Dataset. The table displays the obtained SDRs with p parameters varying from 0.1 to 0.9.

The table presents the SDRs evaluation obtained via the proposed method using various value of p which characterizes the ℓ_p-norm based-minimization measurement method, as can be seen, the SDR marginally increases as p increases and reaches its max when p = 0.7. For improved performance, the parameter p is set to 0.7 in the subsequent experiments. Changing the p parameter value to take advantage of the source sparsity prove that the sparse reconstruction based on ℓ_p-norm based-minimization approach is a flexible framework.

The estimated sources performances are evaluated using the source-to-distortion ratio (SDR), the source-to-artifact ratio (SAR) and the source-to-interference ratio (SIR) criterions, and compared with the estimated sources performances obtained via VDM method [24], adaptive spectrum amplitude estimator and masking method [25] and the nonnegative tensor factorization of

modulation spectrograms method [32]. The SDR, SAR and SIR are defined by the following equation:

$$SDR = 10\log_{10}\frac{\left\|s_{target}\right\|^2}{\left\|e_{interf}+e_{noise}+e_{artif}\right\|^2} \tag{30}$$

$$SAR = 10\log_{10}\frac{\left\|s_{target}+e_{interf}+e_{noise}\right\|^2}{\left\|e_{artif}\right\|^2} \tag{31}$$

$$SIR = 10\log_{10}\frac{\left\|s_{target}\right\|^2}{\left\|e_{interf}\right\|^2} \tag{32}$$

The figure 4 presents a comparison evaluated by the mean square errors (MSEs) [26] between the original signal and the estimated sources obtained via the proposed method and the estimates sources obtained via VDM method [24], adaptive spectrum amplitude estimator and masking method [25] and the nonnegative tensor factorization of modulation spectrograms method [32]. The comparison has been performed for various reverberation conditions where the reverberation time is varied from 100 *ms* to 500 *ms*. As can be observed the proposed method provide in smaller MSE even highly reverberant environment.

The figures 5, 6 and 7 present respectively, a comparison in term of SDR, SAR and SIR between the estimated sources obtained via the proposed method, and the estimates sources via VDM method [24], adaptive spectrum amplitude estimator and masking method [25] and the nonnegative tensor factorization of modulation spectrograms method [32]. The comparison has been performed for various reverberation times where the reverberation time is varied from 100 *ms* to 500 *ms*.

As can be seen, the proposed method results in a better performance in terms of the three performance criteria compared to VDM, the adaptive spectrum amplitude estimator and masking, and the nonnegative tensor factorization of modulation spectrograms methods in reverberant environment. The proposed method results in higher performance criteria even in highly reverberant environment.

7 Conclusion

A new method to solve the SCBSS problem has been presented. The method is combining the adaptive mode separation-based wavelet transform with adaptive mode separation (AMSWT) and the density-based clustering with sparse reconstruction. The SCBSS problem is transformed

into a non-underdetermined. The method operates in the time-frequency domain and in reverberant environment. The proposed method has been tested on speech datasets constructed from TIMIT and NOIZEUS databases for various reverberation time conditions. The simulations results indicate the smaller MSE criteria and the high values of SIR, SAR and SDR. The simulations results demonstrate the effectiveness of the proposed method to solve the SCBSS problem even in highly reverberant environment

