Single-Channel Blind Source Separation using Adaptive Mode Separation-Based Wavelet Transform and ICA

[1]M. KEMIHA, [2]A. KACHA
[1]Radiation Physics and Applications laboratory, University of Jijel, BP 98 Ouled Aissa, 18000 Jijel. ALGERIA
[2]Radiation Physics and Applications laboratory, Electronic Department, University of Jijel, ALGERIA

Abstract: In this paper, a new method to solve the signal-channel blind source separation (SCBSS) problem has been proposed. The method is based on combining the Adaptive Mode Separation-Based Wavelet Transform (AMSWT) and the ICA-based single channel separation. First, the amplitude spectrum of the instantaneous mixture signal is obtained via the Fourier transform. Then, the AMSWT is introduced to adaptively extract spectral intrinsic components (SIC) by applying the variational scaling and wavelet functions. The AMSWT is applied to every mode to obtain the time-frequency distribution. Then the time-frequency distribution of the mixed signal is exploited. The ICA-based single-channel separation has been applied on spectral rows corresponding to different time intervals. Finally, these components are grouped using the β-distance of Gaussian distribution D_β. Objective measure of separation quality has been performed using the scale-invariant (SI) parameters and compared with the existing method to solve SCBSS problem. Experimental results show that the proposed method has better separation performance than the existed methods, and the proposed method present a powerful method to solve de SCBSS problem.

Keywords: Signal-channel blind source separation. Adaptive Mode Separation-Based Wavelet Transform. Spectral decomposition-based method. β- distance of Gaussian distribution

Received: July 17, 2021. Revised: August 12, 2022. Accepted: September 11, 2022. Published: October 6, 2022.

1. Introduction

Blind signal separation (BSS) consists to separate source signal from mixed signals without any information. BSS have wide range of applications such as medical imaging and engineering [1-4], image processing and speech recognition [5, 6], and speech signal processing [7, 8], communication systems [9], astrophysics [10], automatic transcription or speech and musical instrument identification [11], mechanical fault detection [12, 13].

In the literature, many approaches have been proposed to solve the BSS problem. The most popular is the independent component analysis method (ICA). In [14] an algorithm based on phase space reconstruction was proposed. In [15], an algorithm composed of pseudo-multiple input multiple output observation structure and independent component analysis (ICA) was proposed. In [16], an improved empirical mode decomposition method for blind separation of single-channel vibration signal mixtures was proposed. The ICA is characterized by simplicity and results quality. ICA technique is based on linear transformation to find components from multidimensional mixed data. The ICA is performed on the hypothesis that the source signals are statistically independent. The founded components are statistically independent too.

A single channel source separation methods overview is presented in [17]. Methods based on spectral representation of the observed signal are usually known as spectral decomposition-based methods. Spectral decomposition-based methods have been introduced by many authors. In [18] nonnegative matrix factorization (NMF) method has been applied on the Short Time Fourier Transform (STFT) representation of a single-channel observed signal, but the method requires the use of an additional training data. In [19], wavelet transforms and a combination of empirical mode decomposition (EMD) and ICA has been proposed, but the wavelet transforms require some predefined basis functions to represent a signal. The EMD and its improved algorithms are empirical, and there is no complete mathematical theory basis [20]. In [21] the bark scale aligned wavelet packet decomposition has been introduced, after the Fourier transform, the Gaussian mixture model (GMM) has been used in separation step. In [22] a combination of various single channel separation methods, a spectral decomposition based techniques and model based methods has been dissected.

In [23] a new Adaptive Mode Separation-Based Wavelet Transform (AMSWT) has been proposed to seismic time–frequency analysis. The novel time-frequency analysis approach is inspired by the adaptive wavelet bank configuration to empirical wavelet transform (EWT) [24-26] and the spectral mode separation thought from variational mode decomposition (VMD) [27]). The AMSWT method consists to adaptively extract spectral intrinsic components by solving a recursive optimization problem. To obtain the spectral boundaries for wavelets bank configuration, the limited support of every spectral mode is introduced. Then, the obtained spectral boundaries for wavelets bank configuration built to highlight the spectral information. The AMSWT method is a fully adaptive approach without requiring prior information.

In [28] a new method to solve the SCBSS problem is proposed. The method is applied on the time-frequency representation of a single-channel observed signal. The ICA-based single-channel separation has been applied on spectral rows corresponding to different time intervals. The β-distance of Gaussian distribution D_β is used to measure the distance between time-frequency domain components of the mixed signal obtained by ICA, and finally, these components are grouped. The grouping algorithm of the components return to solve the optimization problem by minimizing the negentropy of reconstructed constituent signals.

In this paper a new method has been proposed to solve the SCBSS problem. The method is based on combining the AMSWT [23] and the ICA-based single channel separation method [28]. The time-frequency representation of a signal is considered as a multichannel observed signal and can be separated by ICA. After separation, the statistically independent time-frequency components are then grouped. The grouping using the β-distance of Gaussian distribution D_β

The performance of the proposed method is tested on real speech sounds chosen from available databases and compared to the results obtained via EMD based single-channel separation, the wavelets based-single channel separation introduced in [19] and the single-channel separation audio signals based on variational mode decomposition (VMD). The quality of the obtained separation results was evaluated using the scale-invariant (SI) parameters such as SI-SDR, SI-SAR, SI-SIR, which are particularly recommended for single-channel separation evaluation [29, 30].

