International Journal of Speech Technology (2018) 21:809–823
https://doi.org/10.1007/s10772-018-9544-3

A new efficient backward BSS crosstalk-resistant algorithm for automatic blind speech quality enhancement

Mohamed Djendi[1] · **Meriem Zoulikha**[1]

Received: 26 February 2018 / Accepted: 30 July 2018 / Published online: 13 August 2018

Abstract
In last 10 years, several noise reduction (NR) algorithms have been proposed to be combined with the blind source separation techniques to separate speech and noise signals from blind noisy observations. More often, techniques use voice activity detector (VAD) systems for the optimal solution. In this paper, we propose a new backward blind source separation (BBSS) structure that uses the input correlation properties to provide: (i) high convergence rates and good tracking capabilities, since the acoustic environments imply long and time-variant noise paths, and (ii) low misalignment and robustness against different noise type variations and double-talk. The proposed algorithm has an automatic behavior to enhance noisy speech signals, and do not need any VAD systems to separate speech and noise signals. The obtained results in terms of several objective criteria show the good performance properties of the proposed algorithm in comparison with state-of-the-art algorithms.

Keywords Speech enhancement · Noise reduction · Voice activity detector · BSS · Forward · Backward

1 Introduction

Speech enhancement and acoustic noise reductions applications have been active research fields in the last four decades. The existing speech enhancement techniques aim to improve speech quality by using various algorithms to provide a good convergence speed performance and fast tracking capabilities, since the acoustic environments imply very long and time-variant echo path. A plethora of techniques and algorithms using speech and noise characteristics can be found in the literature (Djendi et al. 2013; Loizou and Kim 2011; Loizou 2013).

Generally the speech enhancement techniques or algorithms can be categorized as single channel, dual channel or multichannel enhancement techniques (Djendi et al. 2009; Ghosh and Tsiartas 2011; Sandoval-Ibarra et al. 2016). Single channel enhancement techniques are used in the situations where only one recorder microphone is available. The single channel speech enhancement techniques still an important field of research because of their simple realization and effectiveness. The single channel is particularly valuable in mobile communication request, where only a single microphone is used due to cost and size constraints (Sandoval-Ibarra et al. 2016). In recent times, several single channel algorithms have been proposed in literature.

Recently in (Upadhyay 2016; Upadhyay and Karmakar 2015), the problem of single channel speech enhancement in stationary environments is discussed and it is proposed the Wiener filtering combined with recursive noise estimation algorithms to enhance speech signals. In Roy et al. (2016), the authors proposed a single channel speech enhancement algorithm using a subband iterative Kalman filter. A wavelet filter bank is first used to decompose the noise corrupted speech into a number of subbands then it is processed by an efficient Kalman filter. In Lee et al. (2017), Cho et al. (2016), the authors proposed new single-channel speech enhancement methods using reconstructive using nonnegative matrix factorization (NMF) with spectro-temporal speech presence probabilities, and outlier detection are also proposed. In order to improve the single channel solution to the problem of speech enhancement, several dual channel and multichannel enhancement techniques have been proposed in literature. For example, several papers have been proposed for dual channel speech enhancement techniques based on

✉ Mohamed Djendi
m_djendi@yahoo.fr

Meriem Zoulikha
m_zoulikha@hotmail.fr

[1] Signal Processing and Image Laboratory (LATSI), University of Blida 1, Route de Soumaa B.P. 270, 09000 Blida, Algeria

the combination between the blind source separation and adaptive filters (Djendi 2010; Ikeda and Sugiyama 1999; Al-Kindi and Dunlop 1989; Gerven and Compernolle 1995). The same dual microphones techniques were used to propose several two-channel or dual adaptive filter that work only on blind noisy speech signals (Sato et al. 2005; Ghribi et al. 2016). We can also cite the machine learning and the active learning techniques and their use in the domain of noisy signal classification and enhancement as given in (Vajda and Santosh 2017; Bouguelia et al. 2018; Zhang et al. 2015). Another direction of research that allows enhancing the speech signal from noisy observations is direction of arrival estimation and localization when multi-speech sources are available (Dey and Ashour 2018a, b, c).

In the approach where multi-channel technique is used for speech enhancement techniques, we can find several technique that are adaptive and not adaptive and all of them aim to improve the single and dual microphones techniques for the same application, i.e. speech enhancement and acoustic noise reduction application. In Marro et al. (1998), the authors concluded that in teleconferencing systems, the use of hands-free sound pick-up reduces speech quality. This is due to ambient noise, acoustic echo, and the reverberation produced by the acoustical environment. The authors of this paper presented a theoretical analysis of noise reduction and dereverberation algorithms based on a microphone array combined with a Wiener post-filter. It is shown that the transfer function of the post-filter depends on the input signal-to-noise ratio (SNR) and on the noise reduction yielded by the array. The use of a directivity-controlled array instead of a conventional beam-former was proposed to improve the performance of the whole system. Several papers based on the multichannel approach were proposed accordingly. Therefore, as multichannel enhancement techniques employ microphone arrays and take advantage of availability of multiple signal inputs to our system, to make possible the use of phase alignment to reject the undesired noise components (Meyer 1997; Lotter et al. 2003; Wang et al. 2016; Mildner and Goetze 2006; Senthamizh Selvi et al. 2017; Qingning and Waleed 2006).

In this paper, we focus our interest on the dual channel approach and we propose a new efficient crosstalk backward blind source separation (BSS) resistant algorithm for automatic blind speech enhancement application. The proposed algorithm is a self-controlled system for automatic speech enhancement application and doesn't need of any voice activity detector to separate speech from very noisy observations.

This paper is organized as follows: after the introduction which is presented in Sect. 1, we present in Sect. 2, the noisy observation model that we adopt in our work. In Sect. 3, we give the principle of backward blind source separation (BSS) structure and two known backward algorithms that are combined with this structure. In Sect. 4, we give the mathematical formulation of the proposed crosstalk backward blind source separation (BSS) resistant algorithm for automatic blind speech enhancement application and its theoretical analysis. In Sect. 5, we show the simulation results of the proposed algorithm in terms of several objective criteria, and finally, in Sect. 6, we conclude our work.

2 Noisy observations model

In this work, we consider two-microphone configurations to make available two noisy observations. The two noisy observations are composed by one speech source signal and one punctual noise. We assume that the speech source signal us placed close to the first microphone, however, the second source of noise is located close to the second microphone (see Fig. 1). The noisy observations of this model are given by the following relations (Ghosh and Tsiartas 2011; Djendi 2010; Gerven and Compernolle 1995):

$$m_1(n) = s(n) + h_{21}(n) \ * b(n) \tag{1}$$

$$m_2(n) = b(n) + h_{12}(n) \ * s(n) \tag{2}$$

The symbol "*" stands for the linear convolution operation. The parameters $h_{12}(n)$ and $h_{21}(n)$ are the cross-coupling effects between the two-channel; s(n) and b(n) are two sources of speech and noise respectively. Note that the sources signals (s(n), b(n)), and the real filters ($h_{12}(n)$, $h_{21}(n)$) are unknown parameters, and only observed signals $m_1(n)$ and $m_2(n)$ are available. In a BSS algorithm, no a priori information are available in the separation process. In practice, we often use the backward BSS (BBSS) structure to retrieve the speech signal from only noisy observations. This BBSS structure is well described in next section.

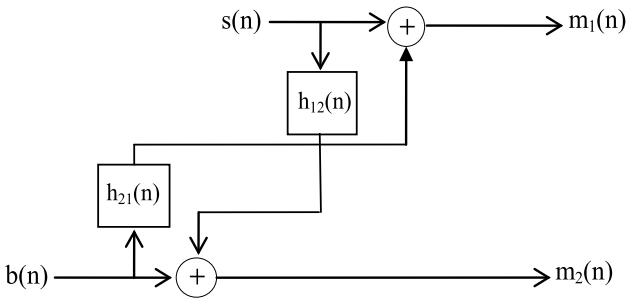

Fig. 1 The simplified mixture model, $s(n)$ and $b(n)$ are the speech signal and the noise respectively. $h_{12}(n)$ and $h_{21}(n)$ represent the impulse responses between the channels

3 Backward BSS (BBSS) structure

The backward blind source separation (BBSS) structure that we consider in this paper is shown by Fig. 2. The noisy input signals of this structure are $m_1(n)$ and $m_2(n)$. The output $s_1(n)$ and $s_2(n)$ of this BSS structure are given by the following equations (Djendi et al. 2013; Djendi 2010; Gerven and Compernolle 1995):

$$s_1(n) = m_1(n) - w_{21}(n) * s_2(n) \tag{3}$$

$$s_2(n) = m_2(n) - w_{12}(n) * s_1(n) \tag{4}$$

Inserting (1) and (2) in (3) and (4) respectively, we get the following outputs signals:

$$s_1(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{12}(n)} * \big(s(n) * \big(\delta(n) - h_{12}(n) * w_{21}(n)\big) + b(n) * \big(h_{21}(n) - w_{21}(n)\big)\big) \tag{5}$$

$$s_2(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{12}(n)} * \big(b(n) * \big(\delta(n) - h_{21}(n) * w_{12}(n)\big) + s(n) * \big(h_{12}(n) - w_{12}(n)\big)\big) \tag{6}$$

To get noise signal at the output $s_2(n)$, and the speech signal at the output $s_1(n)$, we have to satisfied $w_{21}^{opt} = h_{21}$ and $w_{12}^{opt} = h_{12}$). In this case, the outputs of the BBSS structure become as follows $s_1(n) = s(n)$ and $s_2(n) = b(n)$(Djendi et al. 2013).