References

1. Al-Baddai, S.; Al-Subari, K.; Tomé, A.M.; Volberg, G.; Lang, E.W. Combining EMD with ICA to Analyze Combined EEG-fMRI Data. In Proceedings of the MIUA, Egham, UK, 9–11 July 2014; pp. 223–228.
2. Zeng, X.; Li, S.; Li, G.J.; Zhou, Y.; Mo, D.H. Fetal ECG extraction by combining single-channel SVD and cyclostationarity-based blind source separation. Int. J. Signal Process 2013, 6, 367–376.
3. Wilson, S.; Yoon, J. Bayesian ICA-based source separation of Cosmic Microwave Background by a discrete functional approximation. arXiv 2010, arXiv:1011.4018.
4. Draper, B.A.; Baek, K.; Bartlett, M.S.; Beveridge, J.R. Recognizing faces with PCA and ICA. Comput. Vis. Image Underst. 2003, 91, 115–137.
5. A.-K. Takahata, E.-Z. Nadalin, R. Ferrari, L.-T. Duarte, R. Suyama, R.-R. Lopes, J.M.-T. Romano, M. Tygel, Unsupervised processing of geophysical signals: a review of some key aspects of blind deconvolution and blind source separation. IEEE Signal Process. Mag. 29(4), 27-35 (2012).
6. X. Huang, L. Yang, R. Song, and W. Lu, ``Effective pattern recognition and nd-density-peaks clustering based blind identication for underdetermined speech mixing systems," Multimedia Tools Appl., vol. 77, no. 17, pp. 2211522129, Sep. 2018.
7. A. Nagathil, C. Weihs, K. Neumann, and R. Martin, ``Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners," J. Acoust. Soc. Amer., vol. 142, no. 3, pp. 12191228, Sep. 2017.
8. Liu, X.; Srivastava, A.; Gallivan, K. Optimal Linear Representations of Images for Object Recognition. In Proceedings of the 2003 Conference on Computer Vision and Pattern RecognitionWorkshop, Madison, WI, USA, 18–20 June 2003.
9. Yang, J.; Williams, D.B. MIMO Transmission Subspace Tracking with Low Rate Feedback. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA, 23 March 2005.
10. Eronen, A. Musical Instrument Recognition Using ICA-Based Transform of Features and Discriminatively Trained HMMs. In Proceedings of the Seventh International Symposium on Signal Processing and Its Applications, Paris, France, 4 July 2003.
11. R. B. Randall, ``A history of cepstrum analysis and its application to mechanical problems," Mech. Syst. Signal Process., vol. 97, pp. 319, Dec. 2017.
12. M. A. Haile and B. Dykas, ``Blind source separation for vibrationbased diagnostics of rotorcraft bearings," J. Vib. Control, vol. 22, no. 18, pp. 38073820, Oct. 2016.
13. J. Traa, P. Smaragdis, Multichannel source separation and tracking with RANSAC and directional statistics. IEEE/ACM Trans. Audio, Speech, Language Process. 22(12), 2233-2243 (2014).
14. D. Nuzillard, A. Bijaoui, Blind source separation and analysis of multispectral astronomical images. Astron Astrophys Suppl Ser. 147(1), 129–138 (2000).

15. I. Bekkerman, J. Tabrikian, Target detection and localization using mimo radars and sonars. IEEE Trans. Signal Process. 54(10), 3873-3883 (2006).
16. A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing. New York, NY, USA: Wiley, 2003.
17. S. Makino, T. W. Lee, and H. Sawada, Blind Speech Separation. Berlin, Germany: Springer-Verlag, 2007.
18. Duan, Z.; Zhang, Y.; Zhang, C.; Shi, Z. Unsupervised Single-Channel Music Source Separation by Average Harmonic Structure Modeling. IEEE Trans. Audio Speech Lang. Process. 2008, 16, 766–778.
19. C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," Neural Comput., vol. 21, no. 3, pp. 793–830, 2009.
20. Wang, B.; Plumbley, M.D. Investigating Single-Channel Audio Source Separation Methods Based on Non-Negative Matrix Factorization. In Proceedings of the ICA Research Network InternationalWorkshop, Liverpool, UK, 18–19 September 2006; pp. 17–20
21. Mijovic, B.; De Vos, M.; Gligorijevi´c, I.; Taelman, J.; Van Hu_el, S. Source Separation From Single-Channel Recordings by Combining Empirical-Mode Decomposition and Independent Component Analysis. IEEE Trans. Biomed. Eng. 2010, 57, 2188–2196.
22. NE. Huang, Z. Shen, SR. Long, MC. Wu, HH. Shih, Q. Zheng, NC. Yen, CC. Tung, HH. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. In proceedings of The Royal Society A Mathematical Physical and Engineering Sciences 454(1971), 903-995 (1998).
23. Litvin, Y.; Cohen, I. Single-Channel Source Separation of Audio Signals Using Bark Scale wavelet packet decomposition. j. signal process. syst. 2010, 65, 339–350.
24. ya'nan zhang , shengbo qi, and lin zhou. Single Channel Blind Source Separation for Wind Turbine Aeroacoustics Signals Based on Variational Mode Decomposition IEEE access
25. N. Tengtrairat, W. L. Woo, S. S. Dlay, and Bin Gao online Noisy Single-Channel Source Separation Using Adaptive Spectrum Amplitude Estimator and Masking IEEE transactions on signal processing, vol. 64, no. 7, april 1, 2016 1881 Online
26. Lin Li , Charles K. Chui, and Qingtang Jiang Direct Signal Separation via Extraction of Local Frequencies With Adaptive Time-Varying Parameters IEEE transactions on signal processing, vol. 70, 2022 pp 2321- 2333
27. Xun Chen, Aiping Liu, Joyce Chiang, Z. Jane Wang, Martin J. McKeown, and Rabab K. Ward, Removing Muscle Artifacts From EEG Data: Multichannel or Single-Channel Techniques? IEEE SENSORS JOURNAL, VOL. 16, NO. 7, APRIL 1, 2016
28. Fangyu Li , Bangyu Wu , Naihao Liu , Ying Hu, and Hao Wu, '' Seismic Time Frequency Analysis via Adaptive Mode Separation-Based Wavelet Transform'' . IEEE geoscience and remote sensing letters, vol. 17, no. 4, april 2020. pp 696-700.
29. J. Gilles, "Empirical wavelet transform," IEEE Trans. Signal Process., vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
30. K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.
31. J. Yang , Y. Guo , Z. Yang Under-Determined Convolutive Blind Source Separation Combining Density-Based Clustering and Sparse Reconstruction in Time-Frequency Domain. IEEE transactions on circuits and systems–i: regular papers, vol. 66, no. 8, 2019 pp 3015-3027.
32. T. Barker, T. Virtanen, Blind Separation of Audio Mixtures Through Nonnegative Tensor Factorization of Modulation spectrograms, IEEE/ACM transactions on audio, speech, and language processing, vol. 24, no. 12, december 2016.
33. M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, no. 5, pp. 303–320, 1969.
34. I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA, USA: SIAM, 1992.

