International Journal of Speech Technology (2018) 21:355–367
https://doi.org/10.1007/s10772-018-9514-9

An efficient wavelet-based adaptive filtering algorithm for automatic blind speech enhancement

Mohamed Djendi[1]

Received: 21 December 2017 / Accepted: 12 April 2018 / Published online: 21 April 2018

Abstract

In this paper, we address the problem of speech enhancement by adaptive filtering algorithms. A particular attention has been paid to the backward blind source separation (BBSS) algorithm and its use in crosstalk resistant speech enhancement applications. In this paper, we propose to implement the BBSS algorithm in the wavelet-domain. The proposed backward wavelet BBSS (WBBSS) algorithm is then used in speech enhancement application when important crosstalk interferences are presents. The new WBBSS algorithm shows better performances in terms of convergence speed and steady state in comparison with the classical BBSS one. The performances properties of the proposed algorithm are evaluated in term of segmental SNR (SegSNR), segmental mean square error (SegMSE), and cepstral distance (CD) criteria. The obtained results have confirmed the best performance of the proposed WBBSS algorithm in a lot of situations when blind noisy observations are available.

Keywords Speech enhancement · Adaptive algorithm · Segmental SNR · Cepstral distance

Abbreviations

BSS	Blind source separation
BBSS	Backward blind source separation
DFT	Discrete Fourier transform
DWT	Discrete wavelet transform
WBBSS	Wavelet transform of BBSS
ANC	Adaptive noise cancellation
LMS	Least mean square
NLMS	Normalized LMS
TSNR	Two-step noise reduction
MSE	Mean square error
SNR	Signal to noise ratio
SegSNR	Segmental signal to noise ratio
MSE	Mean square error
SegMSE	Segmental mean square error
CD	Cepstral distance
dB	Decibel
VAD	Voice activity detector
E	Expectation operator
H	Mixing matrix
W	Unmixing matrix
m	Delay index
n	Discrete time index
J	DWT scale index
L	Real and adaptive impulse responses length
M	Mean averaging value of CD, SegSNR, SegMSE
f_s	Sampling frequency

Parameters

s(n)	Speech signal
b(n)	Punctual noise
$m_1(n)$	First noisy observation
$m_2(n)$	Second noisy observation
$h_{11}(n)$ and $h_{22}(n)$	Direct impulse responses
$h_{12}(n)$ and $h_{21}(n)$	Cross-coupling impulse responses
$\delta(n)$	Dirac impulse
$v_1(n)$	Estimated speech by forward structure
$v_2(n)$	Estimated noise by forward structure
$w_{12}(n)$ and $w_{21}(n)$	Adaptive coefficients
$\mathbf{w}_{12}(n)$ and $\mathbf{w}_{21}(n)$	Adaptive filter vectors
$\mathbf{P}^{(1)}_{J,K}(n)$	Discrete wavelet transform of $m_1(n)$
$\mathbf{P}^{(2)}_{J,K}(n)$	Discrete wavelet transform of $m_2(n)$
$r_{v_1 v_2}(m)$	Cross-correlation between $v_1(n)$ and $v_2(n)$

✉ Mohamed Djendi
m_djendi@yahoo.fr; m_djendi@univ-blida.dz

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

$r_{v_2v_1}(m)$	Cross-correlation between $v_2(n)$ and $v_1(n)$
θ_1 and θ_2	Fixed step-sizes of BBSS
μ_{12} and μ_{21}	Fixed step-sizes of WBBSS
ς_1 and ς_2	Small positive constant
$\phi(n)$	Discret wavelet function

1 Introduction

Various speech enhancement and acoustic noise reduction techniques have been developed in the previous years as speech enhancement is a core target in many challenging areas such as telecommunications, speech and speaker recognitions, teleconferencing and hand-free telephony (Loizou 2007). In such applications, we aim to recover a speech signal from observations corrupted by different noises components. The unusual noise components can be of various classes that are often present in the environment (Djendi et al. 2013).

Numerous algorithms and techniques were proposed to resolve the problem of corrupted speech signals (Dixit et al. 2014; Bactor and Garg 2012; Scalart and Filho 1996). Moreover, techniques of single or multi-microphones are proposed to improve the behavior of the speech enhancement algorithms and also to reduce the acoustic noise components even in very noisy conditions. The most popular single channel techniques that are widely known in speech enhancement application is the spectrum subtraction (SS) that needs only one channel signal (Boll 1979). It has been embedded in some high-quality mobile phones for the same application. However, the SS technique is only suitable for stationary noise environments. Moreover, it certainly introduces music noise problem. In fact, the higher the noise is reduced, the greater the distortion is brought to the speech signal and accordingly the poorer the intelligibility of the enhanced speech is obtained (Zhang and Zhao 2012; Cappé 1994). As a result, ideal enhancement can hardly be achieved when the signal-to-noise-ratio (SNR) of the noisy speech is relatively low; below 5 dB. In contrast, it has quite good result when the noisy speech SNR is relatively high; above 15 dB.

The SS and others speech enhancement techniques that are based on SS principal have improved the decision directed (DD) techniques in reducing the musical noise components (Ephraim and Malah 1984, 1985; Ephraim et al. 2014; Selva Nidhyananthan et al. 2014). Several and recent algorithms that improve the DD techniques are proposed in (Wolfe and Godsill 2003). In Dong et al. (2009), a speech enhancement algorithm based on high-order cumulant parameter estimation is proposed. In (Davila 1984; Doclo and Moonen 2002), a subspace method, which is based on well-known singular value decomposition (SVD) techniques is proposed; the signal is enhanced when the noise subspace is removed, and accordingly, the clean speech signal is estimated from the noisy speech subspace.

Another approach that has been largely studied in speech enhancement application is the adaptive noise cancellation (ANC) approach that was firstly proposed in (Widrow and Goodlin 1975; Widrow 1985). Furthermore, most important speech enhancement techniques and algorithm use adaptive approaches to get the tracking ability of non-stationary noise properties (Lee and Gan 2004; Weinstein et al. 1993). Several adaptive algorithms have been proposed for speech enhancement application, we can find time domain algorithm (21), frequency domain adaptive algorithms (Plapous et al. 2004; Al-Kindi and Dunlop 1989; Plapous et al. 2005; Djendi et al. 2015, 2016) or adaptive spatial filtering techniques (Tong et al.2015; Goldsworthy 2014) that often use adaptive SVD techniques to separates the speech signal space from the noisy one.

Another direction of research combines the blind source separation techniques with adaptive filtering algorithms to enhance the speech signal and to cancel efficiently the acoustic echo components (Van Gerven and Compernolle 1992; Djendi et al. 2006; Zoulikha and Djendi 2016; Jin et al. 2014). This approach uses at least two microphones configuration to update the adaptive filtering algorithms. Also, a multi-microphone speech enhancement approach has been proposed for the same purpose and have improves the existing one-channel and two-channel speech enhancement and noise reduction adaptive algorithms (Benesty and Cohen 2017; Lee and Dae Na 2016). We can also find several papers that highlighted the problem of speech enhancement on a simple problem of mixing and unmixing signals with convolutive and instantaneous noisy observations (Jutten and Herrault 1991; Nguyen Thi and Jutten 1995; Mansour et al. 1996).

In last ten years, a new direction of research have proven the efficiency of the wavelet domain as a new efficient mean that can improves the techniques of speech enhancement, and several techniques and algorithms have been proposed for the same purpose (Bouzid and Ellouze 2016; Ghribi et al. 2016). In this paper, we propose a new implementation of the backward blind source separation (BBSS) algorithm in the wavelet domain.

The presentation of this paper is given as follows: after the introduction, we present, in Sect. 2, the noisy observations generation process that we use in this paper. Section 3 is reserved to describe the classical and original backward blind source separation (BBSS) algorithm. Section 4 presents the proposed wavelet transform domain of the BBSS (WBBSS) algorithm that works in the wavelet domain and its mathematical formulation and derivation. In Sect. 5, the obtained results of the proposed WBBSS algorithm are presented, and finally we conclude our work in Sect. 6.

