A new speech enhancement adaptive algorithm based on fullband–subband MSE switching

Akila Sayoud[1] · Mohamed Djendi[1] · Abderrezak Guessoum[1]

Abstract

This paper presents a new fullband–subband switching adaptive speech enhancement algorithm, based on mean square error estimation. The proposed algorithm is able to automatically switch between two adaptive filtering algorithms, i.e. the two-channel fullband normalized least mean square (TC-FNLMS) algorithm and the two-channel subband normalized least mean square (TC-SNLMS), where, the proposed switching mechanism leads to a significant improvement in the convergence speed performance of the proposed algorithm. To confirm the efficiency and the good performances of the proposed algorithm in comparison with the fullband and subband versions of the two channel NLMS algorithm, several experiments were carried out in terms of the segmental signal-to-noise-ratio (SegSNR), segmental mean square error (SegMSE), system mismatch (SM) and cepstral distance (CD).

1 Introduction

The advent and wide dissemination of mobile voice communication systems has substantially increased the need for reliable communication systems in noisy environments. The environmental noise is a significant limitation on the performance of hands-free voice communications (Thumchirdchupong and Tangsangiumvisai 2013). To handle this critical issue, i.e. the environmental noise, several noise reduction and speech enhancement techniques have been proposed in the literature (Loizou 2017; Kuo and Peng 1999; Dixit and Mulge 2014). Generally, the approaches can be classified into a single-channel (Upadhyay and Jaiswal 2016; Benesty and Cohen 2018), dual-channel (Man Kima and Kook Kim 2014; Nabi et al. 2017) or multi-channel (Kim

✉ Mohamed Djendi
m_djendi@yahoo.fr

Akila Sayoud
akila.sayoud@gmail.com

Abderrezak Guessoum
aguessoum@yahoo.fr

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

and Hasegawa-Johnson 2012; Habets and Benesty 2013) speech enhancement methods.

In recent years, dual-channel speech enhancement algorithms have been widely used due to their good performance in different noise situations (i.e. stationary or non-stationary noise signal) (Sayoud et al. 2018; Nabi et al. 2016), unlike the single-channel methods that fail in the presence of non-stationary noise (Loizou 2007). Also, they are easy to implement and do not require a large computational complexity compared to multi-channel speech enhancement techniques. Nabi et al. (2018) proposed a dual-microphone noise reduction algorithm based on the coherence function and the bionic wavelet transform using Kalman filter, this method has the ability to deal with two closely spaced microphones and does not require noise statistics estimation. Moreover, numerous dual-channel speech enhancement techniques based on the combination between adaptive algorithms and the forward and backward blind source separation structures have been proposed (Henni et al. 2019; Djendi and Zoulikha 2018; Djendi 2018). Subband adaptive filtering is another efficient approach that is often used for speech enhancement and noise reduction applications due to its capability in improving convergence performance for highly correlated input signals. Lee and Gan (2004) proposed the normalized subband adaptive filtering (NSAF) algorithm, where, it allows to whiten the input and the desired output signals

by divided them into multiple subbands and that improves the convergence rate. For further performances improvement, several modified versions of the NSAF algorithm have been developed (Ni and Li 2010; Yu et al. 2016; Seo and Park 2014). Furthermore, Djendi and Bendoumia (2013) proposed a two-channel subband adaptive filtering algorithm based on the forward blind source separation (FBSS) structure, which has improved the convergence speed performance of the classical FBSS, but, the performance of this algorithm depends on the fixed step size that reflects a compromise between fast convergence rate and low steady-state error. To overcome this conflict, a two-channel subband FBSS structure with variables step-size has been proposed in Djendi and Bendoumia (2016).

In this paper, we propose a new adaptive switching algorithm for noise reduction and speech enhancement applications. The proposed algorithm switches between the two-channel subband normalized least mean square (TC-SNLMS) algorithm (Djendi and Bendoumia 2013) and the two-channel fullband normalized least mean square (TC-FNLMS) one (Van Gerven and Van Compernolle 1992), where, the switching procedure is based on the mean square error (MSE) estimation. After a review of the proposed algorithm, we can say that the proposed switching algorithm exhibits better speed convergence in comparison with the classical two-channel fullband NLMS algorithm and the two- channel subband NLMS one.