3.1 Classical backward BSS (CBBSS) two-channel algorithm

In (Gerven and Van Compernolle 1995), the classical backward BSS (CBBSS) two-channel algorithm is used to adjust the coefficients of the two separation filters $w_{12}(n)$ and $w_{21}(n)$. The update relations, in the minimum mean squared error (MMSE) sense, of both adaptive filters $w_{12}(n)$ and $w_{21}(n)$ are given in a vector form as follows:

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \mu_{12} s_2(n)\, \mathbf{k}_1(n) \tag{7}$$

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \mu_{21} s_1(n)\, \mathbf{k}_2(n) \tag{8}$$

where

$$s_1(n) = m_1(n) - \mathbf{w}_{21}^T(n)\, \mathbf{k}_2(n-1) \tag{9}$$

$$s_2(n) = m_2(n) - \mathbf{w}_{12}^T(n)\, \mathbf{k}_1(n) \tag{10}$$

and $\mathbf{k}_1(n) = \big[s_1(n), s_1(n\text{-}1), ..., s_1(n\text{-}L+1)\big]^T$, $\mathbf{k}_2(n) = \big[s_2(n), s_2(n\text{-}1), ..., s_2(n\text{-}L+1)\big]^T$ are vectors that contain the last L sample of the output $s_1(n)$ and $s_2(n)$ respectively. μ_{12} and μ_{21} are respectively the step sizes of the two adaptive filters $w_{12}(n)$ and $w_{21}(n)$, respectively. To ensure stability and convergence of the two-channel CBBSS algorithm toward optimal solutions, the two step-sizes must be selected between 0 and 2 (Djendi 2010; Gerven and Van Compernolle 1995).

A normalized version of this algorithm is obtained by normalizing the step sizes of each adaptive algorithms by $\mathbf{k}_1^T(n)\, \mathbf{k}_1(n)$ and $\mathbf{k}_2^T(n)\, \mathbf{k}_2(n)$ of the two adaptive filters $w_{12}(n)$ and $w_{21}(n)$, respectively. This algorithm allows to take more simple relation for the step-sizes $0 < \mu_{12} < 2$ and $0 < \mu_{21} < 2$.

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \frac{\mu_{21}}{\mathbf{k}_2^T(n)\ \mathbf{k}_2(n) + \xi_1} s_1(n)\, \mathbf{k}_2(n) \tag{11}$$

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \frac{\mu_{12}}{\mathbf{k}_1^T(n)\ \mathbf{k}_1(n) + \xi_2} s_2(n)\, \mathbf{k}_1(n) \tag{12}$$

where ξ_1 and ξ_2 are two small constants introduced to avoid division by zero. The principle of the CBBSS algorithm is similar to the normalized least mean square (NLMS) algorithm in the dual case, this equivalence has been well shown and proven in (Gerven and Van Compernolle 1995). In Table 1, the CBBSS algorithm is summarized.

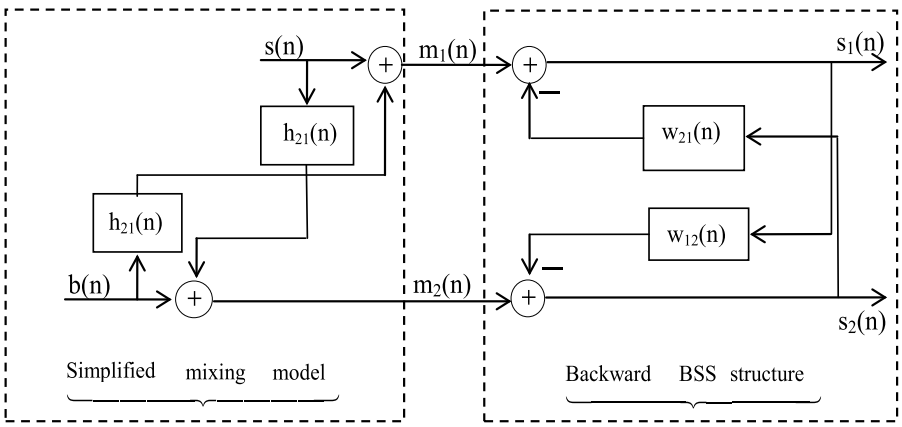

Fig. 2 Backward blind source separation BBSS structure [Left: simplified mixing model], [Right: backward blind source separation (BSS) structure]

4 Proposed robust backward BSS crosstalk-resistant algorithm

4.1 Motivation

In the classical use of the BBSS algorithm, the separating adaptive filters $w_{12}(n)$ and $w_{21}(n)$ have to converge towards the optimal solutions $h_{12}(n)$ and $h_{21}(n)$, respectively, to separate the speech signal and the noise components from the noisy observation $m_1(n)$ and $m_2(n)$ (Djendi et al. 2013; Ghosh and Tsiartas 2011; Djendi 2010). This principle is possible thanks to the use of a voice activity detector (VAD) system. The VAD system allows extracting the source signals from the noisy observation with less distortion (Górriz et al. 2010; Mak 2014; Mukherjee et al. 2018a, b). Usually, the adaptive filters $w_{12}(n)$ and $w_{21}(n)$ are updated alternatively, i.e. if we want to get the speech signal at the output $s_1(n)$, we have to update the adaptive filter $w_{21}(n)$ at only noise presence periods, however the opposite configuration must be adopted for the second adaptive filter $w_{12}(n)$. In this paper, we propose a new automatic BBSS algorithm that update the cross-filters $w_{12}(n)$ and $w_{21}(n)$ automatically and alternatively without need of any VAD system, and is robust for crosstalk presence components.

4.2 Derivation of the proposed algorithm

The mathematic derivation of the proposed algorithm is presented along this section. We recall that the suggested technique principle is based on the use of the intermittent property of the speech signal to adjust the adaptive filter coefficients given by relations (11) and (12) (Djendi et al. 2013). For this reason, we can start from the Newton recurrence (Sayed 2003; Zoulikha and Djendi 2016; Djendi and Zoulikha 2014) applied to the backward blind source separation structure that is given as follows (see Fig. 3):

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21}(n)\frac{\mathbf{p}_{s_1\,\mathbf{k}_2} - \mathbf{R}_{\mathbf{k}_2}\mathbf{w}_{21}(n)}{\zeta_1(n)\,\mathbf{I} + \mathbf{R}_{\mathbf{k}_2}(n)} \tag{13}$$

where $\mathbf{R}_{\mathbf{k}_2}(n)$ represents the autocorrelation matrix of the output vector $\mathbf{k}_2(n)$. It is given by:

$$\mathbf{R}_{\mathbf{k}_2}(n) = E\left[\mathbf{k}_2(n)\,\mathbf{k}_2^{\mathrm{T}}(n)\right] \tag{14}$$

and $\mathbf{P}_{s_1\mathbf{k}_2}(n)$ the cross-correlation vector between the output $s_1(n)$ and the output vector $\mathbf{k}_2(n)$. It is given by:

$$\mathbf{P}_{s1\mathbf{k}2}(n) = E\left[s_1(n)\,\mathbf{k}_2(n)\right] \tag{15}$$

Table 1 Summary of the CBBSS algorithm (16)

CBBSS algorithm steps	Parameters
Initialization	$\boldsymbol{w}_{12}(0) = [\mathbf{0}]^{\mathrm{T}}, \boldsymbol{w}_{21}(0) = [\mathbf{0}]^{\mathrm{T}}, \boldsymbol{k}_1(0) = [\mathbf{0}]^{\mathrm{T}}, \boldsymbol{k}_2(0) = [\mathbf{0}]^{\mathrm{T}}$
	$\mu_{12} = \mu_{21} = 0.98, \xi_1 = \xi_2 = 0.001$
A priori filtering errors	$s_1(n) = m_1(n) - \mathbf{w}_{21}^T(n)\,\mathbf{k}_2(n-1)$
Construction of $\mathbf{k}_1(n)$ and $\mathbf{k}_2(n)$	$\mathbf{k}_1(n) = \left[s_1(n), s_1(n\text{-}1), ..., s_1(n\text{-}L+1)\right]^T$
	$s_2(n) = m_2(n) - \mathbf{w}_{12}^T(n)\,\mathbf{k}_1(n)$
	$\mathbf{k}_2(n) = \left[s_2(n), s_2(n\text{-}1), ..., s_2(n\text{-}L+1)\right]^T$
Cross-filtering updates	$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \frac{\mu_{21}}{\mathbf{k}_2^T(n)\,\mathbf{k}_2(n)+\xi_1} s_1(n)\,\mathbf{k}_2(n)$
	$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \frac{\mu_{12}}{\mathbf{k}_1^T(n)\,\mathbf{k}_1(n)+\xi_2} s_2(n)\,\mathbf{k}_1(n)$

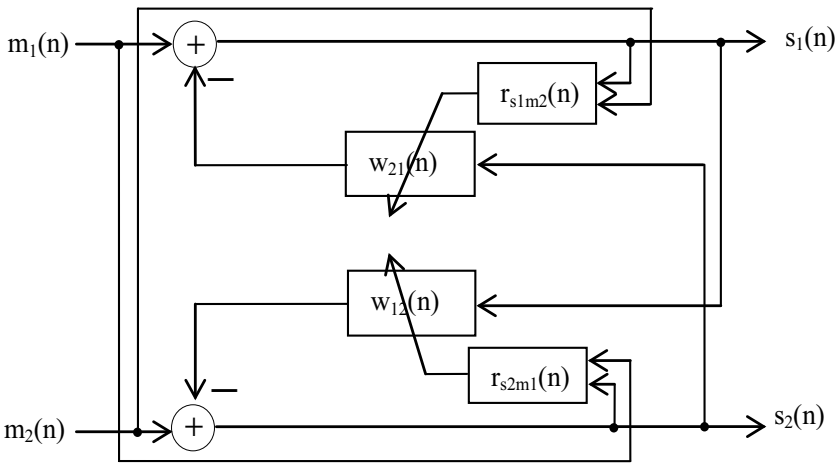

Fig. 3 Proposed algorithm. The new parameters $r_{s1m2}(n)$ and $r_{s2m1}(n)$ are the cross-correlations between the outputs $s_1(n)$ and $s_2(n)$ and the mixing signals $m_1(n)$ and $m_2(n)$ respectively

and $\mathbf{I}$ is the $N \times N$ identity matrix ; and $\zeta_1(n)$ is a small regularization scalar. The step size μ_{21} is a control parameter of relation (13) to ensure stability and convergence. The same thing can be done to relation (12) as follows:

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \mu_{12}(n) \frac{\mathbf{p}_{s_2 \mathbf{k}_1} - \mathbf{R}_{\mathbf{k}_1} \mathbf{w}_{12}(n)}{\zeta_2(n)\, \mathbf{I} + \mathbf{R}_{\mathbf{k}_1}(n)} \tag{16}$$

where $\mathbf{R}_{\mathbf{k}1}(n)$ is the autocorrelation matrix of the output vector $\mathbf{k}_1(n)$, it is given by $\mathbf{R}_{\mathbf{k}_1}(n) = E\left[\mathbf{k}_1(n)\, \mathbf{k}_1^{\mathrm{T}}(n)\right]$. the vector $\mathbf{P}_{s_2 \mathbf{k}_1}$ is the cross-correlation vector between the output $s_2(n)$ and the output vector $\mathbf{k}_1(n)$, it is given by $\mathbf{P}_{s2\mathbf{k}1}(n) = E\left[s_2(n)\, \mathbf{k}_1(n)\right]$. $\zeta_2(n)$ is a small regularization scalar. The step size μ_{12} is a control parameter of relation (16) to ensure stability and convergence.