2 Noisy observations

In this work, we consider two noisy observations $\text{m}_1(\text{n})$ and $\text{m}_2(\text{n})$ available at the output of two sources and two sensors configuration mixing model. The mixing model uses two uncorrelated sources s(n) and b(n) that are the speech signal and the noise respectively. These two signals are considered real and statistically independent, i.e. $E[s(n)\,b(n-m)] = 0$. The two mixing signals $m_1(\text{n})$ and $m_2(n)$ at the sensor outputs of this model are given as follows (Djendi et al. 2013):

$$m_1(\text{n}) = \text{s(n)} + \mathbf{h}_{21}^T(n)\,\mathbf{b}\,(n) \tag{1}$$

$$m_2(n) = b(n) + \mathbf{h}_{12}^T(n)\,\mathbf{s}\,(n) \tag{2}$$

where '*' is the convolution operator. $\mathbf{h}_{12}(n) = \left[h_{12}^{[1]}(n)\,...\,h_{12}^{[L]}(n)\right]^T$ and $\mathbf{h}_{21}(n) = \left[h_{21}^{[1]}(n)\,...\,h_{21}^{[L]}(n)\right]^T$ represent the cross-coupling effects between the channels, and L is their length. The vectors $\mathbf{b}(n) = [b(n)...b(n-L+1)]^T$ and $\mathbf{s}(n) = [s(n)\,...\,s(n-L+1)]^T$ are the punctual noise and the speech source signals. This mixing model allows to say the the first microphone is close to the speech signal, and the punctual noise is close to the second microphone. This configuration is very realistic as in a car or in an office (Djendi et al. 2013).

3 Description of the BBSS algorithm

The general block diagram configuration of the multi-inputs multi-outputs (MIMO) and simplified two-inputs two-outputs (TITO) backward blind source separation (BBSS) algorithm is given in Fig. 1. The aim of this approach is to recover the original sources estimates by using only noisy observations.

In this paper, we study a particular two-inputs two-outputs (TITO or 2×2) BBSS algorithm that aims to separate the original sources estimates of one source signal s(n) and one punctual noise $b(n)$ by using only two available noisy observed signals $\text{m}_1(\text{n})$ and $\text{m}_2(\text{n})$ at the input as given by Fig. 2.

The TITO BBSS algorithm is mainly based on the assumptions that the source signals, $s(n)$ and $b(n)$, are statistically independents, and the mixing and unmixing models H and W respectively, are linear systems (Plapous et al. 2004). In a previous work, we have published a work on the implementation of the forward blind source separation (FBSS) algorithm in the wavelet-domain (Djendi et al. 2006; Mansour et al. 1996). However, in this paper, we will focus our interest on the BBSS algorithm of Fig. 3 and its implementation in the wavelet-domain.

In the blind source separation model of Fig. 2, the output signals $v_1(n)$ and $v_2(n)$ of the BBSS algorithm are given by the following relations (Djendi et al. 2006; Zoulikha et al. 2014):

$$v_1(\text{n}) = m_1(n) - \mathbf{w}_{21}^T(n)\,\mathbf{v}_2(n-1) \tag{3}$$

$$v_2(\text{n}) = m_2(n) - \mathbf{w}_{12}^T(n)\,\mathbf{v}_1(n-1) \tag{4}$$

where $\mathbf{v}_1(n) = \left[v_1(n)...v_1(n-L+1)\right]^T$ and $\mathbf{v}_2(n) = \left[v_2(n)\,...v_2(n-L+1)\right]^T$ are the two outputs vectors of the BBSS algorithm. The vectors $\mathbf{w}_{12}(n) = \left[w_{12}^{[1]}(n)\,...\,w_{12}^{[L]}(n)\right]^T$ and $\mathbf{w}_{21}(n) = \left[w_{21}^{[1]}(n)\,...\,w_{21}^{[L]}(n)\right]^T$ are the adaptive cross-filters of the BBSS algorithm. The output signals $\text{v}_1(\text{n})$ and $v_2(n)$ of Fig. 2 are obtained by inserting relations (1) and (2) into (3) and (4), respectively. In order to get the original speech

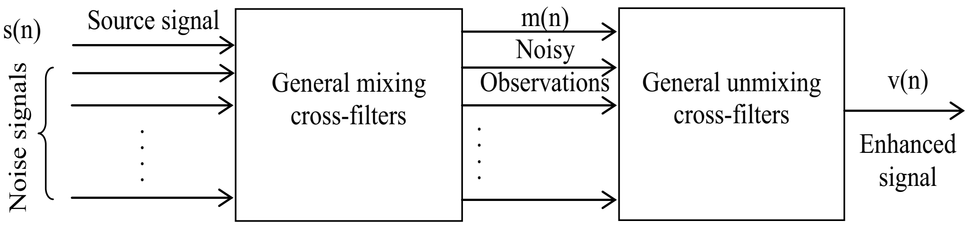

Fig. 1 Block diagram of the BBSS algorithm in the case of MIMO configuration

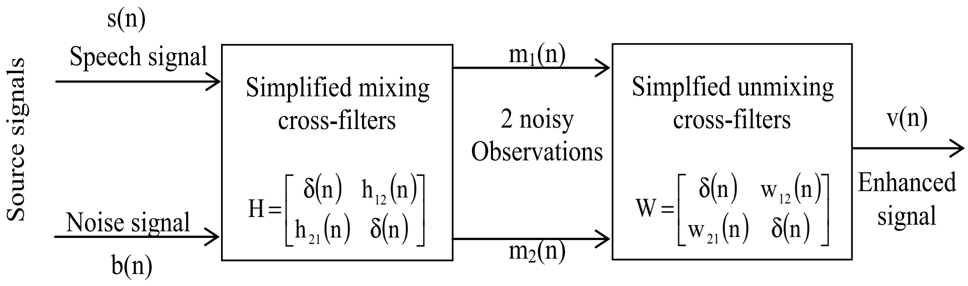
s(n)
Speech signal
Source signals
Noise signal
b(n)
Simplified mixing cross-filters
$H = \begin{bmatrix} \delta(n) & h_{12}(n) \\ h_{21}(n) & \delta(n) \end{bmatrix}$
m₁(n)
2 noisy Observations
m₂(n)
Simplfied unmixing cross-filters
$W = \begin{bmatrix} \delta(n) & w_{12}(n) \\ w_{21}(n) & \delta(n) \end{bmatrix}$
v(n)
Enhanced signal

Fig. 2 Block diagram of the BBSS algorithm in the case of TITO configuration

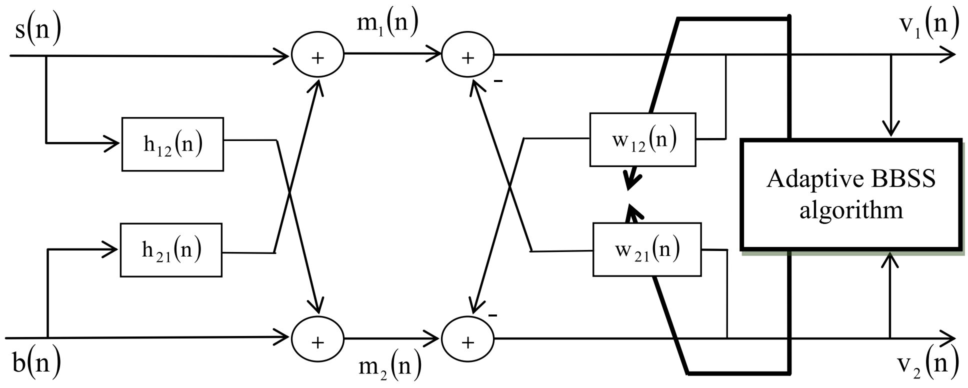

Fig. 3 Backward blind source separation (BBSS) algorithm. Two inputs and two output (TITO) configuration

signal $s(n)$ at the output $v_1(n)$ and the noise component $b(n)$ at the output v_2(n) without distortions, we have to use a voice activity detector (VAD) that updates alternatively the cross-filters $w_{21}(n)$ and $w_{12}(n)$. We recall here that the optimal solutions of the adaptive filters are obtained when $w_{12}(n) = h_{12}(n)$ and $w_{21}(n) = h_{21}(n)$ and we get:

$$v_1(n) = s(n) \tag{5}$$

$$v_2(n) = b(n) \tag{6}$$

According to Fig. 2, and by using a vector notation, the BBSS algorithm updates the cross-filters $w_{21}(n)$ and $w_{12}(n)$ by the following relations:

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + 2\theta_1 \frac{v_2(n)\ \mathbf{v}_1(n)}{\mathbf{v}_1^T(n)\ \mathbf{v}_1(n) + \varsigma_1} \tag{7}$$

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + 2\theta_2 \frac{v_1(n)\ \mathbf{v}_2(n)}{\mathbf{v}_2^T(n)\ \mathbf{v}_2(n) + \varsigma_2} \tag{8}$$

where the scalar θ_1 and θ_2 are the step sizes of the BBSS algorithm and selected as follows: $0<\theta_1<1$, $0<\theta_2<1$. The scalar ς_1 and ς_2 are small constants used to avoid division by zeros. In this paper, we propose to use automatic VAD technique that is inspired from (Zoulikha and Djendi 2016). We recall here that we can alternate the obtained original source signals by inverting the control of the adaptive filters. In Table 1, we summarize the BBSS algorithm.