The organization of this paper is as follows: in Sect. 2, we present mixing and separation model. In Sect. 3 we describe the principle of the proposed algorithm. The simulation results of the proposed switching algorithm in comparison with other competitive algorithms are presented in Sect. 4. Finally, we conclude our work in Sect. 5.

2 Mixing and separation model

In a car environment, the hands-free communication is normally disturbed by the surrounded noise (engine, babbling or street noise). As depicted on Fig. 1, the recorded signals by the two microphones are contaminated by the noise components, these noisy observations are a linear combination between the useful speech signal and the disturbing noise signal. The main problem is to recover the original source signals from only these two available noisy observations. The blind source separation structure is widely used to address this critical issue, (Sayoud et al. 2018; Djendi and Bendoumia 2013; Syskind Pedersen et al. 2007).

In this paper the acoustical environment is modeled by the simplified convolutive mixture shown in Fig. 2a, and in order to retrieve the original signals, i.e. speech signal $s(n)$ and noise signal $b(n)$, from the observed mixed signals, i.e. $m_1(n)$ and $m_2(n)$, we use the forward blind source separation structure (Sayoud et al. 2018; Djendi 2018; Djendi et al. 2016), as depicted in Fig. 2b.

Let $m_1(n)$ and $m_2(n)$ denote the two noisy observation signals that can be modeled by the following formulations:

$$m_1(n) = s(n) * h_{11}(n) + b(n) * h_{21}(n) \tag{1}$$

$$m_2(n) = b(n) * h_{22}(n) + s(n) * h_{12}(n) \tag{2}$$

where $s(n)$ and $b(n)$ are the two independent sources of speech and noise respectively. The symbol '*' stands for the linear convolution operation. $h_{12}(n)$ and $h_{21}(n)$ represent the cross-coupling effects between the channels. We assume that, the first microphone is close to the speaker and the second microphone is close to the noise source, thus, the direct acoustic paths $h_{11}(n)$ and $h_{22}(n)$ are equal to the Kronecker unit impulse $\delta(n)$ (Van Gerven and Van Compernolle 1995). According to this assumption the two relations of the noisy observation signals can be rewritten as follows:

$$m_1(n) = s(n) + b(n) * h_{21}(n) \tag{3}$$

$$m_2(n) = b(n) + s(n) * h_{12}(n) \tag{4}$$

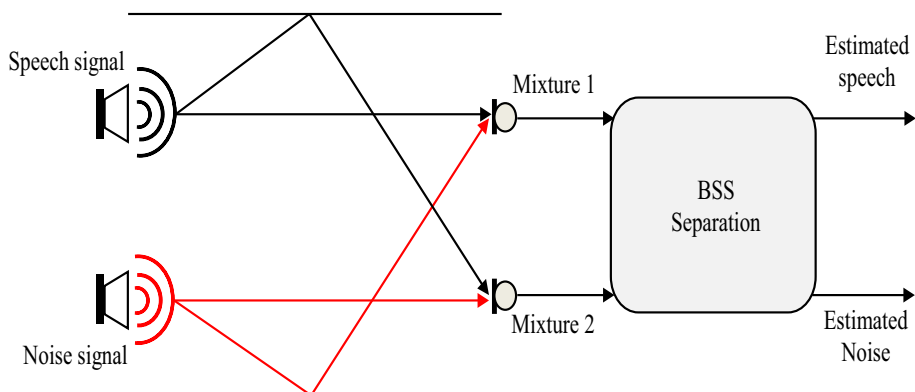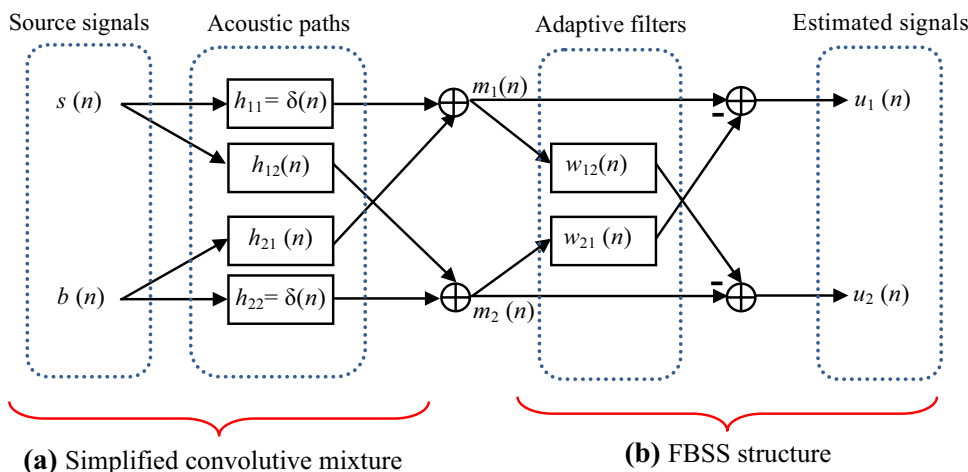**Fig. 1** Illustrative diagram of the mixing and separation model