The remaining content is composed of the following parts: the second section gives the SCBSS problem formulation; the third section introduces adaptive mode separation-based wavelet transform; the fourth section shows the ICA-based single channel separation method; The fifth section present the main steps of the proposed algorithm with the application of this algorithm in the simulation experiments and the comparison results with other algorithms; finally, conclusions and discussions are given in the fifth section.

2. SCBSS Problem Formulation

A general BSS problem can be mathematically defined as follows: Let $\mathbf{s}(t) = [s_1(t),..,s_N(t)]^T$ be a vector of N independent sources at the discrete time instant t. The vector $\mathbf{x}(t) = [x_1(t),..,x_M(t)]^T$of the M observed instantaneous mixtures is modeled as follow:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where $\mathbf{A}$ is the $(M \times N)$ mixing matrix.

In the literature, the main BSS classifications are defined such as: linear and nonlinear BSS; instantaneous and convolutive BSS; over complete and underdetermined BSS. For the last classification, when the number of observed signals M is more than the number of independent sources N, this refers to over complete BSS. On the other hand, when the number of observed signals M is smaller than the number of independent sources N, this becomes to underdetermined BSS.

In general case and for many practical applications only one-channel recording is available. This special case of instantaneous underdetermined source separation problem termed as single channel source separation is discussed in many papers. For this special case, the conventional source separation methods are not suitable.

The SCSS research area where the problem can be simply treated as one observation instantaneous mixed with several unknown sources:

$$x(t) = \sum_{i=1}^{N} a_i s_i(t) \tag{2}$$

where $i = 1,..,N$ denotes number of sources and the goal is to estimate the sources $s_i(t)$ when only the observation signal $x(t)$ is available. In frequency domain, by applying the short time Fourier transform (STFT). The mixture defined in equation (2) becomes:

$$X(f) = \sum_{i=1}^{N} a_i S_i(f) \tag{3}$$

where f denote the frequency. $X(f)$ design the Fourier transform of the mixture signal $x(t)$ and $S(f)$ is a $(Nx1)$ vector whose elements $S_i(f)$ are the Fourier transforms of the source signals $s_i(t)$. Since the separation of the signal is performed frame by frame, the mixing model of each frame can be written as :

$$X(f,m) = \boldsymbol{A}\,\boldsymbol{S}(f,m) \tag{4}$$

where m denotes the frame index.

In [31] the original EMD description, a mode is defined as a signal whose number of local extrema and zero-crossings differ at most by one. In most lately related works, the definition is changed into so-called Intrinsic Mode Functions (IMF), based on modulation criteria [31, 20].

3. Adaptive Mode Separation-based Wavelet Transform

The wavelet $\psi(t)$ is a function localized jointly in time and frequency and with a zero mean. A mother wavelet $\psi_{a,b}(t)$ defined as follow

$$\psi_{a,b}(t) = \frac{1}{\sqrt{b}}\psi(\frac{t-a}{b}) \quad (5)$$

Where a and b denote the translation and dilatation parameters respectively.
The wavelet transform consists to perform the inner product between the family of wavelets $\psi_{a,b}(t)$ and the signal $s(t)$.

$$W_s = \langle s(t), \psi_{a,b}(t) \rangle = \int_{-\infty}^{+\infty} s(t)\frac{1}{\sqrt{b}}\hat{\psi}(\frac{t-a}{b}) \quad (6)$$

The AMSWT perform the time-frequency analysis by the variational scaling and wavelet functions to every mode. So, the method is based on the ADMM [31] solver and then defines a bank of variational scaling functions and wavelets based on the established spectral boundaries.

Therefore, the approximate coefficients and detailed coefficients are obtained by the inner product of the analyzed signal *s* with the variational scaling function, and by the inner product of the analyzed signal *s* with variational wavelets respectively and expressed by the following equations

$$W_s(0,t) = \langle s, \emptyset_1 \rangle = \int s(\tau)\overline{\emptyset}_1(\tau - t)d\tau \quad (7)$$

and

$$W_s(k,t) = \langle s, \psi_k \rangle = \int s(\tau)\overline{\psi}_k(\tau - t)d\tau \quad (8)$$

In [23] the intrinsic modes $u(t)$ have distinguishable features in the frequency domain under the amplitude-modulated frequency-modulated (AM-FM) assumption, using the alternate direction method of multiplier (ADMM) solver, the spectral modes can be adaptively obtained, following how intrinsic mode functions (IMF) are obtained, to estimate compact modes:

$$\min_{u_k,\omega_k} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_K u_k = s(t) \quad (9)$$

Where $s(t)$ 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 empirically determinate using on the following equation:

$$\widetilde{K} = min\{n \in \mathbb{Z}^+ | n \geq 2\rho \ln N\} \quad (10)$$

where N presents the signal length and ρ is the scaling exponent determined by the detrended fluctuation analysis (DFA) [32].