4 Proposed WBBSS algorithm

The proposed Wavelet-domain implementation of the BBSS (WBBSS) algorithm is given in Fig. 4. We use the discrete wavelet transform (DWT) to convert the mixing signals m_1(n) and m_2(n) into wavelet domain with several number of scales then we apply the BBSS structure. The proposed WBBSS algorithm uses approximations and details that are generated by DWT of m_1(n) and m_2(n) to update two cross-adaptive filters, w_{12}(n) and w_{21}(n). The two cross-filters are therefore used to enhances the speech signal at the output v_1(n). To reconstruct the temporal version of the enhanced speech signal, we apply the inverse DWT at the output. We recall here that we propose to use a new voice activity detector (VAD) system to control the adaptive filter w_{21}(n) to get the speech signal at the output v_1(n) because we focus only on the speech enhancement signal recovered at the first output v_1(n).

4.1 Derivation of the proposed WBBSS algorithm

A detailed scheme of the proposed WBBSS algorithm is given in Fig. 5. The formulation of the proposed WBBSS algorithm is done in several steps.

In the first, we apply discrete wavelet transform (DWT) of scale N to the two mixing signal vectors $\mathbf{m}_1$(n) and $\mathbf{m}_2$(n). If the scale DWT number is denoted N, we can write for{J, K} = 0, 1, 2, 3, ..., N-1.:

$$M^{(1)}_{\mathrm{J,\,K}}(\mathrm{n}) = \sum_{\mathrm{J,\,k}} m_1(\mathrm{n})\phi^{\Theta}_{\mathrm{J,K}}(\mathrm{n}), \tag{9}$$

Table 1 The BBSS algorithm (Djendi et al. 2006)

Intialisation	$0<\theta_1<1, 0<\theta_2<1, \varsigma_1 = 0.001, \varsigma_2 = 0.001, \mathbf{v}_1(n) = [\mathbf{0}]^T, \mathbf{v}_2(n) = [\mathbf{0}]^T, \mathbf{w}_{12}(n) = [\mathbf{0}]^T, \mathbf{w}_{21}(n) = [\mathbf{0}]^T.$
Filtering errors	$v_1(\mathrm{n}) = m_1(n) - \mathbf{w}_{21}^T(n)\ \mathbf{v}_2(n-1)$ $v_2(\mathrm{n}) = m_2(n) - \mathbf{w}_{12}^T(n)\ \mathbf{v}_1(n-1)$ $\mathbf{v}_1(n) = \left[v_1(n) \ldots v_1(n-L+1)\right]^T; \mathbf{v}_2(n) = \left[v_2(n) \ldots v_2(n-L+1)\right]^T$
Cross-filters update of BBSS	$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + 2\theta_1 \frac{v_2(n)\ \mathbf{v}_1(n)}{\mathbf{v}_1^T(n)\ \mathbf{v}_1(n)+\varsigma_1}\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + 2\theta_2 \frac{v_1(n)\ \mathbf{v}_2(n)}{\mathbf{v}_2^T(n)\ \mathbf{v}_2(n)+\varsigma_1}$

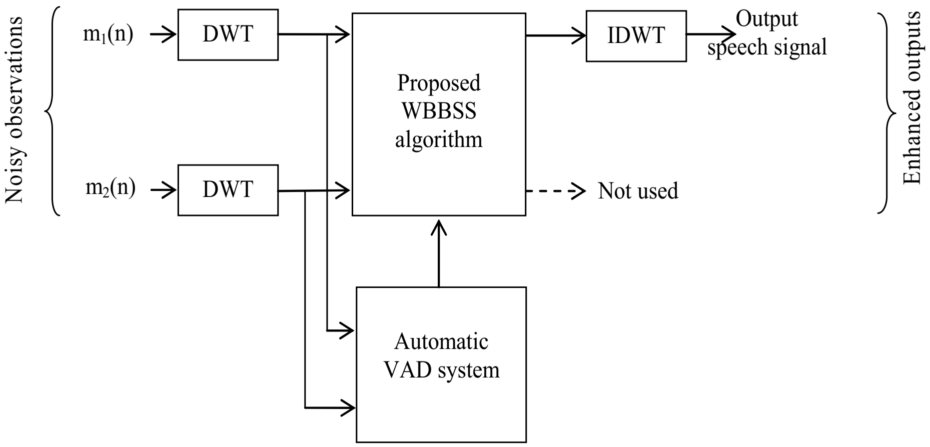

Fig. 4 Bloc diagram of the proposed WBBSS algorithm. *DWT* discrete wavelet transform, *IDWT* the inverse of DWT, *VAD* a voice activity detector

$$M^{(2)}_{\mathrm{J,K}}(\mathrm{n}) = \sum_{\mathrm{J,k}} m_2(\mathrm{n})\, \phi^{\Theta}_{\mathrm{J,K}}(\mathrm{n}), \tag{10}$$

where $\phi_{\mathrm{J,k}}(\mathrm{n}) = 2^{-\mathrm{J}/2}\phi\left(2^{-\mathrm{J}}\mathrm{n\text{-}k}\right)$ is the wavelet basis function, and ‘Θ’ symbolizes complex conjugate. J and k indicate scale and translation respectively of the mother wavelet function ϕ. The two mixing signals $\mathbf{m}_1(\mathrm{n})$ and $\mathbf{m}_2(\mathrm{n})$ of length L are given by $\mathrm{m}_1(\mathrm{n}) = \left[\mathrm{m}_1(\mathrm{n}), \mathrm{m}_1(\mathrm{n}-1), \ldots \mathrm{m}_1(\mathrm{n}-\mathrm{L}+1)\right]^{\mathrm{T}}$, and $\mathrm{m}_2(\mathrm{n}) = \left[\mathrm{m}_2(\mathrm{n}), \mathrm{m}_2(\mathrm{n}-1), \ldots \mathrm{m}_2(\mathrm{n}-\mathrm{L}+1)\right]^{\mathrm{T}}$. The two ($N$ x L) DWT matrices (where L is the length of the DWT vector for J scales) $M^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ and $M^{(2)}_{\mathrm{J,K}}(\mathrm{n})$of the mixing signals $\mathbf{m}_1(\mathrm{n})$ and $\mathbf{m}_2(\mathrm{n})$, respectively, are defined as follows: $M^{(1)}_{\mathrm{J,K}}(\mathrm{n})=\left[m^{(1)}_{0,\mathrm{K}}(\mathrm{n}), m^{(1)}_{1,\mathrm{K}}(\mathrm{n}-1), \ldots m^{(1)}_{\mathrm{N\text{-}1,K}}(\mathrm{n}-\mathrm{L}+1)\right]^{T}$ and $M^{(2)}_{\mathrm{J,K}}(\mathrm{n})=\left[m^{(2)}_{0,\mathrm{K}}(\mathrm{n}), m^{(2)}_{1,\mathrm{K}}(\mathrm{n}-1), \ldots m^{(2)}_{\mathrm{N\text{-}1,K}}(\mathrm{n}-\mathrm{L}+1)\right]^{T}$, where N is the DWT number scale, i.e., {J, K} = 0, 1, 2, 3, ..., *N*-1.

In the second step we apply the FBSS structure to the two converted DWT mixing signals $M^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ and $M^{(2)}_{\mathrm{J,K}}(\mathrm{n})$.This procedure generates the approximations and the detail parts of the two mixing DWT observations, $M^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ and $M^{(2)}_{\mathrm{J,K}}(\mathrm{n})$, then we use them to estimate, at each iteration, the filtering errors vectors which are given as follows:

$$\mathbf{v}_1\,(\mathrm{n})=\mathbf{m}_1(\mathrm{n}) - \sum_{\mathrm{J}} \mathbf{w}_{21,\mathrm{J}}(\mathrm{n})\mathbf{M}^{(1)}_{\mathrm{J,K}}(\mathrm{n}) \tag{11}$$

$$\mathbf{v}_2\,(\mathrm{n}) = \mathbf{m}_2(\mathrm{n}) - \sum_{\mathrm{J}} \mathbf{w}_{12,\mathrm{J}}(\mathrm{n})\mathbf{M}^{(2)}_{\mathrm{J,K}}(\mathrm{n}) \tag{12}$$