Fig. 2 Diagrammatic representation of the convolutive mixing and the FBSS structure

(a) Simplified convolutive mixture   **(b)** FBSS structure

The two estimated output signals of the FBSS structure are given by the following relations:

$$u_1(n) = m_1(n) - m_2(n) * w_{21}(n) \qquad (5)$$

$$u_2(n) = m_2(n) - m_1(n) * w_{12}(n) \qquad (6)$$

where $w_{21}(n)$ and $w_{12}(n)$ are the adaptive filters of the FBSS structure.

By inserting Eqs. (3) and (4) in (5) and (6) respectively, we obtain:

$$u_1(n) = b(n) * \left[h_{21}(n) - w_{21}(n)\right] + s(n) * \left[\delta(n) - h_{12}(n) * w_{21}(n)\right] \qquad (7)$$

$$u_2(n) = s(n) * \left[h_{12}(n) - w_{12}(n)\right] + b(n) * \left[\delta(n) - h_{21}(n) * w_{12}(n)\right] \qquad (8)$$

The evident theoretical solution of the problem is obtained by setting $h_{21}(n) = w_{21}(n)$ and $h_{12}(n) = w_{12}(n)$, the estimated output signals $u_1(n)$ and $u_2(n)$ can be expressed as follows:

$$u_1(n) = s(n) * \left[\delta(n) - h_{12}(n) * h_{21}(n)\right] \qquad (9)$$

$$u_2(n) = b(n) * \left[\delta(n) - h_{21}(n) * h_{12}(n)\right] \qquad (10)$$

From Eqs. (9) and (10), we notice that the estimated output signals of the FBSS structure is distorted by the post filter. The effect of this post filter is important when the two microphones are closely spaced (Djendi et al. 2007). To avoid the post filter effect, we consider in our work the case where the two microphones are loosely spaced.

3 Proposed algorithm

In this section, we present the proposed adaptive switching algorithm, where, the switching mechanism is based on the mean square error estimation. A block diagram of the proposed algorithm is shown in Fig. 3. The proposed algorithm is
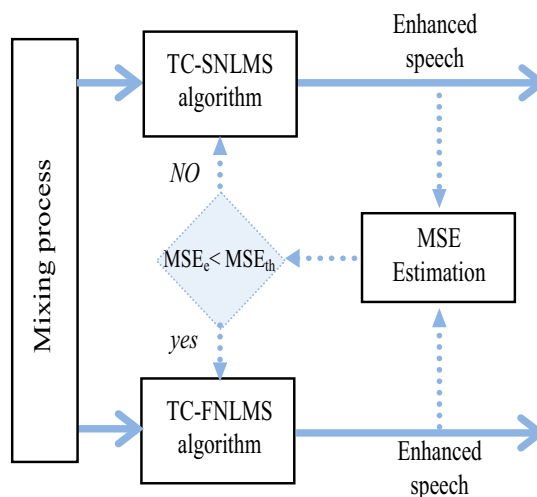

Fig. 3 Flowchart of the proposed algorithm

very well suited to high and low noise level scenarios, where, it is capable of switching between the two-channel subband NLMS algorithm (TC-SNLMS) (Djendi and Bendoumia 2013) and the two-channel fullband NLMS (TC-FNLMS) algorithm (Van Gerven and Van Compernolle 1992). The main idea is to use the TC-SNLMS algorithm for processing high level noise components and alternatively employ the TC-FNLMS algorithm when noise intensity is low. The switching mechanism can be described by the following steps:

- *Step 1* Energy calculation.

In the first step, we compute recursively the energy $E(n)$ of the filtering error $u_1(n)$ by the following relation:

$$E(n) = \beta E(n - 1) + (1 - \beta)u_1(n)^2 \qquad (11)$$

where β is a smoothing factor.