According to [23] the equation is solved using a quadratic penalty term and the parameter λ that denotes the Lagrangian multiplier for rendering the problem unconstrained

$$L(u_k, \omega_k, \lambda) = \eta \sum_k \left\| \delta_t [\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)] e^{j\omega_k t} \right\|_2^2 + \langle \lambda, s - \textstyle\sum_K u_k \rangle + \| s - \textstyle\sum_K u_k \|_2^2. \quad (11)$$

therefore u_k is determined recursively as

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{s}(\omega) - \sum_{i \neq j} \hat{u}_i^{n+1}(\omega) + \frac{\hat{\lambda}^n}{2}}{1 + 2\eta(\omega - \omega_k^n)^2} \quad (12)$$

where $\hat{s}(\omega)$, $\hat{u}_i(\omega)$ and $\hat{\lambda}(\omega)$ denote the Fourier transform of the input signal $s(t)$, 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

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 \, d\omega} \quad (13)$$

Therefore, Instead of using a predefined wavelet bank, we build adaptive wavelets banks using the spectral modes and associated center frequencies represent the intrinsic components.

In [23] authors defined the boundaries between each mode using the mode bandwidth and central frequencies, Whereas, in the literature, some authors are just used the average between the two central frequencies as the spectral boundary, which does not consider the spectral 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

$$\mathbf{\Omega}_k = \frac{\omega_k + \frac{\beta_k}{2} + \omega_{k+1} - \frac{\beta_{k+1}}{2}}{2} \quad (14)$$

we take $\mathbf{\Omega}_k = 0$ and $\mathbf{\Omega}_k = \pi$.
For the variational scaling functions and wavelets based on the spectral boundaries: the authors use the idea used in the construction of both Littlewood–Paley and Meyer's wavelets [33]. $\widehat{\emptyset}_k$ 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.

$$\widehat{\emptyset}_k = \begin{cases} 1, & \omega \leq (1-\gamma)\mathbf{\Omega}_k \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \mathbf{\Omega}_k)\right), & (1-\gamma)\mathbf{\Omega}_k \leq \omega \leq (1+\gamma)\mathbf{\Omega}_k \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

and

$$\hat{\psi}_k = \begin{cases} 1, & (1+\gamma)\mathbf{\Omega}_k \leq \omega \leq (1-\gamma)\mathbf{\Omega}_{k+1} \\ \cos\left(\frac{\pi}{2}\alpha(\gamma, \mathbf{\Omega}_{k+1})\right), & (1-\lambda)\mathbf{\Omega}_{k+1} \leq \omega \leq (1+\lambda)\mathbf{\Omega}_k \\ \sin\left(\frac{\pi}{2}\alpha(\gamma, \mathbf{\Omega}_k)\right), & (1-\lambda)\mathbf{\Omega}_k \leq \omega \leq (1+\lambda)\mathbf{\Omega}_k \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

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

$$\beta(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 1 \\ \beta(x) + \beta(1-x) = 1, & 0 < x < 1 \end{cases} \quad (17)$$

here each
and the full width at half maximum (FWHM) of the peak.

4. ICA-based Single Channel Separation Method

In [28] the authors propose a new method to solve the SCBSS problem. The method is applied on the time-frequency representation of a single-channel observed signal. The time-frequency representation is a non-linear transformation, the use of non-linear ICA would be appropriate, but However, as mentioned in [28, 34] under certain conditions nonlinear BSS problem can be solved using linear ICA.

Let $x(t)$ denote the signal in time domain, using the Short Time Fourier Transform (STFT), the signal is transformed in the frequency domain. The transformation is performed frame by frame and m is the STFT time frame number. The STFT is the mxn complex matrix of time frequency representation, this matrix contain m-rows instantaneous signal spectra,

Let z_i where $i = 1,..,m$ spectral components obtained via the time-frequency representation of a single channel signal. The obtained z_i are statistically independent. In this step, the rows of the TFD^{mix} matrix are treated as individual channels in a multichannel signal. Then the ICA is applied on this multichannel signal.

The ith row of Z denoted z_i can be written as $TFD^i = a^i z_i$ an ith time frequency component of a mixed one-channel signal. The relation between Z where $z = [z_i]$ $i = 1,..,m$ and TFD^{mix} is given as the following equation

$$TFD_{mix} = A.Z = \sum_i a^i z_i = \sum_i TFD^i \quad (18)$$

where A is the $(M \times N)$ mixing matrix whose elements a_{ji} ,where, a^i is an ith column of A.
The z_i present the spectral bases. The columns of A describing time variation of $z_i$are called time bases and denoted by a^i. The matrix TFD^i denote the product of the time basis a^i and the spectral basis z_i is called ith time-frequency component.