In general case, the parameters $\zeta_1(n)\, \mathbf{I}$ and $\zeta_2(n)\, \mathbf{I}$ are introduced in the Newton recursion of (13) and (16) to allow regularization of the two-channel algorithm. However, as these two regularization parameter are constant, the behavior of the Newton algorithm applied to (13) and (16) is similar in the transient and permanent regime. The idea is how to change these parameters to get enhancement in either transient or permanent regime. Enhancement of the Newton algorithm in the transient regime is got by improving the convergence speed of the algorithm, hence enhancing the permanent regime is to make the final mean square error (MSE) small, i.e. we want to get a blind two-channel algorithm that has a faster convergence speed and small final MSE.

In this paper we propose to use the cross-correlation vector of the filtering error $s_1(n)$ and the noisy observation $p_2(n)$ instead of $\zeta_1(n)\, \mathbf{I}$ in (13), and the cross-correlation vector of the filtering error $s_2(n)$ and the noisy observation $p_1(n)$ instead of $\zeta_2(n)\, \mathbf{I}$ in (16). These two modifications allow to the Newton algorithm of relation (13) and (16) to be enhanced in the transient and permanent regimes. The new proposed solution of the automatic speech enhancement by the BBSS algorithm is given by the following relation:

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21}(n) \frac{\mathbf{p}_{s_1 \mathbf{k}_2} - \mathbf{R}_{\mathbf{k}_2}(n)\, \mathbf{w}_{21}(n)}{\left(\zeta_1(n) + \left\|\mathrm{r}_{s1\mathbf{m}2}(n)\right\|^2\right) \mathbf{I} + \mathbf{R}_{\mathbf{k}_2}(n)} \tag{17}$$

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \mu_{12}(n) \frac{\mathbf{p}_{s_2 \mathbf{k}_1} - \mathbf{R}_{\mathbf{k}_1}(n)\, \mathbf{w}_{12}(n)}{\left(\zeta_2(n) + \left\|\mathrm{r}_{s2\mathbf{m}1}(n)\right\|^2\right) \mathbf{I} + \mathbf{R}_{\mathbf{k}_1}(n)} \tag{18}$$

where $\mathbf{r}_{s1\mathbf{m}2}(n)$ is the cross-correlation vector of the output signal $s_1(n)$ and the noisy observation vector $\mathbf{m}_2(n)$, and $\mathbf{r}_{s2\mathbf{m}1}(n)$ is the cross-correlation vector computed between the output signal $s_2(n)$ and the noisy observation vector $\mathbf{m}_1(n)$. They are given as follows:

$$\mathbf{r}_{s1m2}(n) = E\left[s_1(k)\, \mathbf{m}_2(k-n)\right] \tag{19}$$

$$\mathbf{r}_{s2m1}(n) = E\left[s_2(k)\, \mathbf{m}_1(k-n)\right] \tag{20}$$

and

$$\left\|\mathbf{r}_{s1\mathbf{m}2}(n)\right\|^2 = \sum_{k=0}^{L-1} \left|\mathrm{r}_{s1\mathbf{m}2}(n-k)\right|^2 \tag{21}$$

$$\left\|\mathbf{r}_{s2\mathbf{m}1}(n)\right\|^2 = \sum_{k=0}^{L-1} \left|\mathrm{r}_{s2\mathbf{m}1}(n-k)\right|^2 \tag{22}$$

where L is a sample number of the cross-correlation vector norm. In the following, we will drive an automatic and less complex algorithm. We will start by relation (17) then we make an extrapolation for relation (18).

Step 1: In first, we introduce a parameter β_1 that allows controlling the contribution of $\left\|\mathrm{r}_{s1\mathbf{m}2}(n)\right\|^2$ in the regularization of (17). Also, we suppose ergodic and stochastic condition that allows to replace $\mathbf{P}_{s_1\mathbf{k}_2}(n)$ and $\mathbf{P}_{s2\mathbf{k}1}(n)$ by their instantaneous values, i.e. $\mathbf{P}_{s1\mathbf{k}2}(n) = \left[s_1(n)\, \mathbf{k}_2(n)\right]$ and $\mathbf{P}_{s2\mathbf{k}1}(n) = \left[s_2(n)\, \mathbf{k}_1(n)\right]$. The new relation of $\mathbf{w}_{21}(n)$ is given as follows:

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \frac{\mu_{21}(n)}{\beta_1\left(\zeta_1(n) + \left\|\mathbf{r}_{s1\mathbf{m}2}(n)\right\|^2\right) \mathbf{I} + \left(1-\beta_1\right) \mathbf{k}_2(n)\, \mathbf{k}_2^{\mathrm{T}}(n)} \mathbf{k}_2(n)\, s_1(n) \tag{23}$$

Step 2: In the second step, in order to reduce the complexity of the algorithm (23), we aim to reduce the complexity of (23) by using the following matrix inverse lemma:

$$[\mathbf{A} + \mathbf{BCD}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left[\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right]^{-1}\mathbf{DA}^{-1} \tag{24}$$

we make the following equality beteween the denomiator of (23) and (24), we get:

$$[\mathbf{A} + \mathbf{BCD}]^{-1} = \left[\beta_1\left(\zeta_1(n) + \left\|\mathbf{r}_{s1\mathbf{m}2}(n)\right\|^2\right) \mathbf{I} + \left(1-\beta_1\right) \mathbf{k}_2(n)\, \mathbf{k}_2^{\mathrm{T}}(n)\right]^{-1} \tag{25}$$

If we put $\mathbf{A} = \beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I}$, $\mathbf{B} = \mathbf{k}_2(n)$, $\mathbf{C} = \left(1-\beta_1\right)$, and $\mathbf{D} = \mathbf{k}_2^{\mathrm{T}}(\mathrm{n})$, and after applying (24), we get the following relation:

$$\begin{aligned}&\left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I} + \left(1-\beta_1\right)\mathbf{k}_2(\mathrm{n})\mathbf{k}_2^{\mathrm{T}}(\mathrm{n})\right]^{-1}\\ &= \left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I}\right]^{-1}\\ &\quad - \left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I}\right]^{-1}\mathbf{k}_2(n)\\ &\quad \times\left[\left(1-\beta_1\right)^{-1} + \mathbf{k}_2^T(n)\left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I}\,\mathbf{k}_2(n)\right]^{-1}\right]^{-1}\\ &\quad \times\mathbf{s}_2^T(n)\left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I}\right]^{-1}\end{aligned} \tag{26}$$

Step 3: More simplification of (26) can be obtained. We multiply both sides of (26) by $\mathbf{k}_2(n)$ and after some modification and rearrangements we get the following simple relation:

$$\begin{aligned}&\left[\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)\mathbf{I} + \left(1-\beta_1\right)\mathbf{k}_2(\mathrm{n})\,\mathbf{k}_2^{\mathrm{T}}(\mathrm{n})\right]^{-1}\mathbf{k}_2(\mathrm{n})\\ &= \frac{\mathbf{k}_2(\mathrm{n})}{\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right) + \left(1-\beta_1\right)\left\|\mathbf{k}_2(\mathrm{n})\right\|^2}\end{aligned} \tag{27}$$

Step 4: If we replace relation (27) in (23) we get the final update relation of the filter $\mathbf{w}_{21}(\mathrm{n})$:

$$\begin{aligned}&\mathbf{w}_{21}(\mathrm{n+1}) = \mathbf{w}_{21}(\mathrm{n})\\ &+ \frac{\mu_{21}(\mathrm{n})}{\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right) + \left(1-\beta_1\right)\left\|\mathbf{k}_2(\mathrm{n})\right\|^2}\mathbf{k}_2(\mathrm{n})\,s_1(\mathrm{n})\end{aligned} \tag{28}$$

In our proposed algorithm, we expoit the symetric property of the backward blind source separation structure to conclude the derivation of the update relation of the adaptive filter $\mathbf{w}_{12}(\mathrm{n})$ and we get:

$$\begin{aligned}&\mathbf{w}_{12}(\mathrm{n+1}) = \mathbf{w}_{12}(\mathrm{n})\\ &+ \frac{\mu_{12}(\mathrm{n})}{\beta_2\left(\zeta_2(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s2m1}}(n)\right\|^2\right) + \left(1-\beta_2\right)\left\|\mathbf{k}_1(\mathrm{n})\right\|^2}\mathbf{k}_1(\mathrm{n})\,s_2(\mathrm{n})\end{aligned} \tag{29}$$

where $\zeta_1(\mathrm{n})$ and $\zeta_2(\mathrm{n})$ are small positive constants β_1, β_2, $\mu_{21}(\mathrm{n})$, and $\mu_{12}(\mathrm{n})$ are control parameters of the proposed algorithm. These last parameters have to be finely selected to accomplish the best tradeoff between faster convergence speed and low final MSE. The proposed algorithm is summarized in Table 2.