- *Step 2* MSE estimation.

Once the energy has been calculated, the mean square error estimate can be computed as follow:

$$MSE_e(n) = 10 \log_{10}(E) \tag{12}$$

- *Step 3* Switching rule.

The switching mechanism between the two adaptive filtering algorithms is performed by comparing the *MSE* estimate with a *MSE* threshold, where the MSE_{th} value is determined from experimental observations. The switching rule is defined as follows:

$$\begin{cases} \text{if } MSE_e(n) < MSE_{th}, & \text{TC - FNLMS algorithm} \\ \text{else,} & \text{TC - SNLMS algorithm} \end{cases}$$

3.1 Two-channel subband NLMS (TC-SNLMS) algorithm

The TC-SNLMS algorithm is a subband implementation of the forward blind source separation based on the use of the normalized least mean square (NLMS) algorithm (Djendi and Bendoumia 2013). As depicted on Fig. 4, this algorithm consists on dividing the two noisy signals $m_1(n)$ and $m_2(n)$ into a set of subband signals by an M-channel analysis filter banks. The resultant subband signals $m_{1i}(n)$ and $m_{2i}(n)$ for $i = 1, 2, \ldots, M$ are decimated according to the subbands number, then the forward BSS structure is applied to estimate the decimated output sub-signals $u_{1i,D}(p)$ and $u_{2i,D}(p)$ from only the decimated mixing sub-signals $m_{1i,D}(p)$, $m_{2i,D}(p)$, and finally the synthesis filter banks are used to reconstruct the estimated signals to their fullband form $u_1(n)$ and $u_2(n)$.

The decimated output sub-signals of the TC-SNLMS algorithm are given by the following formulas:

$$u_{1i,D}(p) = m_{1i,D}(p) - \mathbf{w}_{21}^T(p)\mathbf{m}_{2i,D}(p) \quad i = 1, 2, \ldots, M. \tag{13}$$

$$u_{2i,D}(p) = m_{2i,D}(p) - \mathbf{w}_{12}^T(p)\mathbf{m}_{1i,D}(p) \quad i = 1, 2, \ldots, M. \tag{14}$$

The decimated mixing sub-signals are defined as:

$$m_{1i,D}(p) = m_{1i}(pM) \quad i = 1, 2, \ldots, M. \tag{15}$$

$$m_{2i,D}(p) = m_{2i}(pM) \quad i = 1, 2, \ldots, M. \tag{16}$$

$$m_{1i}(n) = \mathbf{H}_i^T \mathbf{m}_1(n) \quad i = 1, 2, \ldots, M. \tag{17}$$

$$m_{2i}(n) = \mathbf{H}_i^T \mathbf{m}_2(n) \quad i = 1, 2, \ldots, M. \tag{18}$$

where M is the number of subbands and D is the decimator factor, $(D = M)$. The variable n is used for the time index of the original fullband mixing signals and p is used for the decimated sub-signals. $m_{1i}(pM)$ and $m_{2i}(pM)$ are the outputs of the analysis filters banks. $\mathbf{m}_1(n) = [m_1(n), m_1(n-1), \ldots, m_1(n-l+1)]$, $\mathbf{m}_2(n) =] = [m_2(n), m_2(n-1), \ldots, m_2(n-l+1)]$. l is the length of the analysis filters $\mathbf{H}_i$.

The estimated fullband signals $u_1(n)$ and $u_2(n)$ are given by the following relations:
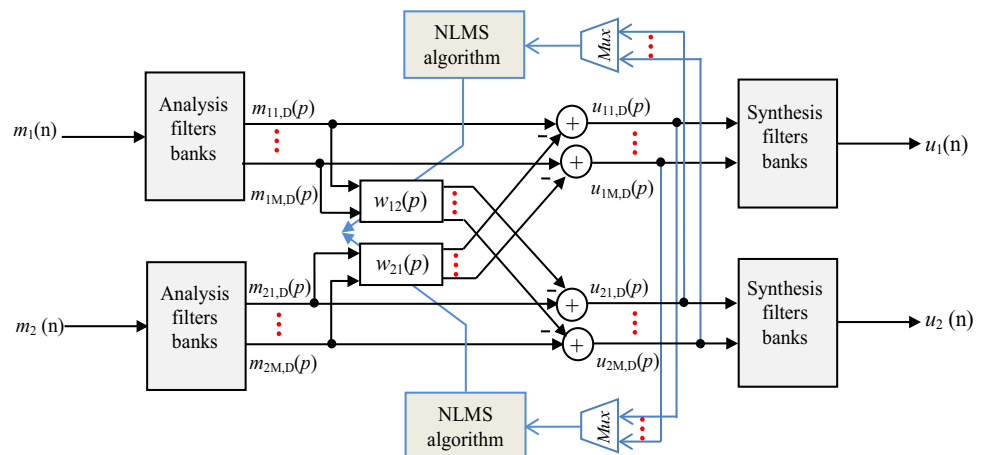
$$u_1(n) = \sum_{i=1}^M \mathbf{G}_i^T \mathbf{U}_{1i}(n) \tag{19}$$