The grouping of TFD^i bases is performed into subgroups by the grouping of time bases a^i and frequency bases z_i as the following equation:

$$\begin{aligned} TFD_{mix} &= \sum_i TFD^i \\ &= \sum_{j_1} TFD^{j_1} \\ &+ \sum_{j_2} TFD^{j_2} + \cdots + \sum_{j_p} TFD^{j_p} \end{aligned} \quad (19)$$

Where $j_1,..,j_p$ are p index sets obtained by grouping TFD^i bases.

In [28], to reduce computational complexity, authors used only the TFD^i bases which have a specified variance of the mixed signal. The grouping of bases consists to collecting elements into clusters. The clustering is based on the maximization of negentropy of separated components. The ICA-based single-channel separation methods primarily use component grouping based on similarity in time or frequency domain. In [28] authors suggest the use of a time-frequency structure to measure the similarity features in both time and spectral domain.

5. Grouping Process

The grouping process is performed by clustering the ith time-frequency distribution TFD^i bases, or the distance between TFD^i bases, using the β-distance of Gaussian distribution D_β [28].

The generalized Gaussian distribution is expressed as following:

$$p(y|\mu,\sigma,\beta) = \frac{\omega(\beta)}{\sigma} exp\left[-c(\beta)\left|\frac{y-\mu}{\sigma}\right|^{2/(1+\beta)}\right] \quad (20)$$

where μ denote the expected value. β describes the type of a random variable y, i.e., its deviation from normal distribution where $-1 \leq \beta \leq 0$. σ present the standard deviation of a random variable y. The parameters $\omega(\beta)$ and $c(\beta)$ are given by the following expression: $\omega(\beta) = \frac{\Gamma\left[\frac{3}{2}(1+\beta)\right]^{1/2}}{(1+\beta)\Gamma\left[\frac{1}{2}(1+\beta)\right]^{3/2}}$ and $c(\beta) = \left[\frac{\Gamma\left[\frac{3}{2}(1+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]}\right]^{1/(1+\beta)}$ where Γ is the Gamma-Euler function.

The time-frequency is considered as a random variable, its distribution is given in parametric terms. Therefore, it is possible to estimate the parameters μ, σ , β based on the model defined in Equation (20). The β-distance of Gaussian distribution D_β is defined as the following equation

$$D_\beta = \left|\beta_{i,org} - \beta_i(TFD_{rec,i})\right| \quad (21)$$

the β_i parameter is estimated by a posteriori determination of the maximum of the β, where , the a posteriori distribution of the β parameter is given as [32, 35]: $p(\beta|y) \propto p(y|\beta)p(\beta)$, where $p(y|\beta)$ denotes a data likelihood [32] and is given as the following equation

$$p(y|\beta) = \prod_N \frac{\omega(\beta)}{\sigma} exp\left[-c(\beta)\left|\frac{y_N-\mu}{\sigma}\right|^{2/(1+\beta)}\right] \quad (24)$$

where $p(\beta)$ present the a priori distribution of the β parameter [18, article de khedamy bih].

The statistically independent constituent signals have the maximum negentropy [10,50]. So, the grouping or the TFD^i bases consists in maximizing negentropy (negative entropy) of reconstructed constituent signals $TFD_{rec,i}$ by finding of reconstructed constituent signals $TFD_{rec,i} = \sum_{ji} TFD^{ji}$ with the maximum negentropy, the TFD^i bases can be grouped.

Let ν is the normalized Gaussian random variable $(\mu = 0, \sigma = 1)$ and $G(.)$ is a nonlinear function of the random variable usually having the form $G(y) = \frac{1}{a}\log\cos h\,(ay)$,

$a \in (1,2)$ or $G(y) = -e^{(-\frac{y^2}{2})}$. The negentropy function $J(y)$ is given by the following equationexpression [35]: $J(y) \sim [E(G(y)) - E(G(v))]$. The negentropy function $J(y)$ approximation has numerous advantages such as conceptual simplicity and rapid calculation rate [35]. As a result, it is very often used as a cost function in algorithms for solving ICA problems [28].

6. Results and Discussions

To evaluate the performance of the proposed approach, simulations are performed. The proposed method has been applied on speech datasets selected from TIMIT [36] and NOIZEUS [37] databases. The instantaneous mixture is simulated by the recordings of three sentences $s_1(\mathrm{t}), s_2(\mathrm{t})$ and $s_3(\mathrm{t})$. The signals are pronounced by male and female speakers and were recorded at the sampling frequency F_s. The instantaneous mixture is defined by the following equation:

$$x(t) = a_1\, s_1(\mathrm{t}) + a_2\, s_2(\mathrm{t}) + a_3\, s_3(\mathrm{t}) \qquad (22)$$

where a_1, a_2 and a_3 are constants parameters.