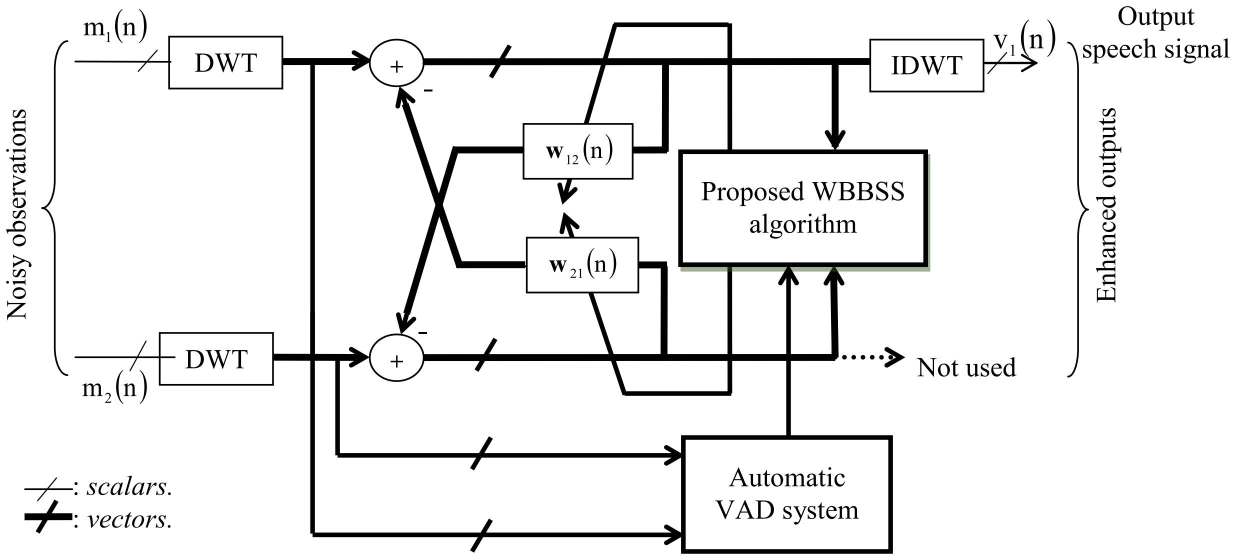

Fig. 5 Proposed WBBSS algorithm. *DWT* discrete wavelet transform, *IDWT* the inverse of DWT, *VAD* a voice activity detector

After the filtering errors vector computation, then we apply the BBSS algorithm to the two wavelets-domain mixing signals $\mathbf{M}^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ and $\mathbf{M}^{(2)}_{\mathrm{J,K}}(\mathrm{n})$. We use approximations and details information of noisy observations, $\mathbf{M}^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ and $\mathbf{M}^{(2)}_{\mathrm{J,K}}(\mathrm{n})$, to update the cross-adaptive filters $\mathbf{w}_{12,\mathrm{J}}(\mathrm{n})$, and $\mathbf{w}_{21,\mathrm{J}}(\mathrm{n})$ at each sub-scale J:

$$\mathbf{w}_{12,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{12,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{12}\mathbf{v}_2(\mathrm{n})\frac{M^{(1)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_1 + M^{(1)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(1)}_{\mathrm{J,K}}(\mathrm{n})} \tag{13}$$

$$\mathbf{w}_{21,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{21,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{21}\mathbf{v}_1(\mathrm{n})\frac{M^{(2)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_2 + M^{(2)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(2)}_{\mathrm{J,K}}(\mathrm{n})} \tag{14}$$

The step-sizes μ_{12} and μ_{21} are used to control the adaptation process of the convergence of the cross adaptive filters$\mathbf{w}_{12,\mathrm{J}}(\mathrm{n})$, and $\mathbf{w}_{21,\mathrm{J}}(\mathrm{n})$ in each DWT scale, respectively. The two-step-sizes μ_{12} and μ_{21} must verify the following relations: $0 < \mu_{12} < 1$ and $0 < \mu_{21} < 1$. The parameter ξ is a small scalar used to avoid division by zero. In last step, and in order to extract the speech signal at the output $\mathrm{v}_1(\mathrm{n})$ and the noise component at the output $\mathrm{v}_2(\mathrm{n})$, we propose to do a modification in relations (13) and (14) by applying an automatic VAD system that is proposed recently in (Jin et al. 2014). We normalized each relations (13) and (14) of the cross adaptive filters $\mathbf{w}_{12,\mathrm{J}}(\mathrm{n})$, and $\mathbf{w}_{21,\mathrm{J}}(\mathrm{n})$ by the norm of the corresponding error vectors $\mathbf{v}_2(\mathrm{n})$ and $\mathbf{v}_1(\mathrm{n})$, respectively. The new update formulas of the cross adaptive filter $\mathbf{w}_{12,\mathrm{J}}(\mathrm{n})$ and $\mathbf{w}_{21,\mathrm{J}}(\mathrm{n})$ at each wavelet scale are given as follows:

$$\mathbf{w}_{12,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{12,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{12}\mathbf{v}_2(\mathrm{n})\mathbf{D}_{12}(\mathrm{n}) \tag{15}$$

$$\mathbf{w}_{21,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{21,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{21}\mathbf{v}_1(\mathrm{n})\mathbf{D}_{21}(\mathrm{n}) \tag{16}$$

where the $\mathbf{D}_{12}(\mathrm{n})$ and $\mathbf{D}_{21}(\mathrm{n})$ are given by the following relations:

$$\mathbf{D}_{12}(\mathrm{n})\frac{M^{(1)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_1 + \left\|v_2(\mathrm{n})\right\|_2^2 + M^{(1)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(1)}_{\mathrm{J,K}}(\mathrm{n})} \tag{17}$$

$$\mathbf{D}_{21}(\mathrm{n})\frac{M^{(2)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_2 + \left\|v_1(\mathrm{n})\right\|_2^2 + M^{(2)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(2)}_{\mathrm{J,K}}(\mathrm{n})} \tag{18}$$

where $\|.\|_2^2$ is the mathematical squared norm operator. $\mathbf{v}_j(n) = \left[v_j(n) \ldots v_j(n-L+1)\right]^T$ with $j = \{1, 2\}$ are the output BBSS signal vectors. The scalar ξ_1 and ξ_2 are small constants used to avoid division by zeros. In Table 2, we summarize the proposed WBBSS algorithm.

5 Analysis of simulation results

We have done intensive experiments on the proposed WBBSS algorithm in comparison with the original version [i.e. BBSS (Djendi et al. 2006)], and the two-step noise reduction (that is called TSNR) algorithm (Plapous et al. 2004) that use only one microphone to enhance the corrupted speech signal, also this TSNR algorithm is mainly based on the use of the Wiener function gain to estimate in two times the correcting gain of the corrupted speech signal, this algorithm I well explained in (Plapous et al. 2004). The comparison of the proposed algorithm with BBSS (Djendi et al. 2006) and TSNR (Weinstein et al. 1993) is based on the evaluation of the convergence speed performance and the segmental mean square error of the output speech signals of each algorithm. The comparison is based on the following criteria: (i) the segmental mean square error (SegMSE), (ii) the segmental signal-to-noise-ratio (SegSNR), and (iii) the cepstral distance (CD) (Hu and Loizou 2007, 2008). All of the simulations are used the same parameters as given in

Table 2 The proposed WBBSS algorithm [In this paper]

Initialization	$0 < \mu_{12} < 1, 0 < \mu_{21} < 1, \xi_1 = 0.001, \xi_2 = 0.001, \mathbf{v}_1(n) = [\mathbf{0}]^T,$ and $\mathbf{v}_2(n) = [\mathbf{0}]^T$, $\mathbf{m}_1(n) = [\mathbf{0}]^T$, and $\mathbf{m}_2(n) = [\mathbf{0}]^T$ $\mathbf{w}_{12,J}(n) = [\mathbf{0}]^T, \mathbf{w}_{21,J}(n) = [\mathbf{0}]^T$. Selection of $\phi^{\Theta}_{\mathrm{J,K}}(\mathrm{n})$.
Noisy inputs DWT transformations	$M^{(1)}_{\mathrm{J,K}}(\mathrm{n}) = \sum_{\mathrm{J,k}} m_1(\mathrm{n})\,\phi^{\Theta}_{\mathrm{J,K}}(\mathrm{n})$, $M^{(2)}_{\mathrm{J,K}}(\mathrm{n}) = \sum_{\mathrm{J,k}} m_2(\mathrm{n})\,\phi^{\Theta}_{\mathrm{J,K}}(\mathrm{n})$
Output filtering errors	$\mathbf{v}_1(\mathrm{n}) = \mathbf{m}_1(\mathrm{n}) - \sum_{\mathrm{J}} \mathbf{w}_{21,\mathrm{J}}(\mathrm{n})M^{(1)}_{\mathrm{J,K}}(\mathrm{n})$ $\mathbf{v}_2(\mathrm{n}) = \mathbf{m}_2(\mathrm{n}) - \sum_{\mathrm{J}} \mathbf{w}_{12,\mathrm{J}}(\mathrm{n})M^{(2)}_{\mathrm{J,K}}(\mathrm{n})$
Intermediate automatic parameters	$\mathbf{D}_{12}(\mathrm{n})\frac{M^{(1)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_1+\left\|v_2(\mathrm{n})\right\|_2^2+M^{(1)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(1)}_{\mathrm{J,K}}(\mathrm{n})}$ $\mathbf{D}_{21}(\mathrm{n})\frac{M^{(2)}_{\mathrm{J,K}}(\mathrm{n})}{\xi_2+\left\|v_1(\mathrm{n})\right\|_2^2+M^{(2)}_{\mathrm{J,K}}(\mathrm{n})M^{\mathrm{T}(2)}_{\mathrm{J,K}}(\mathrm{n})}$
Cross-filters updates	$\mathbf{w}_{12,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{12,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{12}\mathbf{v}_2(\mathrm{n})\,\mathbf{D}_{12}(\mathrm{n})$ $\mathbf{w}_{21,\mathrm{J}}(\mathrm{n}+1) = \mathbf{w}_{21,\mathrm{J}}(\mathrm{n}) + 2\,\mu_{21}\mathbf{v}_1(\mathrm{n})\,\mathbf{D}_{21}(\mathrm{n})$

Table 3 for the algorithms, i.e. the TSNR, the BBSS and the proposed WBBSS.