$$u_2(n) = \sum_{i=1}^M \mathbf{G}_i^T \mathbf{U}_{2i}(n) \tag{20}$$

where

$$u_{1i}(n) = \begin{cases} u_{1i,D}(p/I), & n = 0, \pm I, \pm 2I, \ldots \\ 0 & otherwise \end{cases} \quad For\ i = 1, 2, \ldots, M. \tag{21}$$

$$u_{2i}(n) = \begin{cases} u_{2i,D}(p/I), & n = 0, \pm I, \pm 2I, \ldots \\ 0 & otherwise \end{cases} \quad For\ i = 1, 2, \ldots, M. \tag{22}$$

Fig. 4 Descriptive scheme of the TC-SNLMS algorithm

and $U_{1i}(n) = \left[u_{1i}(n), u_{1i}(n-1), \ldots, u_{1i}(n-l+1)\right]$,
$U_{2i}(n) = \left[u_{2i}(n), u_{2i}(n-1), \ldots, u_{2i}(n-l+1)\right]$,
$U_{2i}(n) = \left[u_{2i}(n), u_{2i}(n-1), \ldots, u_{2i}(n-l+1)\right]$

I is the interpolator factor, in our case we take $I = D = M$. l is the length of the synthesis filters G_i

Adopting the NLMS algorithm to update the filters $w_{21}(p)$ and $w_{12}(p)$, we get in a vector notation the following relations:

$$w_{21}(p+1) = w_{21}(p) + \mu_1 \sum_{i=1}^{M} \frac{m_{2i,D}(p) u_{1i,D}(p)}{\left|\varepsilon + m_{2i,D}^T(p) m_{2i,D}(p)\right|} \quad (23)$$

$$w_{12}(p+1) = w_{12}(p) + \mu_2 \sum_{i=1}^{M} \frac{m_{1i,D}(p) u_{2i,D}(p)}{\left|\varepsilon + m_{1i,D}^T(p) m_{1i,D}(p)\right|} \quad (24)$$

where $m_{1i,D}(p) = \left[m_{1i,D}(p), m_{1i,D}(p-1), \ldots, m_{1i,D}(p-L+1)\right]$ and $m_{2i,D}(p) = \left[m_{2i,D}(p), m_{2i,D}(p-1), \ldots, m_{2i,D}(P-L+1)\right]$. L is the length of the adaptive filters. The step-sizes $\mu_1, \mu_2 (0 < \mu_1, \mu_2 < 2)$, are the control parameters of the TC-SNLMS algorithm, which adjusts respectively, the convergence direction of the adaptive filters $w_{21}(p)$ and $w_{12}(p)$. The parameter ε is a small positive constant which allows avoiding division by very small values in absence of the input signal (silence periods).

3.2 Two-channel fullband NLMS (TC-FNLMS) algorithm

The TC-FNLMS algorithm is a two channel adaptive filtering algorithm based on the combination between the forward blind source separation structure and the normalized least mean square algorithm (Van Gerven and Van Compernolle 1992). A detailed scheme of the TC-FNLMS algorithm is presented on Fig. 5.