4.3 Theoretical analysis of the proposed algorithm

In this analysis, we adopt a new notation of the proposed algorithm of relations (28) and (29). Hence, the new two-channel update of the cross-adaptive filters $\mathbf{w}_{12}(\mathrm{n})$ and $\mathbf{w}_{21}(\mathrm{n})$ of the proposed algorithm can be rewritten as follows:

$$\mathbf{w}_{21}(\mathrm{n+1}) = \mathbf{w}_{21}(\mathrm{n}) + \nabla_1(\mathrm{n})\,\mathbf{k}_2(\mathrm{n})\,s_1(\mathrm{n}) \tag{30}$$

$$\mathbf{w}_{12}(\mathrm{n+1}) = \mathbf{w}_{12}(\mathrm{n}) + \nabla_2(\mathrm{n})\,\mathbf{k}_1(\mathrm{n})\,s_2(\mathrm{n}) \tag{31}$$

where the two new step-sizes $\nabla_1(\mathrm{n})$ and $\nabla_2(\mathrm{n})$ are given by the following relations:

$$\nabla_1(\mathrm{n}) = \frac{\mu_{21}(\mathrm{n})}{\beta_1\left(\zeta_1(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right) + \left(1-\beta_1\right)\left\|\mathbf{k}_2(\mathrm{n})\right\|^2} \tag{32}$$

$$\nabla_2(\mathrm{n}) = \frac{\mu_{12}(\mathrm{n})}{\beta_2\left(\zeta_2(\mathrm{n}) + \left\|\mathbf{r}_{\mathrm{s2m1}}(n)\right\|^2\right) + \left(1-\beta_2\right)\left\|\mathbf{k}_1(\mathrm{n})\right\|^2} \tag{33}$$

Table 2 The proposed algorithm [In this paper]

Proposed algorithm steps	Parameters
Initialization	$\boldsymbol{w}_{12}(0)=[\mathbf{0}]^{\mathrm{T}}$, $\boldsymbol{w}_{21}(0)=[\mathbf{0}]^{\mathrm{T}}$, $\boldsymbol{k}_1(0)=[\mathbf{0}]^{\mathrm{T}}$, $\boldsymbol{k}_2(0)=[\mathbf{0}]^{\mathrm{T}}$ $\mu_{21}(\mathrm{n})=\mu_{12}(\mathrm{n})=0.5$, $\zeta_1(\mathrm{n})=\zeta_2(\mathrm{n})=0.001$, $\beta_1=\beta_2=0.89$
A priori filtering errors	$s_1(n)=\mathrm{m}_1(\mathrm{n})-\mathbf{w}_{21}^T(n)\,\mathbf{k}_2(n-1)$
Construction of $\boldsymbol{k}_1(n)$ and $\boldsymbol{k}_2(n)$ Computation of $\left\|\mathbf{k}_1(\mathrm{n})\right\|^2$, $\left\|\mathbf{k}_2(\mathrm{n})\right\|^2$, $\left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2$, and $\left\|\mathbf{r}_{\mathrm{s2m1}}(n)\right\|$	$\mathbf{k}_1(\mathrm{n})=\left[\mathrm{s}_1(\mathrm{n}),\mathrm{s}_1(\mathrm{n\text{-}1}),\ldots,\mathrm{s}_1(\mathrm{n\text{-}L+1})\right]^T$, $\left\|\mathbf{k}_1(\mathrm{n})\right\|^2=\sum_{\mathrm{k}=0}^{\mathrm{L}-1}\left\|s_1(n-k)\right\|^2$ $s_2(n)=\mathrm{m}_2(\mathrm{n})-\mathbf{w}_{12}^T(n)\,\mathbf{k}_1(n)$ $\mathbf{k}_2(\mathrm{n})=\left[\mathrm{s}_2(\mathrm{n}),\mathrm{s}_2(\mathrm{n\text{-}1}),\ldots,\mathrm{s}_2(\mathrm{n\text{-}L+1})\right]^T$, $\left\|\mathbf{k}_2(\mathrm{n})\right\|^2=\sum_{\mathrm{k}=0}^{\mathrm{L}-1}\left\|s_2(n-k)\right\|^2$ $\mathbf{r}_{\mathrm{s1m2}}(n)=E\left[\mathrm{s}_1(k)\,\mathbf{m}_2(k-n)\right]$, $\left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2=\sum_{\mathrm{k}=0}^{\mathrm{L}-1}\left\|\mathrm{r}_{\mathrm{s1m2}}(n-k)\right\|^2$ $\mathbf{r}_{\mathrm{s2m1}}(n)=E\left[\mathrm{s}_2(k)\,\mathbf{m}_1(k-n)\right]$, $\left\|\mathbf{r}_{\mathrm{s2m1}}(n)\right\|^2=\sum_{\mathrm{k}=0}^{\mathrm{L}-1}\left\|\mathrm{r}_{\mathrm{s2m1}}(n-k)\right\|^2$
Cross-filters update	$\mathbf{w}_{21}(\mathrm{n+1})=\mathbf{w}_{21}(\mathrm{n})+\frac{\mu_{21}(\mathrm{n})}{\beta_1\left(\zeta_1(\mathrm{n})+\left\|\mathbf{r}_{\mathrm{s1m2}}(n)\right\|^2\right)+\left(1-\beta_1\right)\left\|\mathbf{k}_2(\mathrm{n})\right\|^2}\mathbf{k}_2(\mathrm{n})\,s_1(\mathrm{n})$ $\mathbf{w}_{12}(\mathrm{n+1})=\mathbf{w}_{12}(\mathrm{n})+\frac{\mu_{12}(\mathrm{n})}{\beta_2\left(\zeta_2(\mathrm{n})+\left\|\mathbf{r}_{\mathrm{s2m1}}(n)\right\|^2\right)+\left(1-\beta_2\right)\left\|\mathbf{k}_1(\mathrm{n})\right\|^2}\mathbf{k}_1(\mathrm{n})\,s_2(\mathrm{n})$

In order to analysis the behavior of the proposed algorithm, a particular attention is made to the step-sizes of relation (32) and (33). From relation (32), we can note that the step size ∇_1 (n) of the adaptive filter w_{21}(n) is large when the cross-correlation factor $\mathbf{r}_{s1\mathbf{m}2}(n)$ is small, i.e. the step size ∇_1 (n) takes large values when the speech signal is absent, and gets small values in the opposite case. This configuration allows to the adaptive filter w_{21}(n) to be adjusted in the speech absence periods and be frozen in the opposite situation. Furthermore, this automatic mechanism of adjusting the adaptive filter w_{21}(n) allows to formulate an adaptive noise cancellation (ANC) system with noise-only reference, and make possible to cancel the noise components at the output s_1 (n).

In the other hand, an invert relation between the variation of the step-size ∇_2 (n) and the cross-correlation factor $\mathbf{r}_{s2\mathbf{m}1}(n)$ is observed. i.e. the step size ∇_2 (n) is large when $\mathbf{r}_{s2\mathbf{m}1}(n)$ takes small values in speech presence periods. This automatic mechanism allows to the adaptive filter w_{12}(n) to be adjusted to suppress the speech signal at the output s_2 (n) and to get the noise source components in the same output, i.e. s_2 (n).

This automatic mechanism that makes an alternative update of the adaptive filters w_{21}(n) and w_{12}(n), leads to a blind system separation of the speech and the noise components at the outputs s_1 (n) and s_2 (n) without any *priori* information about them, i.e. only the mixing signals are available at the input of the algorithm. A demonstration of these conclusions and theoretical analysis will be given in the simulations part of (See Subsection 5.5).

5 Simulation results

In this section, we analyze the behavior of the proposed algorithm in comparison with two two-channel adaptive BSS-based algorithm, which are the classical BSS (CBBSS) algorithm (16), and the variable step-size backward source separation (VSS-BBSS) algorithm (Djendi and Zoulikha 2014).

5.1 Description of the experimental model and the used signals

We have generated the simulated impulse responses by the model proposed in (Djendi 2010; Ikeda and Sugiyama 1999; Al-Kindi and Dunlop 1989; Gerven and Van Compernolle 1995; Sato et al. 2005; Ghribi et al. 2016; Vajda and Santosh 2017; Bouguelia et al. 2018; Zhang et al. 2015; Dey and Ashour 2018a, b, c; Marro et al. 1998; Meyer 1997; Lotter et al. 2003; Wang et al. 2016; Mildner and Goetze 2006; Senthamizh Selvi et al. 2017; Qingning and Waleed 2006; Vlaj and Kačič 2012; Djendi et al. 2006), i.e. $h_{12}(n) = \delta(n) + \psi_1(n)$ and $h_{21}(n) = \delta(n) + \psi_2$, where $\delta(n)$ is the first sample of the impulse response that represents the direct acoustic path from each source to the cross-coupled microphone. ψ_1 and ψ_2 are exponentially weighted tail that model the room effect (Djendi et al. 2006). Figure 4 shows an example of each impulse responses $h_{12}(n)$ (left of Fig. 4) and $h_{21}(n)$ (right of Fig. 4) that corresponds to spaced microphones; with a sampling period $Ts = 125$ µs, the corresponding reverberation time is 30.8 ms, and the size of the impulse responses is $L = 128$ (Djendi et al. 2006).

The speech and the noises signals are real, sampled at $fs = 8$ kHz, and obtained from AURORA database (Zue et al. 1990; Varga and Steeneken 1993; ITU-T 2003). The noises that we use are White noise, USASI (United State of America Standard Institute now ANSI), street, car and babble. The mixing signals $m_1(n)$ and $m_2(n)$ are generated for different input SNRs, i.e. −6, 0, and 6 dB. We give an example of a speech signal, a noise and mixing ones $m_1(n)$ in Fig. 5. The input SNR is selected to be 0 *dB* at the two microphones, respectively.