5.1 Description of the signals used in simulation

In Fig. 6 we show the original speech signal and the USASI (United State of America Standard Institute, now ISI) noise signals that are used in simulation. On Fig. 7, we show an example of real impulse responses that we used in our experiments. These source signals are sampled at 16 kHz and coded on 16 bit. The speech signals are taken from the AURORA database. These signals are measured in real situations and often used in the domain of adaptive filtering and speech enhancement to test the performance of algorithms. We recall here that the used voice activity detector in the BBSS algorithm is manual.

In order to show the performance of the proposed WBBSS algorithm in comparison with the classical BBSS (Djendi et al. 2006), and the TSNR (Plapous et al. 2004), we have reported on Figs. 8 and 9 the output speech signal v_1(n) obtained by the two algorithms (i.e. the classical BBSS, TSNR, and the proposed WBBSS). In the proposed WBBSS algorithm, the DWT scale is N = 2 (the results are shown on Fig. 8) and N = 4 (the results are shown on Fig. 9). The simulation parameters for each algorithm are given in Table 3.

According to Figs. 8 and 9, we can well see that the estimated speech signal v_1(n) is quickly denoised by the proposed WBBSS algorithm in comparison with TSNR and BBSS algorithms that take more times to converge especially in the case of N = 4. This experiment leads to say that the three algorithms (BBSS, TSNR, and proposed WBBSS) are good enough, to reduce acoustic noise components at the output. To give insight of the proposed

Table 3 Simulation parameters of BBSS (Djendi et al. 2006), TSNR (Plapous et al. 2004), and proposed WBBSS algorithms

	Parameters
Input signals	Speech signal: Aurora data base Noise: USASI noise Sampling frequency rate: fs = 16 kHz
Mixing signals	Real filters length: $L = 128$ Input SNR = -6, 0, 6 dB
BBSS algorithm (Djendi et al. 2006)	Adaptive filter length: $L = 128$ Fixed step-sizes: $\theta_1 = 0.8$, $\theta_1 = 0.8$ $\varsigma_1 = \varsigma_2 = 10^{-6}$
TSNR (Plapous et al. 2004)	Frame length = 256; overlapping length = 50%, Wiener filter is used as a gain function
Proposed WBBSS algorithm [In this paper]	Adaptive filter length: $L = 128$ Fixed step-sizes: $\mu_{12} = 0.\ 8$, $\mu_{21} = 0.8$ $\xi_1 = \xi_2 = 10^{-6}$ Adaptive filters: relations (15) and (16) DWT scale, N = 2 and 4

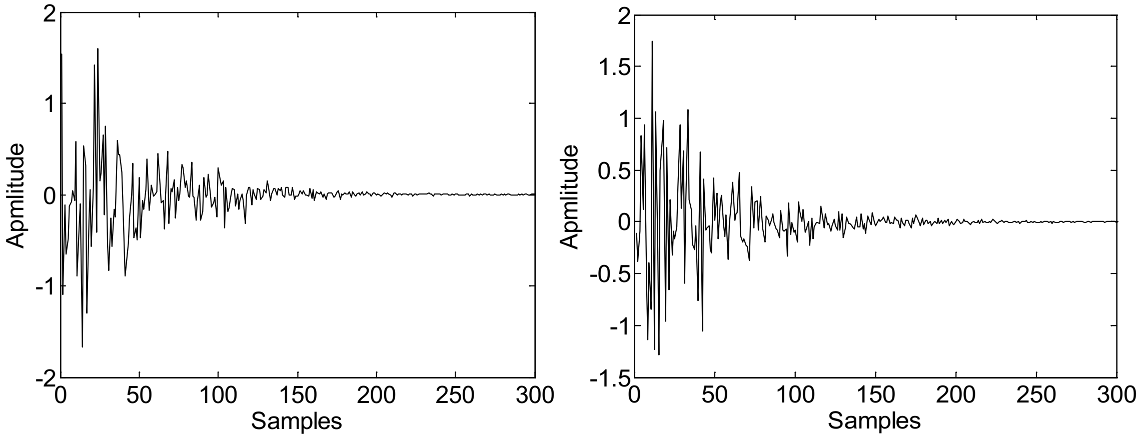

Fig. 6 Simulated impulse responses in the spaced microphones case; [Left]: $h_{12}(n)$, [Right]: $h_{21}(n)$. The real filters length is $L = 128$. fs = 8 kHz

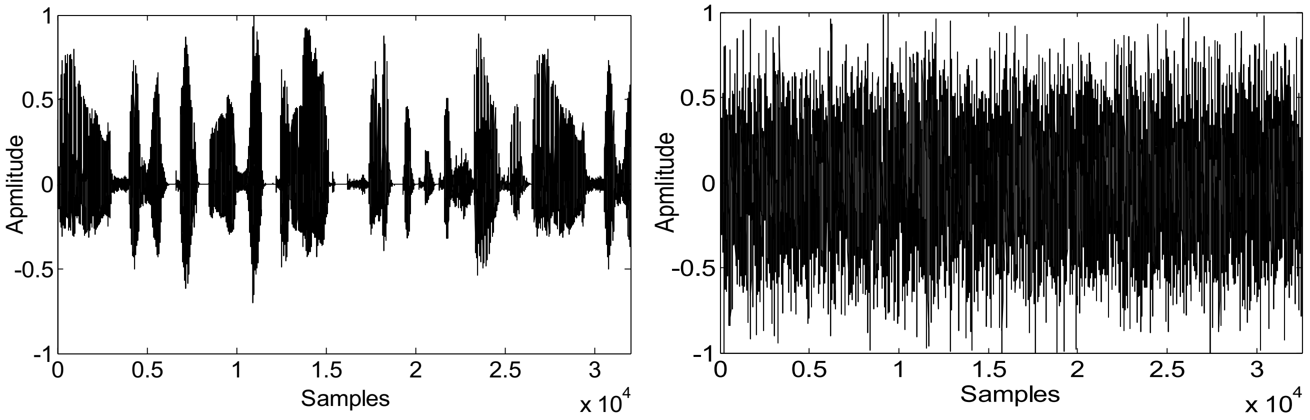

Fig. 7 Original speech signal taken from Aurora data-base [In left], and the USASI (United State of America Standard Institute now ANSI) noise [In right]

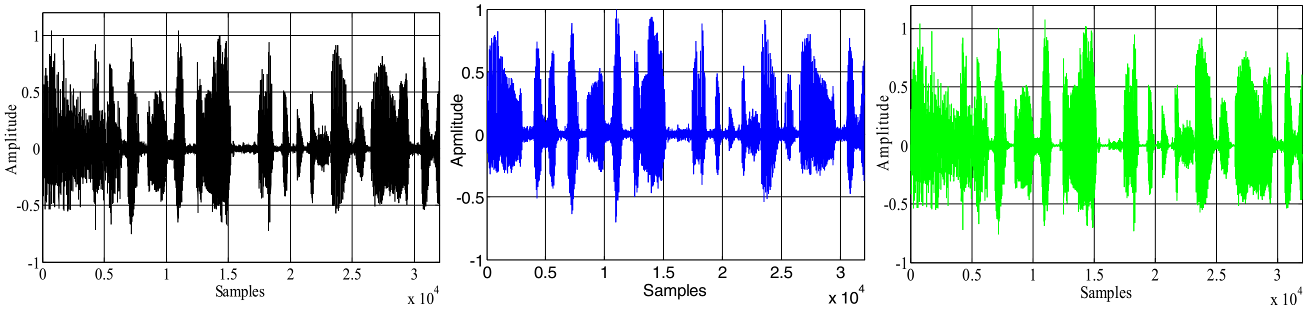

Fig. 8 Outputs of BBSS [in black], TSNR [in blue] and proposed WBBSS [in green] for N = 2. (Color figure online)