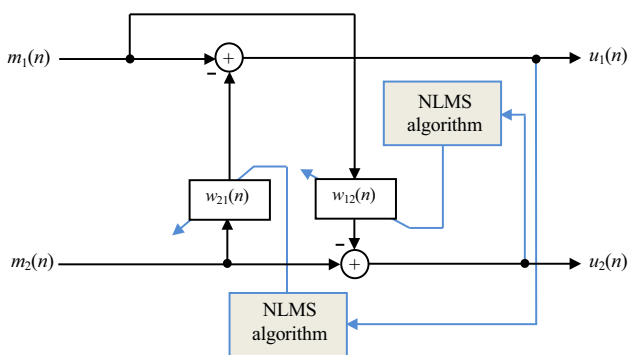The enhanced output signals of the TC-FNLMS algorithm are given by the following relations

Fig. 5 Detailed scheme of the TC-FNLMS algorithm

$$u_1(n) = m_1(n) - w_{21}^T(n) m_2(n) \quad (25)$$

$$u_2(n) = m_2(n) - w_{12}^T(n) m_1(n) \quad (26)$$

where $m_1(n) = \left[m_1(n), m_1(n-1), \ldots, m_1(n-L+1)\right]^T$ and $m_2(n) = \left[m_2(n), m_2(n-1), \ldots, m_2(n-L+1)\right]^T$ are the vectors that contain the last L samples of the inputs $m_1(n)$ and $m_2(n)$, respectively.

The update relations of the adaptive filters $w_{21}(n)$ and $w_{12}(n)$ are given as follows:

$$w_{21}(n+1) = w_{21}(n) + \mu_{21} \frac{m_2(n) u_1(n)}{\left|\alpha + m_2^T(n) m_2(n)\right|} \quad (27)$$

$$w_{12}(n+1) = w_{12}(n) + \mu_{12} \frac{m_1(n) u_2(n)}{\left|\alpha + m_1^T(n) m_1(n)\right|} \quad (28)$$

where $0 < \mu_{12}, \mu_{21} < 2$ are the two step-size that control the convergence behavior of the cross-adaptive filters $w_{21}(n)$ and $w_{12}(n)$. The parameter α is a small constant introduced to avoid division by zero.

4 Simulation results

4.1 Description of the used signals

In this simulation, the original speech signal $s(n)$ is a French sentence of about 4 s length, taken from AURORA database (Combescure 1981), and it is presented on Fig. 6 with its manual segmentation. For the punctual noise source signal $b(n)$ the USASI noise (United state of America Standard Institute now (ANSI)) is used (see Fig. 6). These signals are sampled at 8 kHz and coded on 16 bits. To generate the impulse responses $h_{12}(n)$ and $h_{21}(n)$, we have used the physical model described in Djendi et al. (2006). An example of these impulse responses is given in Fig. 7.

We have used the signals described above to generate the noisy observations $m_1(n)$ and $m_2(n)$ according to the simplified convolutive mixture model (describe in Sect. 2). In Fig. 8 we show the time evolution of the two noisy observations $m_1(n)$ and $m_2(n)$. The input signal-to-noise-ratio (SNR) is selected to be $SNR_1 = 0\ dB$ and $SNR_2 = 0\ dB$ at the first and the second microphone, respectively.