The proposed method operates in the time-frequency domain, and is summarized by the following steps for each frame:

1. Compute the Short Time Fourier Transform (STFT) of the observed signal $x(t)$
2. Apply the variational scaling and wavelet functions to every mode to obtain the time-frequency distribution using equation (7) and (8).
3. The input data for ICA is a spectrogram. The ICA is applied on this multichannel signal (applied on spectral rows corresponding to different time intervals)
4. The β-distance of Gaussian distribution D_β is used to measure the distance between time-frequency domain components of the mixed signal obtained by ICA.
5. Solve the optimization problem by minimizing the negentropy of reconstructed constituent signals.
6. Reconstruct the appearance of the particular source in the original signal.

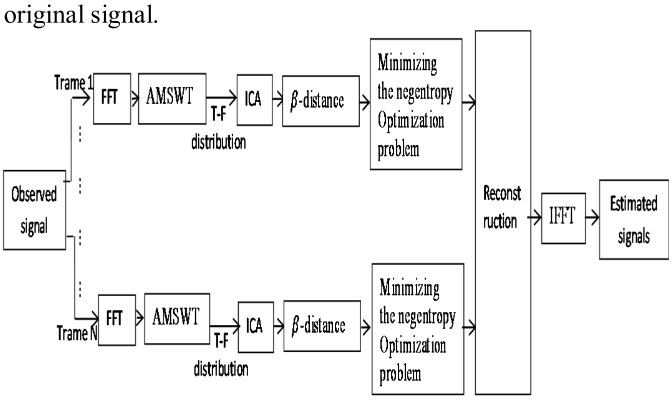
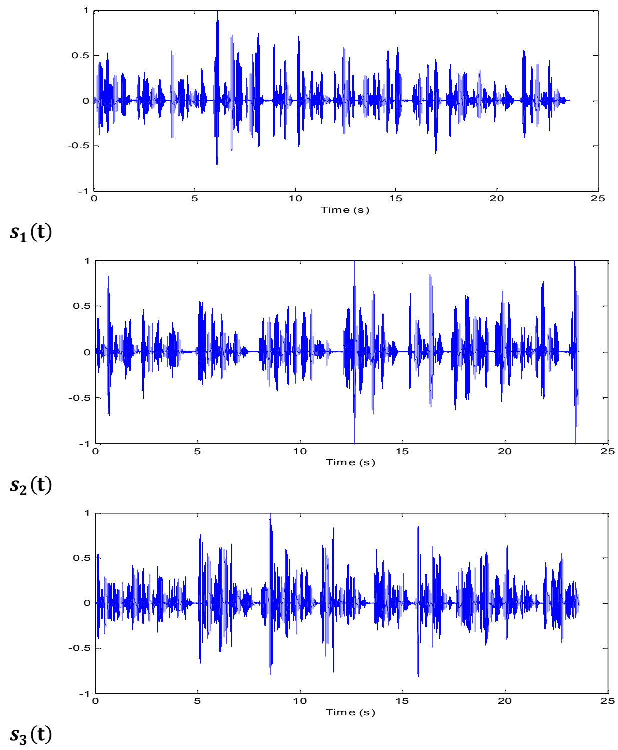


Figure 1. Flowchart of the proposed method

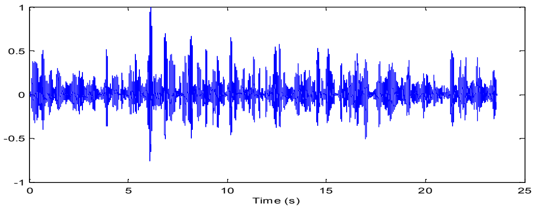
Thereafter, as an illustration example, the proposed method is applied to separate an instantaneous mixture defined by equation (22). The Fig.2 (a) shows the three speech signals representation in time domain. First, the observed single-channel presented in Fig.2 (b) was transformed to the frequency domain using the STFT. The Fig.2 (c) presents the STFT of a frame of the observed mixture. Then, for each frame, the AMSWT method is introduced to obtain optimal spectral mode separation; we apply the variational scaling and wavelet functions to every mode to obtain the time-frequency distribution using equation (7) and (8) as illustrated by Fig. 2 (d).

Once the T-F distribution is obtained, the spectrogram which is considered as a multichannel observed signal is used as the input data for ICA-based single channel separation. Then, as mentioned in step 4 of the algorithm, the β-distance of Gaussian distribution D_β is used to measure the distance between time-frequency domain components. Solving the optimization problem as mentioned in step 5. For our example and for a mode, the estimated spectral components are shown in Fig. 2(e). For this frame, collecting elements into clusters, the estimates frame of the signal $s_1(t)$is illustrated in Fig.2(f). The estimated signals are illustrated in Fig.2(g). As can be seen, the estimated signals were similar to the original signal showed in figure Fig.2 (a)

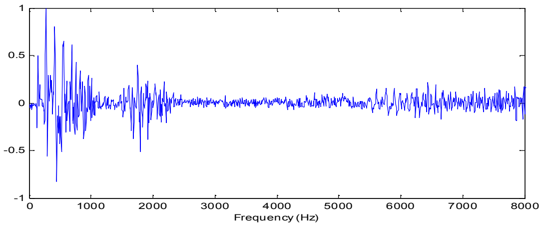


(a) Original sources time-Domain représentation.