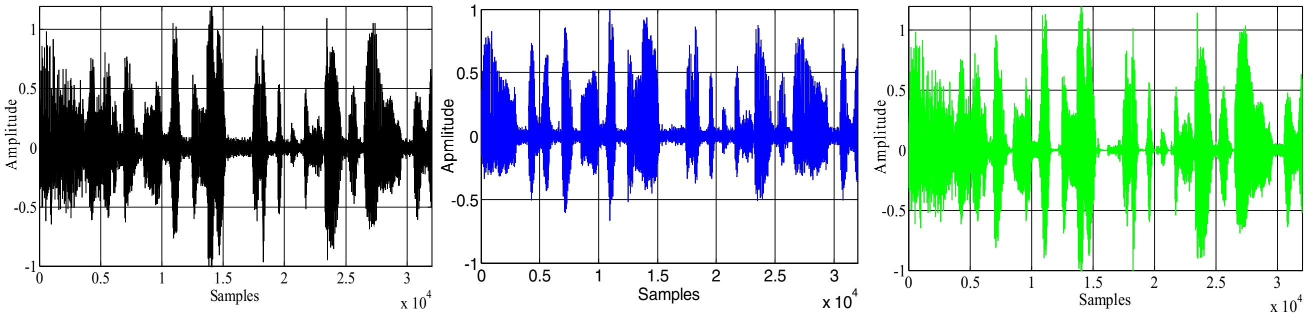

Fig. 9 Outputs of BBSS [in black], TSNR [in blue] and proposed WBBSS [in green] for N = 4. (Color figure online)

WBBSS algorithm, we present in the next sections intensive simulations that are carried out with other objective criteria.

5.2 Segmental MSE (SegMSE) evaluation

In order to quantify the convergence speed performance of the adaptive filters of the proposed algorithm and its original version, i.e. the BBSS (Djendi et al. 2006), we have used the SegMSE criterion. As the output speech signal $v_1(n)$ is controlled by the adaptive filters $w_{21}(n)$, we concentrate on the evaluation of the SegMSE criterion calculated on the output of this filter, i.e. $w_{21}(n)$. This SegMSE criterion is evaluated by the following relation (Hu and Loizou 2007, 2008):

$$\mathrm{SegMSE_{dB}} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\left(\frac{1}{Z}\sum_{n=Zm}^{Zm+Z-1}\left|v_1(n)\right|^2\right) \quad (19)$$

where ‘Z’ is the time averaging frame length of the output $v_1(n)$ or the segment length, and ‘M’ is the number of segments when speech signal is absent. We note that the SegMSE criterion is estimated in noise-only presence periods. This good property is achieved thanks to the speech signal intermittence property. All the simulation parameters of both algorithms are summarized in Table 3. The used noise types are White, Babble, USASI (United Sate of America Signal Institute now ANSI), and street. All the noises that we use in all the simulation of this paper are real and sampled at 8 kHz.

The obtained results of the SegMSE for four noises types are reported on Fig. 8. According to this experiment of Fig. 10, we can say that the proposed WBBSS algorithm behaves more efficiently than the BBSS, especially when the DWT scale is selected high (N = 4). This is true for all of the other DWT scales and for all noise types.

5.3 SegSNR evaluation

We have evaluated the performance of the two-step noise reduction (TSNR) algorithm (Plapous et al. 2004), BBSS (Djendi et al. 2006) and the proposed WBBSS algorithm in terms of output segmental signal-to-noise ratio (SegSNR) criterion. The SegSNR allows quantifying the acoustic noise reduction amounts at the output. As we are interested on speech enhancement signal, we focus only on the SegSNR output of the output $v_1(n)$. The SegSNR criterion is evaluated according to the following relation (Hu and Loizou 2007, 2008):

$$\mathrm{SegSNR_{dB}} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\left(\frac{\sum_{n=Zm}^{Zm+Z-1}|s(n)|^2}{\sum_{n=Zm}^{Zm+Z-1}|s(n)-v_1(n)|^2}\right) \tag{20}$$

where s(n) and $v_1(n)$ are the original and the enhanced speech signals, respectively. The parameters ‘M’ and ‘Z’ are the number of segments and the segment 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 | . | represents the magnitude operator. We recall here that all the ‘M’ segments correspond to only speech signal presence periods. The *log10* symbol is the base 10 logarithm of a given number. The same parameters of each algorithm

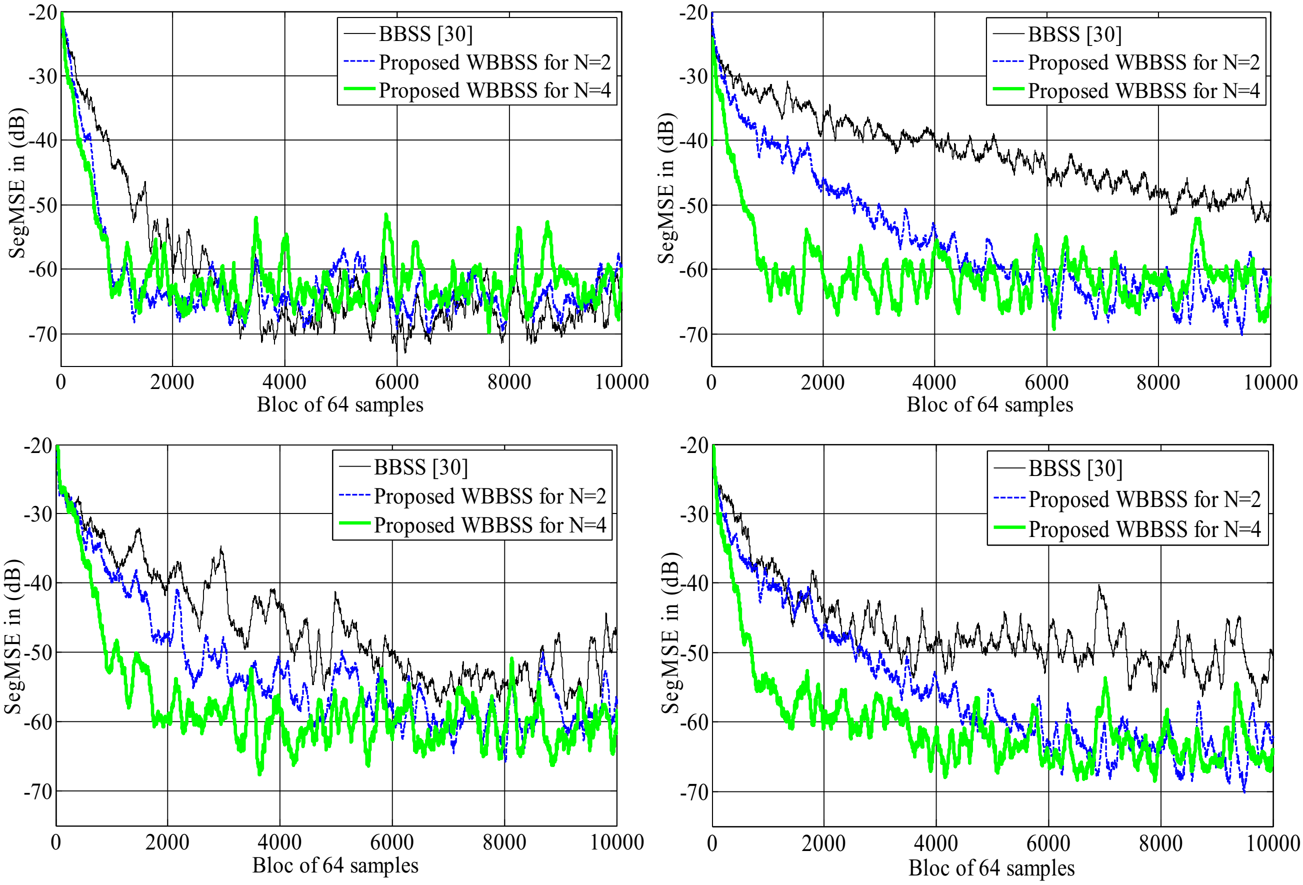

Fig. 10 SegMSE of BBSS (Djendi et al. 2006) and proposed WBBSS algorithms for DWT scales N = 2 and 4. [Top left]: white; [Top right]: USASI; [Bottom left]: Babble; [Bottom right]: Street

that are summarized in Table 3 are used in the simulation. The obtained results of the SegSNR are given in Fig. 11 for three input SNR values, i.e. -6 dB, 0 dB, and 6 dB.

According to Fig. 11, we note the best performances of the proposed WBBSS algorithm when the DWT scale is selected equal to N = 2. However, the SegSNR values decrease when the DWT scale increases. We have also noted that in low and high DWT scales, the general order values of the output SegSNR are superior to 40 dB, which means that the proposed algorithms (with different DWT scales) and both of TSNR (Plapous et al. 2004) and BBSS (Djendi et al. 2006) algorithms lead to a good reduction of the acoustic noise components. The important advantage of the proposed WBBSS algorithm is the speech convergence property that was highlighted in the SegSNR and well shown in the SegMSE criterion of Sect. 5.2 for each input SNR.