5.2 Simulation parameters of the algorithms

In order to objectively compare our proposed algorithm against the performances of two other competitive ones,

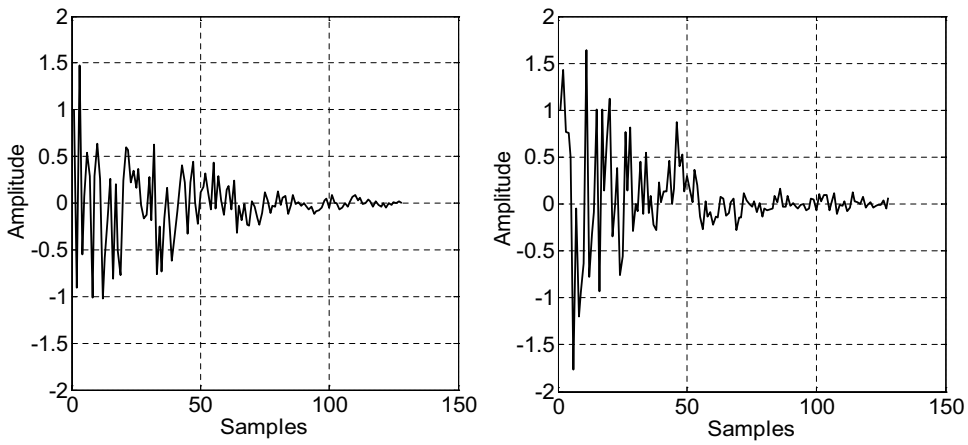

Fig. 4 Simulated impulse responses in the spaced microphones case; [Left]: $h_{12}(n)$, [Right]: $h_{21}(n)$. The real filters length is $L = 128$. $fs = 8$ kHz

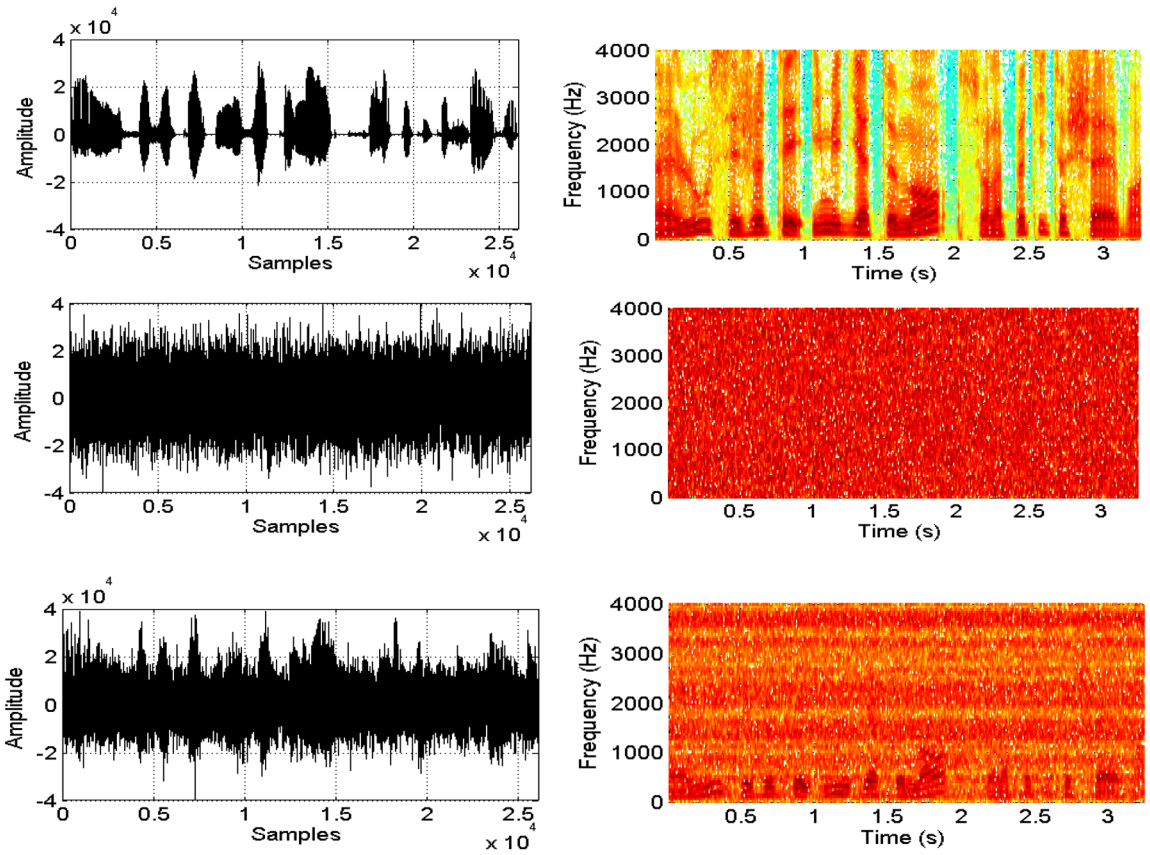

Fig. 5 Source, noise and mixing signal samples. [Top]: the speech signal and its spectrogram. [Middle]: the noise (white) and its spectrogram. [Bottom]: the mixing signal $m_1(n)$ and its spectrogram. The input SNR is selected to be 0 dB at the two microphones, and the real filters length is $L = 128$

i.e. the conventional blind source separation (CBBSS) (Gerven and Van Compernolle 1995), and the variable step-size blind source separation (VSS-BBSS) algorithms (Djendi and Zoulikha 2014), we have selected the best parameters of each algorithm to achieve the best behavior with speech signals. The parameters of each algorithm are summarized in Table 3. We recall here that the CBBSS algorithm (Gerven and Van Compernolle 1995) uses a manual voice activity detector (MVAD) mechanism to control the adaptation of both adaptive estimated filters $w_{12}(n)$ and $w_{21}(n)$, however the VSS-BBSS (Djendi and Zoulikha 2014), which is an improved version of CBBSS, uses a variable step-sizes technique that performs as an automatic voice activity detector (AVAD) mechanism. Recall that the adaptation process of the estimated filters $w_{12}(n)$ and $w_{21}(n)$ by the proposed algorithm is done automatically thanks to the variable step-sizes that are given by relations (32) and (33), respectively. This modification allows to our algorithm to be adapted automatically without need to any VAD system. We note that these parameters are used in all the simulations that we have done and are presented along this paper.

From Table 3, we can see that the proposed and simulated algorithms share some parameters. The shared parameters of these algorithms are the adaptive filters length of $w_{12}(n)$ and $w_{21}(n)$ which is selected to be equal to $L = 128$, or 256 (for more details about these parameters, see Table 3). The considered situation of the simulation is exact modelization of the adaptive filter, i.e. the adaptive filters length is equal to the real ones. The other parameters are specific for each algorithm. Moreover, the control parameters of our algorithm are the optimal ones, and several simulations are carried out to get these optimal values. Finally, we note that control parameters of Table 3 are used along all the carried out simulations and experiments. All the presented simulations are carried out with speech signal and noise components sampled at 8 kHz and coded on 16 bits.

5.3 Time-domain outputs of the proposed algorithm

Simulated and proposed algorithms aim to extract speech at the first output $s_1(n)$ and the noise components at the second output $s_2(n)$. As we are interested on speech enhancement,

Table 3 Control parameters of the conventional BBSS (CBBSS), the variable ste-size BBSS (VSS-BBSS), and the proposed algorithms

Conventionnel and proposed algorithms	Parameters
Conventional BBSS (CBBSS) (Gerven and Van Compernolle 1995)	Length of the filters: $L = 128$ or 256 Step sizes of the filters: $\mu_{12} = 0.4$; $\mu_{21} = 0.4$
Variable step-size BBSS (VSS-BBSS) (Djendi and Zoulikha 2014)	Length of the filters: $L = 128$ or 256 Step sizes of the filters: $\mu_{12} = 0.4$; $\mu_{21} = 0.4$ Step sizes of the sub filters: $\mu_{w_{cont1}} = 0.4$; $\mu_{w_{cont2}} = 0.001$ Step sizes of the main filters: $\mu_{w_{12}}\text{min} = 0$; $\mu_{w_{12}}\text{max} = 0.02$ $\mu_{w_{21}}\text{min} = 0$; $\mu_{w_{21}}\text{max} = 0.2$ Signal-to-noise ratios threshold $\text{SNR}_{1\text{min}} = -40$ dB; $\text{SNR}_{1\text{max}} = -15$ dB $\text{SNR}_{2\text{min}} = -1$ dB; $\text{SNR}_{2\text{max}} = 6$ dB
Proposed algorithm (In this paper)	Length of the filters: $L = 128$ or 256 Step sizes of the main filters: $\mu_{12} = 0.4$; $\mu_{21} = 0.4$ $\beta_1 = 0.5$; $\beta_2 = 0.5$ $\alpha_1 = 0.5$; $\alpha_2 = 0.5$

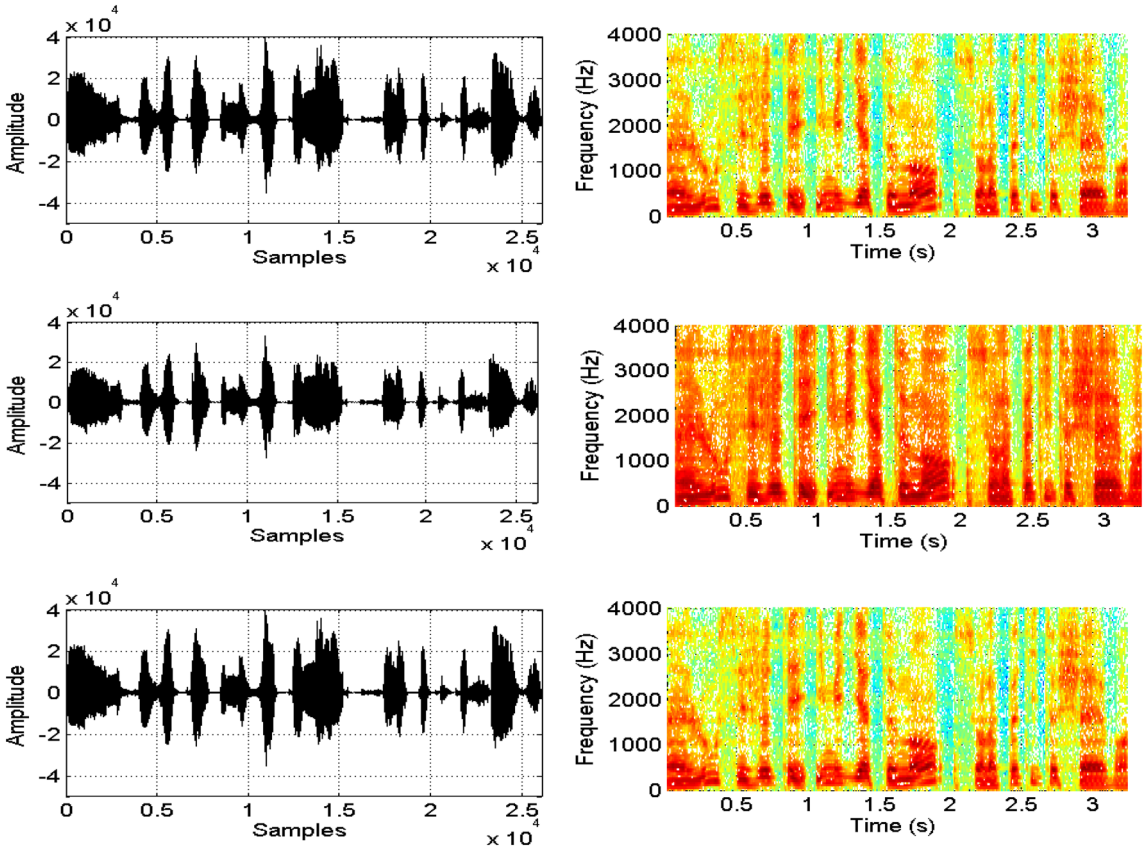

Fig. 6 The output speech signals of, [Top]: the CBBSS, [middle]: the VSS-BBSS and [Bottom]: the proposed algorithm. Each output has its spectrogram in the right. $L = 256$

we only focus on the output $s_1(n)$ and the behavior of the adaptive cross-filter $w_{21}(n)$. In Fig. 6, we illustrate the output $s_1(n)$ of the proposed algorithm, CBBSS and VSS-BBSS algorithms with the parameters of Table 3. This figure shows the good performance of each algorithm in reducing the acoustic noise components at the output $s_1(n)$. No further performance comparisons between the algorithms can be done according to this figure.