35. J.-J. Yang, H.-L. Liu, “Blind identification of the underdetermined mixing matrix based on K-weighted hyperline clustering,” *Neurocomputing*, vol. 149, pp. 483–489, Feb. 2015.
36. TIMIT database. https://catalog.ldc.upenn.edu/ldc93s1
37. NOIZEUS database http://ecs.utdallas.edu/loizou/speech/noizeus/
38. E. Habets, “Room impulse response (RIR) generator,” 2008
39. Wei Liu, Siyuan Cao, and Yangkang Chen. Applications of variational mode decomposition in seismic time-frequency analysis GEOPHYSICS, VOL. 81, NO. 5 (SEPTEMBER-OCTOBER 2016); P. V365–V378,
40. C. Févotte, R. Gribonval, E. Vincent, BSS EVAL toolbox user guide, IRISA (2005), http://www.irisa.fr/metiss/bss_eval

Figures

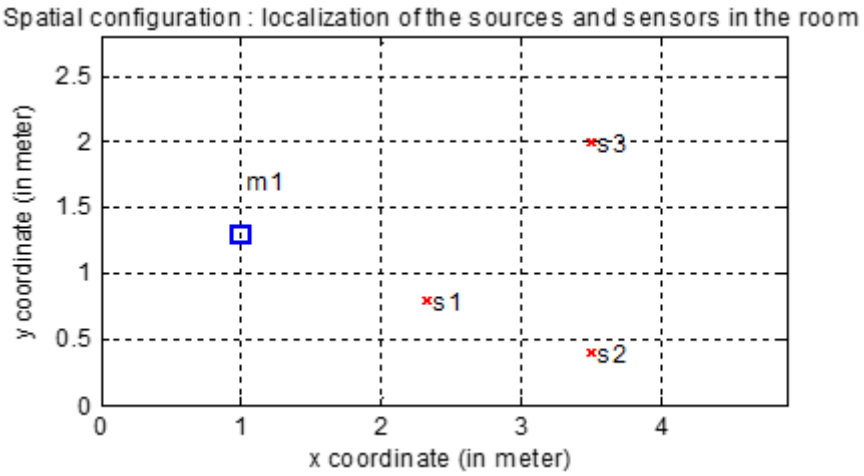

Figure 1

Sources-microphone configurations.