5.4 CD evaluation

In order to quantify the distortion amount introduced in the output speech signal by the proposed WBBSS algorithm in comparison with the BBSS (Djendi et al. 2006) and the TSNR (Plapous et al. 2004) ones, we have evaluated the cepstral distance criterion. The CD criterion is computed by the log-spectrum distance between the original speech signal s(n)and the output of the proposed WBBSS and BBSS algorithms, i.e. $v_1(n)$. The CD criterion is evaluated by the following relation (Hu and Loizou 2007, 2008):

$$\mathrm{CD_{dB}} = \frac{10}{\mathrm{M}}\sum_{m=0}^{M-1}\log 10\sum_{n=Zm}^{Zm+Z-1}\left(c_s(n) - \left(c_{v1}(n)\right)^2\right) \tag{21}$$

where $c_s(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log|S(\omega)|e^{j\omega n}d\omega$ and $c_{v1}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|V_1(\omega)\right|e^{j\omega n}d\omega$ are the nth real cepstral coefficients of the signals s(n) and $v_1(n)$, respectively. We recall here that $S(\omega)$ and $V_1(\omega)$ present the short Fourier transform of the original speech signal s(n) and the enhanced one$v_1(n)$, respectively. The parameter 'Z' is the mean averaging value of the CD criterion and 'M' represents the number of segment where only speech is present. The simulations parameters of this experiment are the same as given in Table 3. In order to well see the output speech signal distortion evolution of the proposed WBBSS algorithm, we have selected the DWT scale to be equal to N = 2 and then to N = 4. The simulation results of this experiment for three input SNR values, i.e. − 6, 0, and 6 dB are shown on Fig. 12.

From this experiment and according to Fig. 12, we conclude that the proposed WBBSS algorithm has almost the

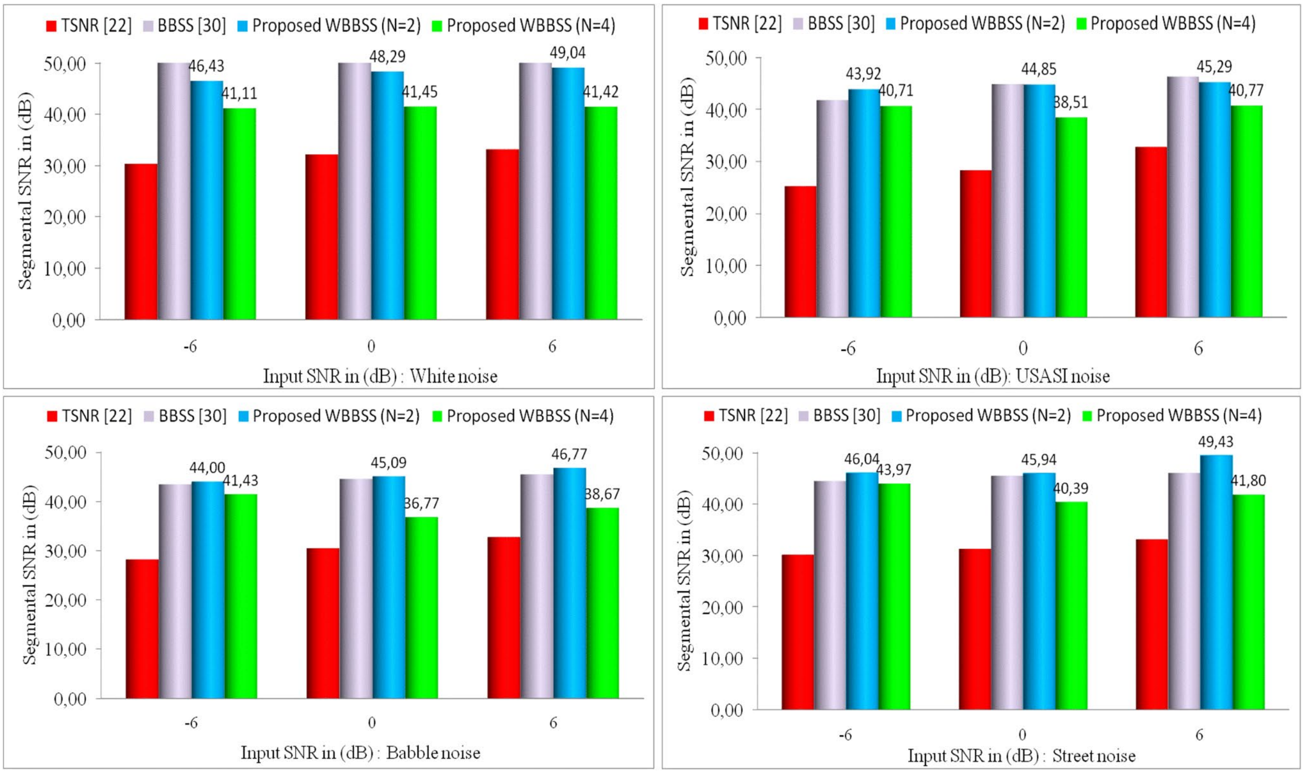

Fig. 11 SegSNR evaluation of the BBSS (Djendi et al. 2006), TSNR (Plapous et al. 2004), and the proposed WBBSS algorithms for WDT scale N = 2 and 4. [Top left]: white; [Top right]: USASI; [Bottom left]: Babble; [Bottom right]: Street

lower CD values for different scale of the DWT, with a slight superiority in the case when the DWT scale is selected N = 2. These results show that the proposed WBBSS algorithm outperforms the conventional BBSS (Djendi et al. 2006) and TSNR (Plapous et al. 2004) for each DWT scale. However, we observe a decrease of the CD values when the number of DWT scale is chosen low, i.e. N = 2. We have also noted that for high input SNR situations, the CD values are low thus signifying a superior intelligibility performance of the output speech signal of the proposed WBBSS algorithm. We have also observed that the enhanced speech signals of the proposed WBBSS algorithm sound much clearer with much less speech distortion than the BBSS (Djendi et al. 2006) and TSNR (Plapous et al. 2004) algorithms.

6 Conclusion

In this paper, we have proposed a new wavelet based BSS (WBBSS) algorithm that is based on a wavelet implementation of the backward blind source separation (BBSS) algorithm. The theoretical derivation and intensive experimental validation of the proposed WBBSS algorithm have been presented along this paper. The proposed WBBSS algorithm has shown best performances in terms of segmental signal-to-noise ratio, segmental mean square error and cepstral distance criteria. We have also shown the best convergence speed performance property of the WBBSS algorithm through the evaluation of the system mismatch criterion. This new work allows improving the convergence speed of the WBBSS algorithm even when very noisy observations are available. The drawback of the proposed WBBSS algorithm is a degradation of the criterion CD, SegMSE, SegSNR when the discrete wavelet transform (DWT) scale is selected high. Also, we have noted that the proposed WBBSS outperforms the BBSS and the TSNR algorithms with low DWT scales, and comparable performances when high DWT scales are selected. We have also noted that both of BBSS and TSNR algorithms use a manual VAD to give such performances, however, the proposed WBBSS algorithm detect activity and silence periods of the speech signal automatically thanks to the modifications that we have reported on the update relation of the cross-filters (See relations 15, 16). This means that the new WBBSS algorithm can be used directly in practice as its work automatically without need of any VAD systems. In order to improve the behavior of the proposed WBBSS algorithm when the DWT