5.4 Evaluation of the system mismatch (SM) criterion

The system mismatch (SM) criterion is often used to evaluate the convergence speed performance behavior of any algorithm. The SM criterion evaluates the distance between the estimated adaptive filtering coefficients and the real ones. As we are interested only on the output $s_1(n)$, we focus on the adaptive filter $\boldsymbol{w}_{21}(n)$ and we compute the SM by the following relation (Hu and Loizou 2008):

$$SM_{\text{dB}} = 10\, log_{10}\left(\frac{\left\|\boldsymbol{h}_{21} - \boldsymbol{w}_{21}(n)\right\|^2}{\left\|\boldsymbol{h}_{21}\right\|^2}\right) \tag{34}$$

where h_{21} is the real impulse response, and the symbol $\|\cdot\|$ is the mathematical Euclidean norm. We have done much experiments to evaluate the SM criterion of the three algorithms, i.e. CBBSS, VSS-BBSS, and the proposed RBBSS. The real and adaptive filters length is the same equal to $L = 128$, and 256. Four noise types from AURORA database (Zue et al. 1990) are used, i.e. white, USASI, babble, and street. The obtained results by CBBSS, VSS-BBSS and our proposed algorithms are represented on Fig. 7 for inputs SNR equal to − 3 dB at the two microphones, respectively. We can easily see, from this figure, the superiority of our proposed algorithm in terms of convergence speed performance in comparison with the other ones. We have used the same control parameters of each algorithm as given in Table 3, and the same input signals as explained in Sect. 5.1.

5.5 Step-sizes analysis of the proposed algorithm

In order to analysis the behavior of the proposed algorithm, and as we are interested on speech enhancement problem at

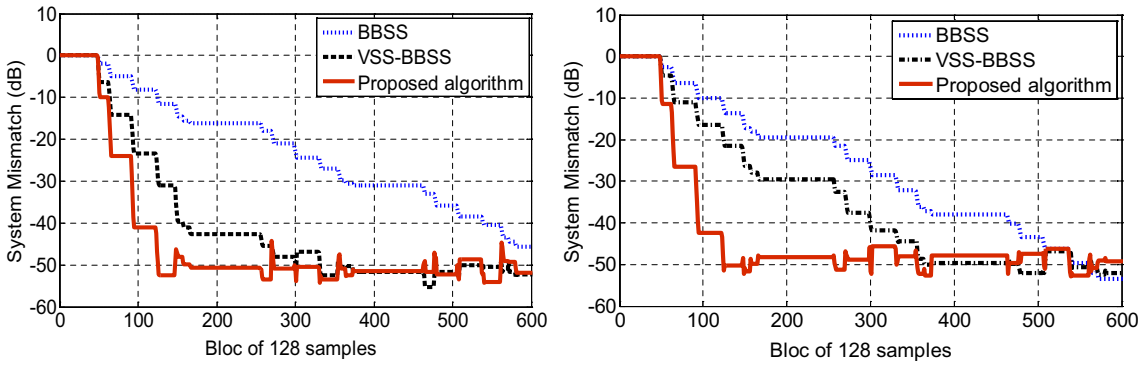

Fig. 7 The system mismatch (SM) comparison between the CBBSS, VSS-BBSS and the proposed algorithms for $L = 128$ [In left], and $L = 256$ [in right]. The parameters of each algorithm are given in Table 3

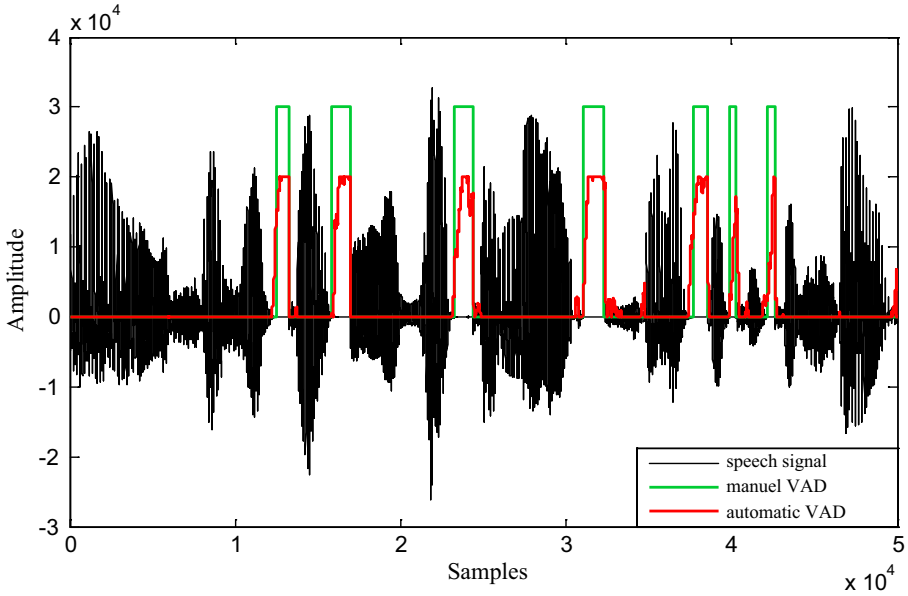

Fig. 8 Original speech signal (in black), Manual VAD (in green), and the automatic VAD obtained by relation (32) [in red]. The control parameters are the same as given in Table 3 for the proposed algorithm. The adaptive and real filter length is $L = 128$. (Color figure online)

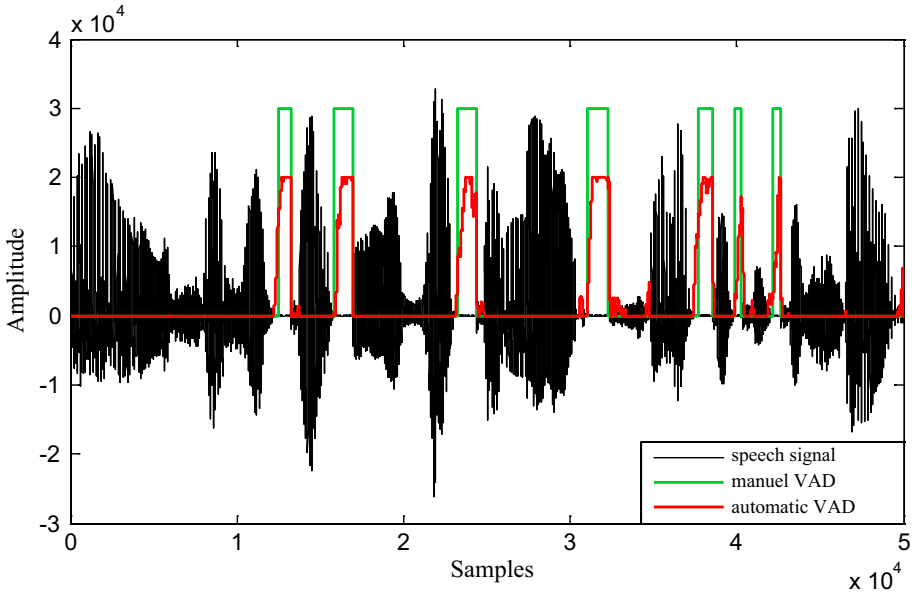

Fig. 9 Original speech signal (in black), Manual VAD (in green), and the automatic VAD obtained by relation (32) [in red]. The control parameters are the same as given in Table 3 for the proposed algorithm. The adaptive and real filter length is $L=256$. (Color figure online)

the output s_1(n), we will focus our interest on relation (32) and its evolution in time domain. Under the same simulation conditions of Sects. (5.2) and (5.3), we have shown the evolution of the step-size ∇_1 (n). In Figs. 8 and 9, we give the time evolutions of the step-size of relation (32) in the cases of two values of the adaptive filters L, i.e. $L=128$, and 256. On the same figures, we show the input speech signal.

From Fig. 8 (for $L=128$), and 9 (for $L=256$), we can observe that the step size ∇_1 (n) of the filter w_{21}(n) is large when the cross-correlation factor $\mathbf{r}_{s1m2}(n)$ is small, i.e. the step size ∇_1 (n) takes large values when the speech signal is absent, and gets small values in the opposite case. This configuration allows to the filter w_{21}(n) to be adjusted in the speech absence periods and be frozen in the opposite situation. This automatic mechanism of adjusting the adaptive filter w_{21}(n) allows to formulate an adaptive noise cancellation (ANC) system with noise-only reference, and make possible to cancel the noise components at the output s_1 (n). In the other hand, an invert relation between the variation of the step-size ∇_2 (n) and the cross-correlation factor $\mathbf{r}_{s2m1}(n)$ is concluded, i.e. ∇_2 (n) is large when $\mathbf{r}_{s2m1}(n)$ takes small values in speech presence periods. This automatic mechanism allows to the adaptive filter w_{12}(n) to be adjusted to suppress the speech signal at the output s_2 (n) and to get the noise source components in the same output, i.e. s_2 (n). This automatic mechanism that makes an alternative update of the adaptive filters w_{21}(n) and w_{12}(n), leads to a blind system separation of the speech and the noise components at the outputs s_1 (n) and s_2 (n) respectively, without any a priori information about them, i.e. only the mixing signals are available at the proposed algorithm inputs.