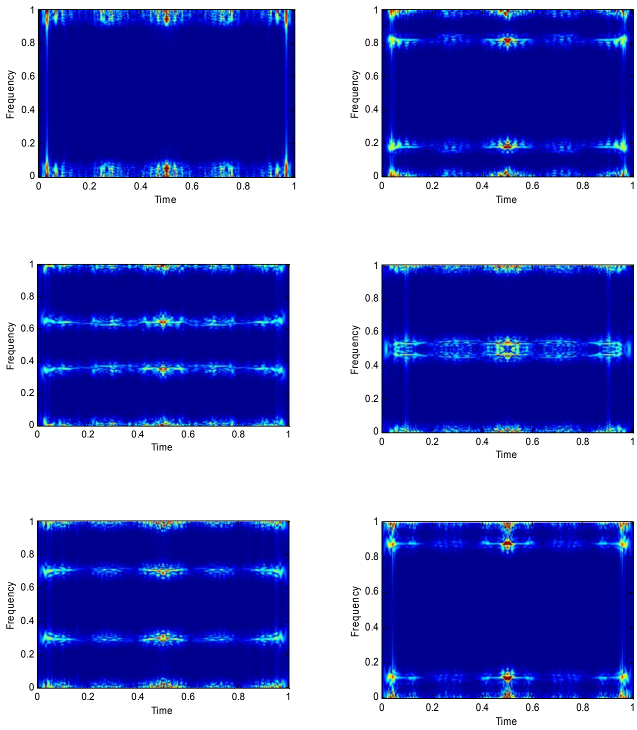
FIGURE 2. ILLUSTRATION EXAMPLE



(B) OBSERVED SIGNAL TIME-DOMAIN REPRESENTATION.

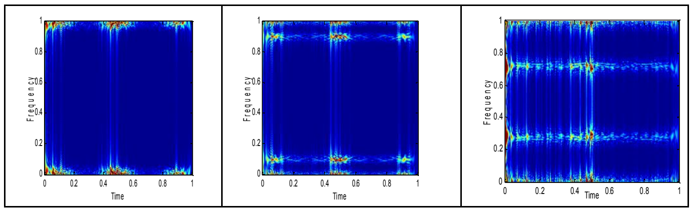


(C) FFT OF THE FRAME OF THE OBSERVED SIGNAL.

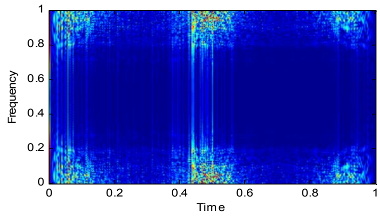


(D) TIME-FREQUENCY DISTRIBUTION OF ONE ARBITER CHOSEN MODE

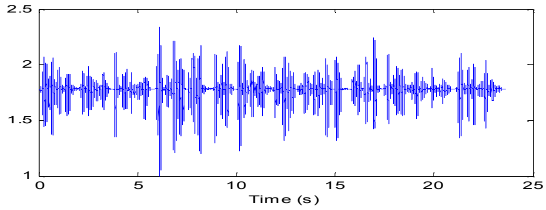
FIGURE 2. CONTINUED



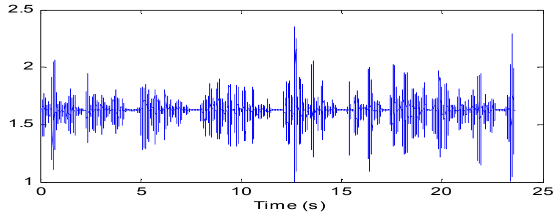
(e) The time-frequency distribution of one arbiter chosen mode.



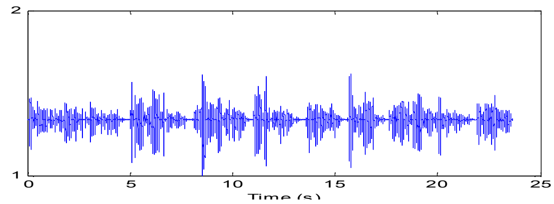
(f) The time-frequency distribution of one arbiter chosen mode



Estimated source $\hat{s}_1(t)$



Estimated source $\hat{s}_2(t)$



Estimated source $\hat{s}_3(t)$

(G) THE TIME-FREQUENCY DISTRIBUTION OF ONE ARBITER CHOSEN MODE

FIGURE 2. CONTINUED

Objective measure of separation quality has been performed. The performances of the proposed method are compared with existing methods in the literature such as the EMD signal-channel separation [19], the wavelets signal-channel separation presented in [19], and the single-channel separation audio signals based on variational mode decomposition (VMD).