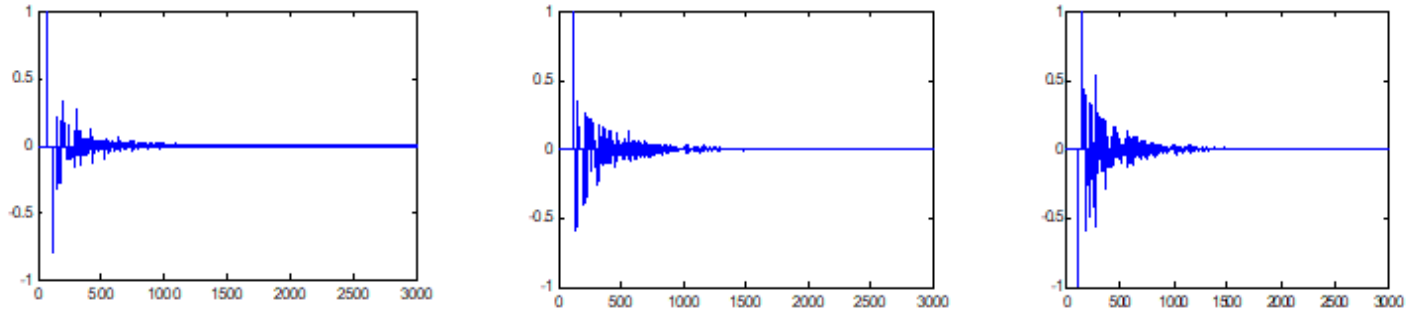

Figure 2

Room impulse responses from source to microphone.

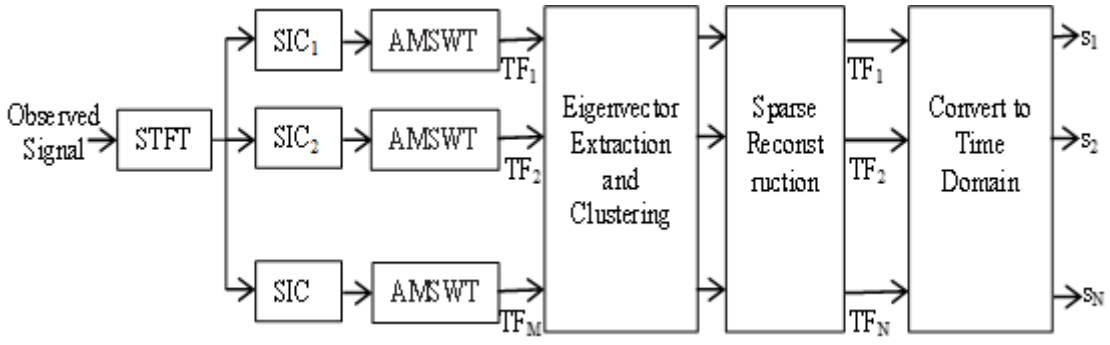

Figure 3

Proposed method flowchart

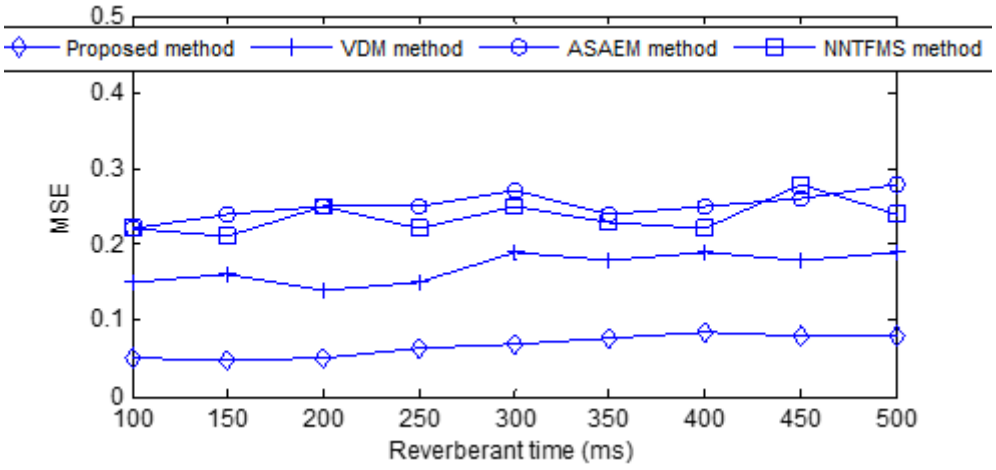

Figure 4

Comparison in term of mean square errors (MSEs) between the estimated sources obtained via the proposed method and the estimates sources obtained via VDM method, adaptive spectrum amplitude estimator and masking method (ASAEM), and the nonnegative tensor factorization of modulation spectrograms method (NNTFMS)