5.6 Evaluation of the cepstral distance (CD) criterion

The cepstral distance (CD) criterion is used in this Section to quantify the output speech signal processing distortion of each algorithm, i.e. CBBSS, VSS-BBSS and the proposed algorithm. The CD criterion is evaluated by the log-spectrum distance between the original speech signal $s(n)$ and the output speech signal $s_1(n)$ of each algorithm (Hu and Loizou 2008). The CD is computed only in speech presence periods and is given by the following relation:

$$CD_{dB} = \frac{10}{M}\sum_{m=0}^{M-1} log10 \sum_{n=Tm}^{Tm+T-1} \left(cp_s(n) - cp_{e_1}(n)\right)^2 \qquad (35)$$

where $cp_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S(\omega)|e^{j\omega n}d\omega$ and $cp_{s1}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|S_1(\omega)|e^{j\omega n}d\omega$ are the nth real cepstral coefficients of the signals $s(n)$ and$s_1(n)$, respectively. We recall here that $S(\omega)$ and $S_1(\omega)$ are the short Fourier transform (SFTF) of the original speech signal $s(n)$ and the enhanced one $s_1(n)$, respectively. $'T'$ is the mean averaging value of the CD criterion and $'M'$ represents the number of segment where only speech is present. We have estimated the CD criterion for three inputs SNRs at the two microphones are -6 dB, 0 dB and 6 dB. In addition, we have used four types of noise components from AURORA database (Zue et al. 1990; Varga and Steeneken 1993; ITU-T 2003) to generate the noisy observations, which are white, USASI, babble, and street noises. The simulation parameters of ach algorithm are similar to those of the previous experiments and are also

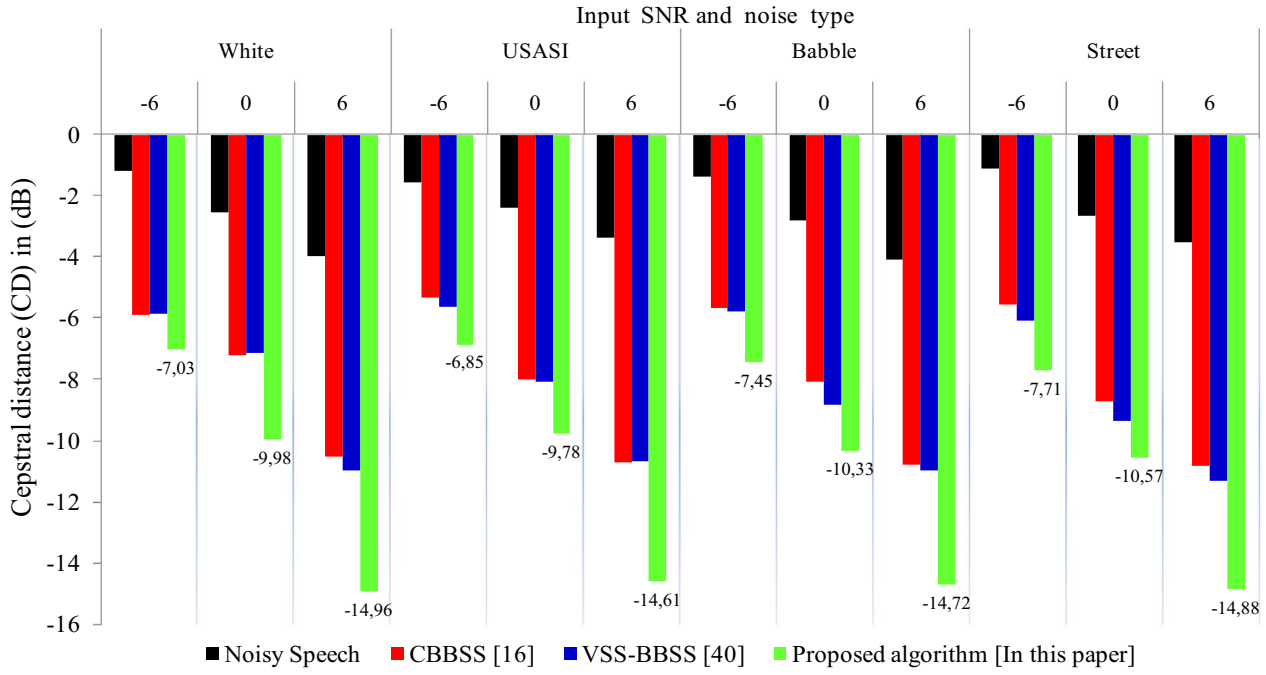

Fig. 10 The cepstral distance (CD) evaluation by: (1) BBSS algorithm, (2) the VSS-BSS algorithm, and (3) the proposed algorithm. The simulation parameters of each algorithm are the same as reported on Table 3 except the length of the adaptive filters $L=128$. The input SNRs are −6 dB, 0 dB and 6 dB

summarized in Table 3. The obtained results of the CD criterion are reported on Fig. 10.

The obtained results of Fig. 10 show clearly the efficiency of the proposed RBBSS algorithm in providing an output speech signal that is very close to the original one and with minimal spectral distortions. Also, we have noted that the proposed algorithm is the one that alters less the speech signal in comparison with the other ones.

5.7 Evaluation of the segmental SNR (SegSNR) criterion

In this section, we analyze the noise reduction performance of the proposed algorithm in terms of segmental signal to noise (SegSNR) criterion. The SegSNR criterion is computed on frames of $'N'$ samples between the original speech signal $s(n)$ and its enhanced version for each algorithm $s_1(n)$. This SegSNR criterion is estimated as follows (Sayed 2003; Zoulikha and Djendi 2016):

$$SegSNR_{dB} = \frac{10}{M}\sum_{m=0}^{M-1} log_{10}\left(\frac{\sum_{n=Nm}^{Nm+N-1}|s(n)|^2}{\sum_{n=Nm}^{Nm+N-1}\left|s(n)-s_1(n)\right|^2}\right) \quad (36)$$

where the parameters $'M'$ and $'N'$ are the number of frame and the frame length, respectively. We note that at the output, we get $'M'$ values of the *SegSNR* criterion, each one is mean averaged on $'N'$ samples. The symbol $|\ .|$ stands for the absolute operator. We recall here that all the $'M'$ frames correspond to only speech signal presence periods. The log_{10} symbol is the base 10 logarithm. The simulation parameters are the same as given in Table 3. We have evaluated the SegSNR criterion for three inputs SNRs, i.e. −6 dB, 0 dB and6 dB. Moreover, four types of noise are used to generate the noisy observations. These noise components which are white, USASI, babble, and street noises are taken from AURORA database (Zue et al. 1990; Varga and Steeneken 1993; ITU-T 2003). The obtained results are reported on Fig. 11.

According to the obtained results, we can easily see that the proposed RBBSS algorithm behaves more efficiency than the other algorithms, and leads to an important SNR at the output. This means that the proposed algorithm suppresses more noise at the output in comparison with the state-of-the-art algorithms, i.e. CBBSS, and VSS-BBSS algorithms. We also conclude that the proposed algorithm has a good performance in different situations when correlated and uncorrelated noises are present. At the end, we can claim that the obtained SegSNR results are another proof performances superiority of the proposed algorithm when combined with BSS structure to restore speech source signal in blind situation when no a priori informations are available about the target.

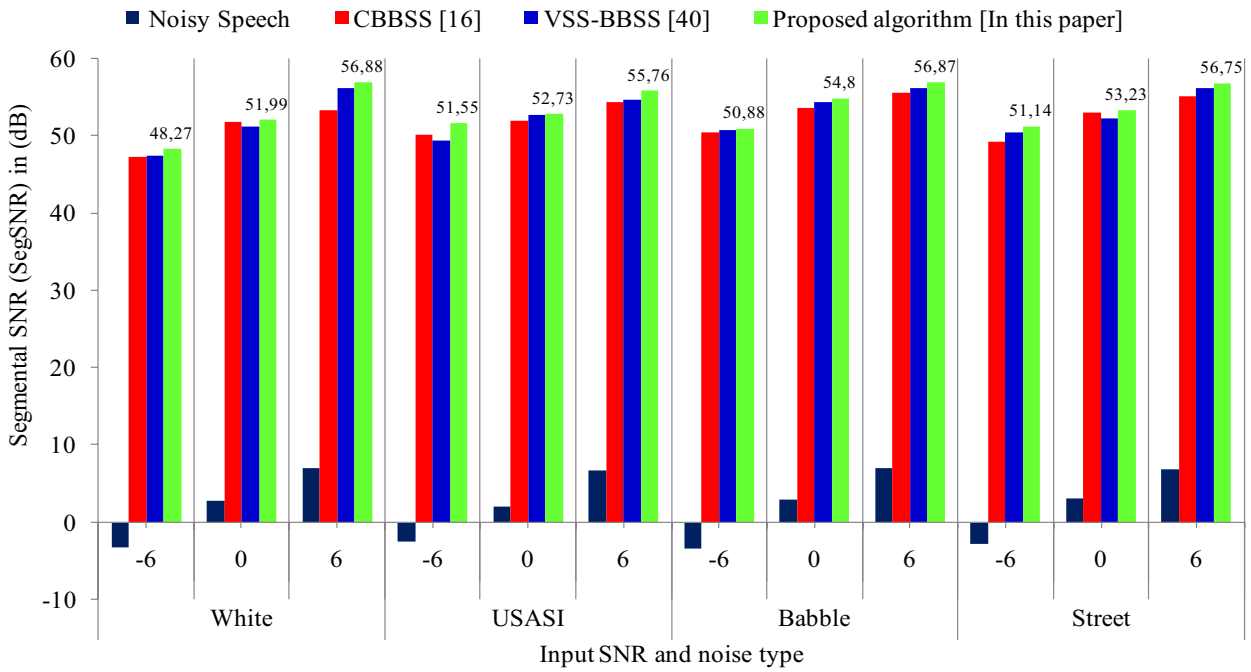

Fig. 11 The Segmental SNR (SegSNR) criterion evaluation by: (1) CBBSS algorithm, (2) the VSS-BSS algorithm, and (3) the proposed algorithm. The parameters of the simulation are the same as reported on Table 3 except the length of the adaptive filter $L = 128$. The input SNRs are -6 dB, 0 dB and 6 dB

6 Conclusion

In this paper, we have proposed a new approach for speech enhancement application. The proposed approach is adaptive and based on the combination between a new automatic adaptive algorithm with the backward blind source separation structure, and allows to automatically adjust the coefficients of the cross-filters.