In [29, 30] a new method has been proposed, the method is a simpler scale-invariant alternative for single-channel separation evaluation by the introduction a new parameters. These parameters are called scale-invariant (SI) such as SI-SDR, SI-SAR, SI-SIR, and they are particularly recommended single-channel separation evaluation. These parameters are defined by the usage of a single coefficient α to account for scaling discrepancies. Let $s(t)$ is the original sources, and $\hat{s}(t)$ is the estimated source expressed as $\hat{s} = e_{target} + e_{res}$ where e_{res} can be decomposed as $e_{res} = e_{interf} + e_{artif}$, where s_{target} are the source signals, and e_{interf} denotes the interferences from other sources, and e_{artif} includes all other artifacts introduced by the separation algorithm. The $SI-SDR$ is given as the following equation : $SI-SDR=\frac{|\alpha s|^2}{|\alpha s-\hat{s}|^2}$ where $\alpha=\text{argmin}_{\alpha}|\alpha s-\hat{s}|^2$. The optimal scaling factor for the target is obtained as $\alpha=\frac{\hat{s}^T s}{\|s\|^2}$, and the scaled reference is defined as $e_{target}=\alpha s$. The performance criteria are given by the following equation:

$$SI-SDR=10log_{10}\left(\frac{\|e_{target}\|^2}{\|e_{res}\|^2}\right)$$
$$SI-SDR=10log_{10}\left(\frac{\|\alpha s\|^2}{\|\alpha s-\hat{s}\|^2}\right) \quad (23)$$

The scale-invariant signal to interference ratio (SI-SIR) is given by the following equation:

$$SI-SIR=10log_{10}\left(\frac{\|e_{target}\|^2}{\|e_{interf}\|^2}\right) \quad (24)$$

and the scale-invariant signal to artifacts ratio (SI-SAR) is defined as follows:

$$SI-SAR=10log_{10}\left(\frac{\|e_{target}\|^2}{\|e_{artif}\|^2}\right) \quad (25)$$

Another performance measure has been evaluated; the measure is expressed in terms of the relative root mean squared error ($RRMSE$) given by the following equation

$$RRMSE=\frac{RMS\left(s_i(t)-\widehat{s_i}(t)\right)}{RMS\left(s_i(t)\right)}100[\%] \quad (26)$$

where $s_i(t)$ denote the signal we want to extract and $\widehat{s_i}(t)$ is the estimate of the signal. (In our case $i=1,..,3$). The speech dataset is corrupted at a signal-to-noise ratio SNR=5 dB, then the SI-SIR, SI- SAR and SI-SDR are evaluated. The obtained results are showed in the Fig.3. A set of 4 noisy mixtures are simulated by corrupting the clean mixture at a signal-to-noise ratio (SNR) ranging from 5 dB to 20 dB with a step of 5 dB, then the RRMSE is evaluated. The Fig.4 shows the obtained results.

To discuss the relation between the frame length and the β-distance of Gaussian distribution D_β, the mean of the mean β-distance of Gaussian distribution D_β has been evaluated for different frame length (512, 1024, 2048, 4096 frame length), the obtained results are showed in Fig.5

As shown, the proposed method presents a better separation quality then the exiting methods expressed by the scale-invariant SI-SDR, SI-SAR, SI-SIR parameters values. On the other hand, the relative roots mean squared error ($RRMSE$) of the proposed method is better than the $RRMSE$ of the existing methods for different SNR values. The mean β-distance of Gaussian distribution D_β for different frame length shows that the combination of the AMSWT method and the ICA-based single channel separation method allows having better results and better separation compared to existing methods. So, the proposed method allows having better separation results then the exiting methods in the literature. The use of the AMSWT allow to generate a superior time–frequency resolution because the wavelet bank is adaptively built on the intrinsic spectral modes; and the use of a time-frequency structure allows measuring the similarity features in both time and spectral domain also the β-distance of Gaussian distribution is a distance measure based on the knowledge of the statistical nature of spectra of original constituent signals of the mixed signal.

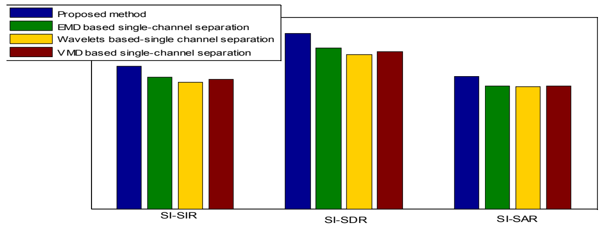


Figure 3. Comparison between the proposed method and the EMD based single-channel separation and the wavelets based-single channel separation and the single-channel separation audio signals based on variational mode decomposition (VMD) in term of SI-SIR, SI-SAR, SI-SDR for SNR=5 dB.