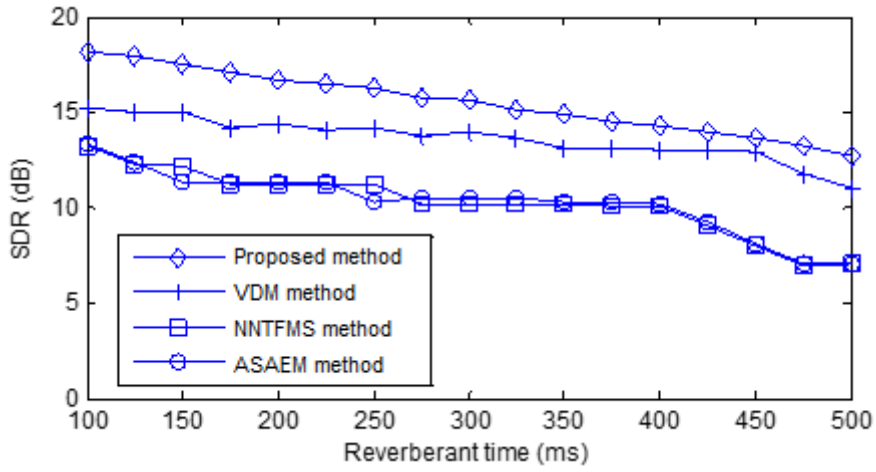

Figure 5

Comparison in term of SDR between the estimated sources obtained via the proposed method and the estimates sources obtained via VDM method, adaptive spectrum amplitude estimator and masking method (ASAEM), and the nonnegative tensor factorization of modulation spectrograms method (NNTFMS)

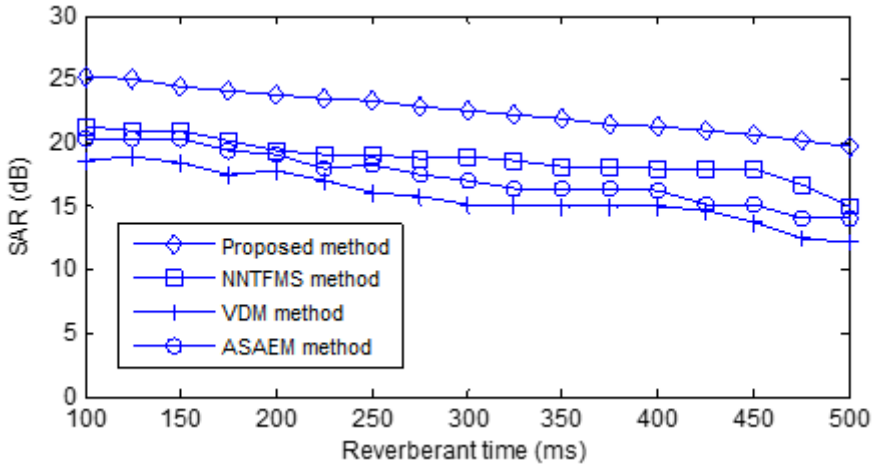

Figure 6

Comparison in term of SAR between the estimated sources obtained via the proposed method and the estimates sources obtained via VDM method, adaptive spectrum amplitude estimator and masking method (ASAEM), and the nonnegative tensor factorization of modulation spectrograms method (NNTFMS)

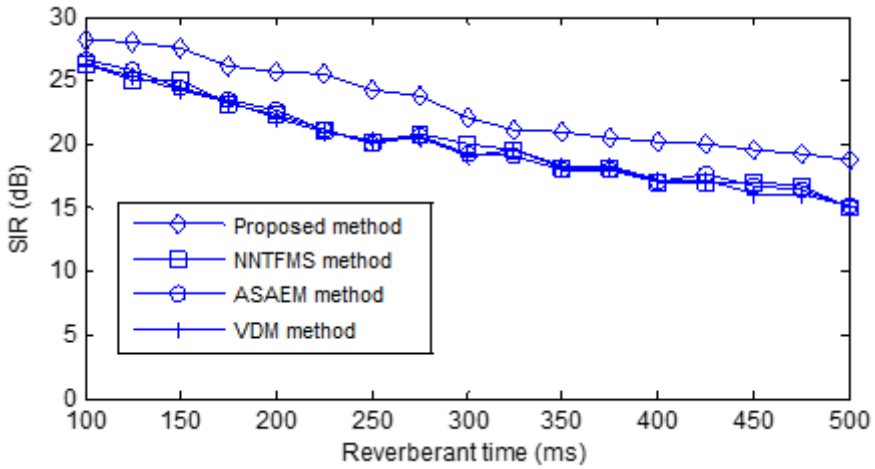

Figure 7

Comparison in term of SIR between the estimated sources obtained via the proposed method and the estimates sources obtained via VDM method, adaptive spectrum amplitude estimator and masking method (ASAEM), and the nonnegative tensor factorization of modulation spectrograms method (NNTFMS)