Intensive experiments were conducted to validate the performance of the proposed algorithm in comparison with two state-of-the-art algorithms, i.e. the classical BBSS and its variable step-size version (VSS-BBSS). The obtained results, expressed in terms of system mismatch, have shown that the proposed algorithm converges quickly to the optimal solutions and this behavior is obtained thanks to the normalization by the norm of the output filtering errors. The obtained CD values have confirmed that the proposed algorithm does not distort the output speech signal especially in the case of loosely spaced microphones (about -14 dB of minimum CD values). The SegSNR results have also shown that the proposed algorithm reduces the acoustic noise components by about 50 dB at the output in several input SNR conditions. The residual noise amount is very small in the case of our proposed algorithm and it do not affect the speech intelligibility at the output.

Finally, we conclude that all the obtained results in terms of CD and SegSNR criteria have shown the superiority of the proposed algorithm in comparison with the other ones. The obtained results have proven the efficiency of the proposed algorithm and show that it can be a good candidate and alternative for speech enhancement and acoustic noise reduction applications. As a future work, the proposed algorithm can be combined with active learning techniques to be used for live stream audio (speech) analysis and can be the one of the contemporary issues in the domain.

References

Al-Kindi, M. J., & Dunlop, J. (1989). Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel. *Signal Processing, 17*(3), 241–250.

Bouguelia, M. R., Nowaczyk, S., Santosh, K. C., & Verikas, A. (2018). Agreeing to disagree: Active learning with noisy labels without crowdsourcing. *International Journal of Machine Learning and Cybernetics, 9*(8), 1307–1319.

Cho, E., Lee, B., & Schafer, R., Widrow, B. (2016). Single channel speech enhancement using outlier detection. Computer Science. https://arxiv.org/pdf/1605.01329.pdf

Dey, N., Ashour, A. S. (2018a). Challenges and future perspectives in speech-sources direction of arrival estimation and localization. In *Direction of arrival estimation and localization of multi-speech sources. SpringerBriefs in electrical and computer engineering* (pp. 49–52). Cham: Springer.

Dey, N., & Ashour, A. S. (2018b). *Direction of arrival estimation and localization of multi-speech sources. SpringerBriefs in Speech Technology*. Cham: Springer.

Dey, N., & Ashour, A. S. (2018c). Applied examples and applications of localization and tracking problem of multiple speech sources. In *Direction of arrival estimation and localization of multi-speech sources. SpringerBriefs in Electrical and Computer Engineering* (pp. 35–48). Cham: Springer.

Djendi, M., Scalart, P., & Gilloire, A. (2006). Noise cancellation using two closely spaced microphones: Experimental study with a specific model and two adaptive algorithms. In *Proceedings of ICASSP*, Vol. 3, pp. 744–747.

Djendi, M. Advanced techniques for two-microphone noise reduction in mobile communications, Ph.D. Dissertation (in French). University of Rennes 1. France 2010, n°19012010.

Djendi, M., Scalart, P., & Gilloire, A. (2013). Analysis of two-sensors forward BSS structure with post-filters in the presence of coherent and incoherent noise. *Speech Communication, 55*(10), 975–987.

Djendi, M., Scalart, P., Gilloire, A. (2009). Comparative study of new blind source separation structures for two-channel acoustic noise cancellation. In *Proceedings of the IEEE*, Glasgow, Scotland, pp. 24–28.

Djendi, M., & Zoulikha, M. (2014). New automatic forward and backward blind sources. Separation algorithms for noise reduction and speech enhancement. *Computer and Electrical Engineering, 40*, 2072–2088.

Fukuda, T., Ichikawa, O., & Nishimura, M. (2010). Long-term spectro-temporal and static harmonic features for voice activity detection. *IEEE Journal on Selected Topics in Signal Processing, 4*(5), 834–844.

Ghosh, P. K., & Tsiartas, A., Narayanan, S. (2011). Robust voice activity detection using long-term signal variability. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(3), 600–613.

Ghribi, K., Djendi, M., & Berkani, D. (2016). A New wavelet-based forward BSS algorithm for acoustic noise reduction and speech quality enhancement. *Applied Acoustics, 105*, 55–66.

Górriz, J. M., Ramírez, J., Lang, E. W., Puntonet, C. G., & Turias, I. (2010). Improved likelihood ratio test based voice activity detector applied to speech recognition. *Speech Communication, 52*(7–8), 664–677.

Hu, Y., & Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech and Language Processing, 16*(1), 229–238.

Ikeda, S., & Sugiyama, A. (1999). An adaptive noise canceller with low signal distortion in the present of crosstalk. In *IEICE Transactions on Fundamentals*, Vol. 82.a, No. 8.

ITU-T P.835.2003. (2003). Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm. ITU-T Recommendation, p. 835.

Lee, S., Han, D. K., & Ko, H. (2017). Single-channel speech enhancement method using reconstructive NMF with spectrotemporal speech presence probabilities. *Applied Acoustics, 117*(B), 257–262.

Loizou, P. C. (2013). *Speech enhancement: Theory and practice* (2nd Ed.). Boca Raton: Taylor & Francis.

Loizou, P. C., & Kim, G. (2011). Reasons why current speech-enhancement algorithms do not improve speech inelligibility and suggested solutions. *IEEE Transactions on Audio, Speech, and Language Processing. 19*(1), 47–56.

Lotter, T., Benien, C., & Vary, P. (2003). Multichannel speech enhancement using Bayesian spectral amplitude estimation. In *Proceedings of ICASSP*, Hong-Kong, pp. 20–24.

Mak, M. W., Yu, H. B. (2014). A study of voice activity detection techniques for NIST speaker recognition evaluations. *Computer Speech and Language, 28*(1), 295–313.

Marro, C., Mahieux, Y., & Simmer, K. U. (1998). Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering. *IEEE Transactions on Speech and Audio Processing, 6*(3), 240–259.

Meyer, J., Uwe, K. (1997). Simmer multi-channel speech enhancement in a car environment using wiener filtering and spectral subtraction. In Proceedings of ICASSP, IEEE, pp. 1–4.

Mildner, V., Goetze, S., Kammeyer, K.-D. (2006). Multi-channel speech enhancement using a psychoacoustic approach for a post-filter. In *Proceedings of ITG-Fachtagung Sprachkommunikation*, Kiel, Germany, pp. 1–4.

Mukherjee, H., Obaidullah, S. M., & Phadikar., S. (2018a). MISNA—A musical instrument segregation system from noisy audio with LPCC-S features and extreme learning. *Multemedia Tools Applications*. https://doi.org/10.1007/s11042-018-5993-6.

Mukherjee, H., Obaidullah, S. M., Santosh, K. C. (2018b). Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. *International Journal on Speech Technology*, https://doi.org/10.1007/s10772-018-9525-6.

Qingning, Z., & Waleed, A. (2006). Speech enhancement by multi-channel crosstalk resistant adaptive noise cancellation. In *Proceedings of IEEE ICASS*, Vol. 1, pp. 485–488.

Roy, S. K., Zhu, W. P., & Champagne, B. (2016). Single channel speech enhancement using subband iterative Kalman filter. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 22–26.

Sandoval-Ibarra, Y., Diaz-Ramirez, V. H., & Kober, V. I. (2016). Speech enhancement with adaptive spectral estimators. *Journal of Communications Technology and Electronics. 61*(6), 672–678.

Sato, M., Sugiyama, A., & Ohnaka, A. (2005). An adaptive noise canceller with low signal-distortion based on variable step size sub filter for human-robot communication. In *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. e88-a, No. 8, pp. 2055–2061.

Sayed, A. H. (2003). *Fundamentals of adaptive filtering*. New York: Wiley.

Senthamizh Selvi, R., & Suresh, G. R., Kanaga Suba Raj, S. (2017). Speech enhancement using harmonic-model with multichannel Wiener Filter. *Journal of Advanced Research in Dynamical and Control Systems, 9*(3), 48–54.

Upadhyay, N., Jaiswal, K. (2016). Single channel speech enhancement: Using Wiener filtering with recursive noise estimation. *Procedia Computer Science, 84*, 22–30.

Upadhyay, N., & Karmakar, A. (2015). Speech Enhancement using spectral subtraction-type algorithms: A comparison and simulation study. In *Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015). Procdia Computer Science*. Vol. 4, pp. 574–584.

Vajda, S., & Santosh, K. C. (2017). A fast k-nearest neighbor classifier using unsupervised clustering. In *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2016. Communications in Computer and Information Science*, Vol. 709, pp. 185–193. Singapore: Springer.

Van Gerven, S., & Van Compernolle, D. (1995). Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness. *IEEE Transactions on Signal Processing, 74*(3), 1602–1612.

Varga, A., & Steeneken, H. J. (1993). Assessment for automatic speech recognition: II. Noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication, 12*(3), 247–251.

Vlaj, D., Kačič, Z., & Kos, M. (2012). Voice activity detection algorithm using nonlinear spectral weights, hangover and hang before criteria. *Computers and Electrical Engineering, 38*(6), 1820–1836.

Wang, X., Guo, Y., Fu, Q., & Yan, Y. (2016). Speech enhancement using multi-channel post-filtering with modified signal presence

probability in reverberant environment. *Chinese Journal of Electronics, 25*(3), 512–519.
Zhang, J., Wu, X., & Shengs, V. S. (2015). Active learning with imbalanced multiple noisy labeling. *IEEE Transactions on Cybernetics, 45*(5), 1095–1107.
Zoulikha, M., & Djendi, M. (2016). A new regularized forward blind source separation algorithm for automatic speech quality enhancement. *Applied Acoustics, 112*, 192–200.
Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication, 9*(4), 351–356.