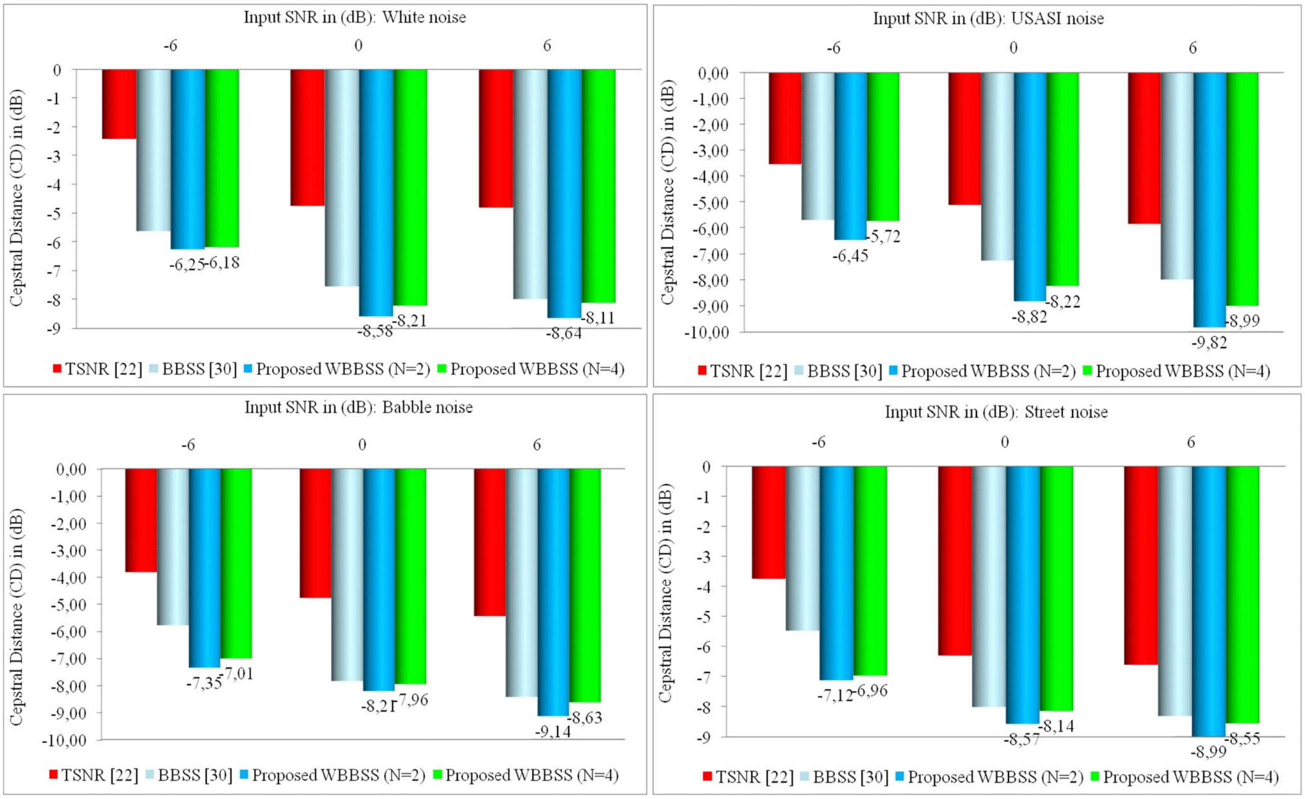

Fig. 12 CD evaluation of the BBSS (Djendi et al. 2006), the TSNR (Plapous et al. 2004), and the proposed WBBSS algorithms. [Top left]: white; [Top right]: USASI; [Bottom left]: Babble; [Bottom right]: Street

scales are selected high, we can envisage using a variable step-size to correct this problem in our future works.

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.

Bactor, P., & Garg, A. (2012). Different techniques for the enhancement of the intelligibility of a speech signal. *International Journal of Engineering Research and Development, 2*(2), 57–64.

Benesty, J., & Cohen, I. (2017). Multichannel speech enhancement in the STFT domain. In J. Benesty, & I. Cohen (Eds.), *Canonical correlation analysis in speech enhancement, Springer briefs in electrical and computer engineering* (pp. 79–101). New York: Springer.

Boll, SF (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing, 27*, 113–120.

Bouzid, A., & Ellouze, N. (2016). Speech enhancement based on wavelet packet of an improved principal component analysis. *Journal Computer Speech and Language, 35*, 58–72.

Cappé, O. (1994). Elimination of the musical noise phenomenon with the Ephraïm and Malah noise suppressor. *IEEE Transactions on Speech Audio Processing, 2*(2), 345–349.

Davila, C. E. (1984). A subspace approach to estimation of autoregressive parameters from noisy measurements. *IEEE Transaction on Signal processing, 46*, 531–534.

Dixit, S., & Mulge, M. Y. (2014). Review on speech enhancement techniques, *International Journal of Computer Science and Mobile Computing, 3*(8), 285–290.

Djendi, M., Bensafia, S., & Safi, M. (2016). A frequency co-channel adaptive algorithm for speech quality enhancement, In *International Conference on Engineering and MIS (ICEMIS).*

Djendi, M., Khemies, F., & Morsli, A. (2015). A Frequency Domain Adaptive Decorrelating Algorithm for Speech Enhancement. In *International Conference on Speech and Computer, SPECOM 2015*, pp. 51–54.

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 the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 744–747.

Djendi, M., & Scalart, P. (2012). Double Pseudo Affine Projection algorithm for speech enhancement and acoustic noise reduction. *2012 Proceedings of the 20th European, Romania, Bucharest*, Vol. 1, pp. 2080–2084.

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

Doclo, S., & Moonen, M. (2002). GSVD-based optimal filtering for signal and multi-microphone speech enhancement. *IEEE Transaction on Signal processing, 50*, 2230–2244.

Dong, J., Wei, X. P., & Zhang, Q. (2009). Speech enhancement algorithm based on high-order Cumulant parameter estimation. *International Journal of Innovative Computing information and Control, 5*, 2725–2733.

Ephraim, Y., LevAri, H., Roberts, W. J. J. (2014). A brief survey of speech enhancement. *IEEE Signal Processing Letters, 10*, 104–106.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 32*(6), 1109–1121.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-33*(2), 443–445.

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

Goldsworthy, R. L. (2014). Two-microphone spatial filtering improves speech reception for cochlear-implant users in reverberant conditions with multiple noise sources. *Trends in Hearing, 18*, 1–13.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America, 122*(3), 1777–1786.

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.

Jin, Y. G., Shin, J. W., & Kim, N. S. (2014). Spectro-temporal filtering for multichannel speech enhancement in short-time Fourier transform domain. *IEEE Signal Processing Letters, 21*(3), 352–355.

Jutten, C., & Herrault, J. (1991). Blind separation of sources: an adaptive algorithm based on neuromimetic architecture. *Signal Processing, 24*, 1–10.

Lee, G., & Dae Na, S. (2016). Seong K2, Cho JH3, Nam Kim M4. Wavelet speech enhancement algorithm using exponential semi-soft masks filtering. *Bioengineered, 7*(5), 352–356.

Lee, K. A., & Gan, W. S. (2004). Improving convergence of the NLMS algorithm using constrained subband updates. *IEEE Signal Processing Letters, 11*(9), 736–739.

Loizou, P. C. (2007). *Speech enhancement: Theory and practice* (pp. 589–599). Boca Raton, FL: Taylor and Francis.

Mansour, A., Jutten, C., & Loubaton, P. (1996). Subspace method for blind separation of sources and for a convolutive mixture model. Signal processing VIII, theories and applications (pp. 2081–2084).

Nguyen Thi, H. L., & Jutten, C. (1995). Blind sources separation for convolutive mixtures. *Signal Processing, 45*, 209–229.

Plapous, C., Marro, C., Scalart, P. (2005). Speech enhancement using harmonic regeneration, In *IEEE International Conference on Acoustics, Speech, Signal Processing*, Philadelphia, PA, USA, 1, pp. 157–160.

Plapous, C., Marro, C., Scalart, P., Mauuary, L., & Two-Step, A. (2004). Noise reduction technique. In *IEEE International Conference on Acoustics, Speech, Signal Processing*, Montral, Quebec Canada, 1, pp. 289–292.

Scalart, P., Filho, J. (1996). Speech enhancement based on a priori signal to noise estimation. In *International Conference on Acoustics, Speech, and Signal Processing*. pp. 629–632.

Selva Nidhyananthan, S., Shantha Selva Kumari, R., & Arun Prakash, A. (2014). A review on speech enhancement algorithms and why to combine with environment classification. *International Journal of Modern Physics C, 25*(10), 210–225.

Tong, R., Bao, G., & Ye, Z. (2015). A higher order subspace algorithm for multichannel speech enhancement. *IEEE Signal Processing Letters, 22*(11), 2004–2008.

Van Gerven, S., & Van Compernolle, D. (1992). Feed forward and feedback in symmetric adaptive noise canceller: Stability analysis in a simplified case. In *European Signal Processing Conference*, Brussels, Belgium. pp. 1081–1084.

Weinstein, E., Feder, M., & Oppenheim, A. V. (1993). Multi-channel signal separation by decorrelation. *IEEE Transactions on Speech Audio Processing, 1*(4), 405–413.

Widrow, B., & Stearns, S. D. (1985). *Adaptive signal processing*, Upper Saddle River: Prentice-Hall.

Widrow, B., Goodlin, R. C. (1975). Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE, 63*, 1692–1716.

Wolfe, P. J., & Godsill, S. J. (2003). Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement. *EURASIP Journal on Applied Signal Processing, 10*, 1043–1051.

Zhang, Y., & Zhao, Y. (2012). Real and imaginary modulation spectral subtraction for speech enhancement. *Journal on Speech Communication, 55*(6), 509–522.

Zoulikha, M., & Djendi, M. (2016). A new regularized forward blind source separation algorithm for automatic speech quality enhancement. *Applied Acoustics, 112*, 192–200.

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