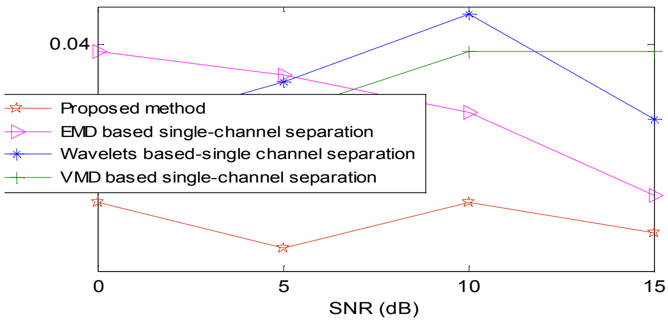


Figure 4. Comparison of algorithms performances for different SNR values

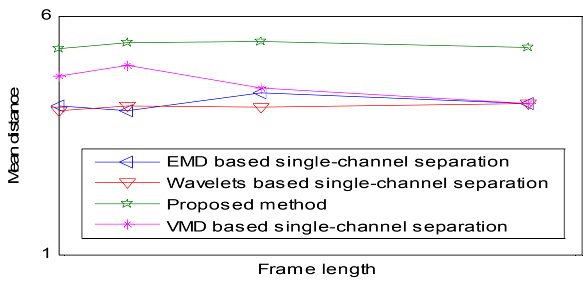


Figure 5. Comparison of the mean β-distance of Gaussian distribution D_β for different frame length.

7. Conclusion

A new method to solve the signal-channel blind source separation problem has been proposed. The method is based on combining two powerful methods such as the Adaptive Mode Separation-Based Wavelet Transform (AMSWT) and the ICA-based single channel separation. A new objective measure of separation quality has been introduced to evaluate the performance of the proposed method. The evaluation parameters are called scale-invariant (SI) such as SI-SDR, SISAR, SI-SIR. Simulation results showed the good performance of the proposed method compared to the exiting method.

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]. James, C.J.; Hesse, C.W. Independent component analysis for biomedical signals. Physiol. Meas. 2005, 26, R15– R39.

[3]. Jimenéz-Gonzaléz, A.; James, C. Source separation of Foetal Heart Sounds and maternal activity from singlechannel phonograms: A temporal independent component analysis approach. In Proceedings of the 2008 Computers in Cardiology, Bologna, Italy, 14–17 September 2008; pp. 949–952.

[4]. 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.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.

[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]. Wilson, S.; Yoon, J. Bayesian ICA-based source separation of Cosmic Microwave Background by a discrete functional approximation. arXiv 2010, arXiv:1011.4018.

[11]. Eronen, A. Musical Instrument Recognition Using ICABased 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.

[12]. R. B. Randall, ``A history of cepstrum analysis and its application to mechanical problems," Mech. Syst. Signal Process., vol. 97, pp. 319, Dec. 2017.

[13]. 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.

[14]. P.He et al., "Single channel blind source separation on the instantaneous mixed signal of multiple dynamic sources," Mechanical Systems & Signal Processing, vol.113, pp. 22-35, December 2018.

[15]. X.Cai et al., "Single Channel Blind Source Separation of Communication Signals Using Pseudo-MIMO Observations," IEEE Communications Letters, vol.22, no.8, pp.1616-1619, Aug 2018.

[16]. D.Wang, W.Guo, and P.W.Tse, "An enhanced empirical mode decomposition method for blind component separation of a singlechannel vibration signal mixture," Journal of Vibration & Control, vol.22, no.11, 2015.

[17]. 39. Gao, B. Single Channel Blind Source Separation. Ph.D. Thesis, Newcastle University, Newcastle, UK, 2011.

[18]. 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.

[19]. 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.

[20]. 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).

[21]. 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.

[22]. 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.

[23]. 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.

[24]. J. Gilles, “Empirical wavelet transform,” IEEE Trans. Signal Process., vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[25]. W. Liu, S. Cao, and Y. Chen, “Seismic time–frequency analysis via empirical wavelet transform,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, pp. 28–32, Jan. 2016.

[26]. N. Liu, Z. Li, F. Sun, Q. Wang, and J. Gao, “The improved empirical wavelet transform and applications to seismic reflection data,” IEEE Geosci. Remote Sens. Lett., to be published.

[27]. [13] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.

[28]. Dariusz Mika, Grzegorz Budzik, and Jerzy Józwik , Single Channel Source Separation with ICA-Based Time- Frequency Decomposition. Sensors.

[29]. J. L. Roux, S.Wisdom, H. Erdogan, and J. R. Hershey, “SDR - half-baked or well done?,” in Proc. IEEE Int. Conf. Acoust. Speech, Signal, Process., 2019, pp. 626– 630.

[30]. M. Torcoli , T. Kastner , and J. Herre. "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence". IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 29, 2021. pp. 1530-1541.

[31]. Konstantin Dragomiretskiy, Dominique Zosso, ‘‘Variational Mode Decomposition ’’. IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 62, NO. 3, FEBRUARY 1, 2014. Pp 531- 544

[32]. C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, “Mosaic organization of DNA nucleotides,” Phys. Rev., vol. 49, no. 2, p. 1685, 1994.

[33]. I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA, USA: SIAM, 1992.

[34]. Isomura, T.; Toyoizumi, T. On the achievability of blind source separation for high-dimensional nonlinear source mixtures. arXiv 2018, arXiv:1808.00668.

[35]. Hyvarinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis; JohnWiley & Sons: New York, NY, USA, 2001.

[36]. TIMIT database. https://catalog.ldc.upenn.edu/ldc93s1

[37]. NOIZEUS database http://ecs.utdallas.edu/loizou/speech/noizeus/