4.2 Time evolution of the output speech signal

In this section, we present a simple visual test on the output signal obtained by the proposed switching algorithm. As we are interested on speech enhancement, we focus only on the first output $u_1(n)$. Parameters setting of the proposed
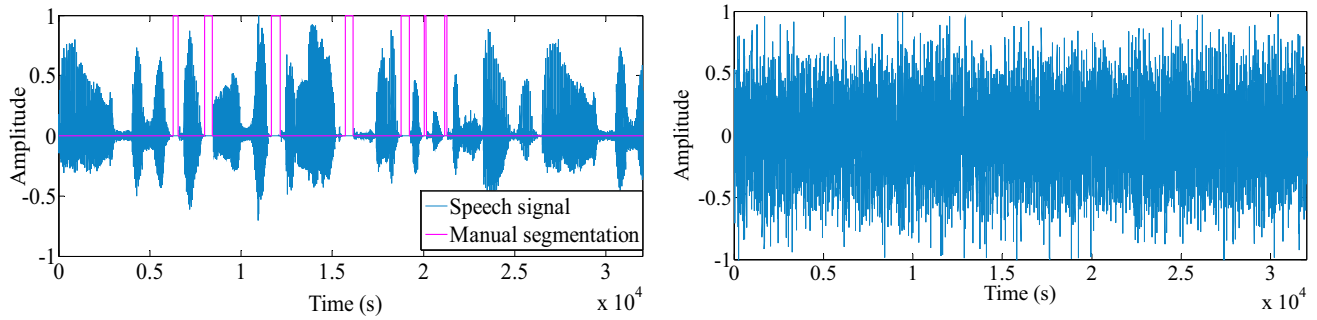
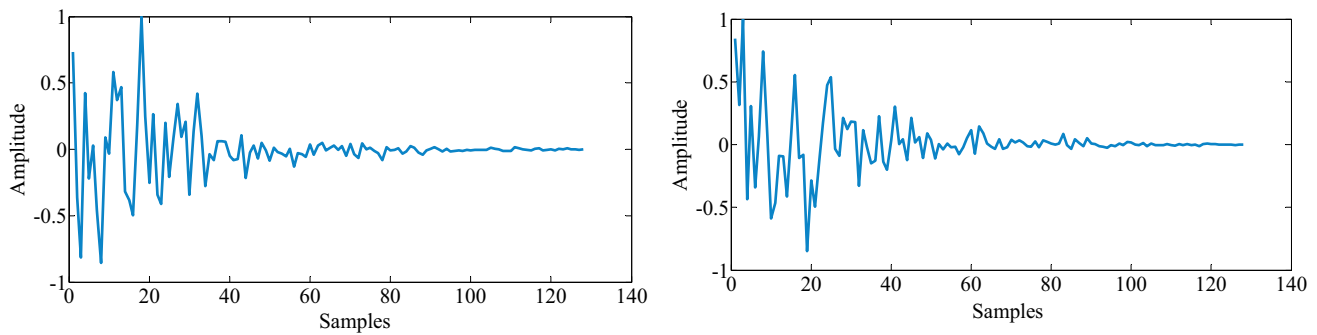**Fig. 6** (in left) The original speech signal with its segmentation, (in right) the USASI noise

Fig. 7 Example of the simulated impulse responses, (in left) $h_1(n)$ and (in right) $h_2(n)$, with $L = 128$
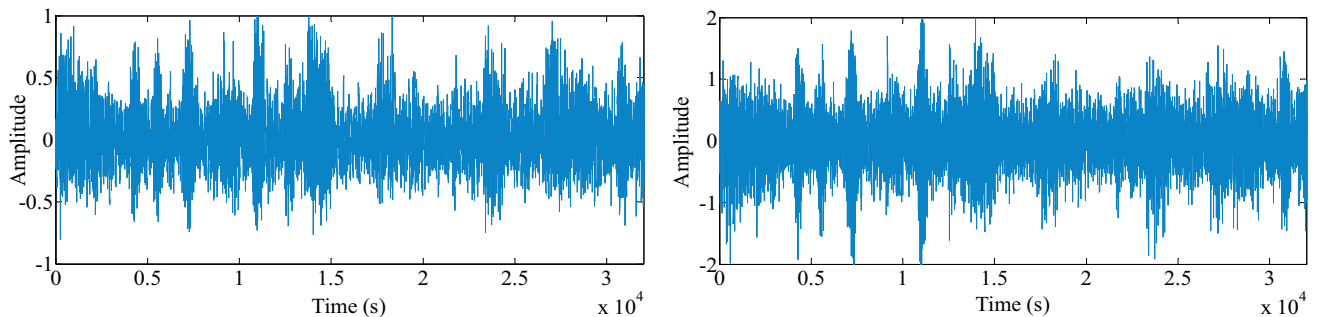

Fig. 8 Time evolution of the two noisy observations, (in left) $m_1(n)$ and (in right) $m_2(n)$

algorithm are given as follows: the adaptive filters length is $L = 128$, the subband filters length for $M = 2$, $M = 4$, $M = 8$ are respectively: $l = 16$, $l = 32$, $l = 64$, the MSE threshold value is: $MSE_{th} = -45$ dB, the input SNR is selected to be 0 dB at the two microphones. In Fig. 9, we show the time evolution of the output speech signal $u_1(n)$ obtained by the proposed switching algorithm with 2, 4 and 8 subband configurations. This figure shows the good behavior of the proposed algorithm in reducing the acoustic noise components with all subband configurations.

4.3 Objective evaluation

In order to evaluate objectively the performances of the proposed switching algorithm in comparison with the classical two-channel fullband NLMS (TC-FNLMS) algorithm and the two-channel subband NLMS (TC-SNLMS) algorithm, intensive experiments have been done in terms of the following objective criteria:

(i)   The system mismatch (SM),
(ii) The segmental mean square error (SegMSE),
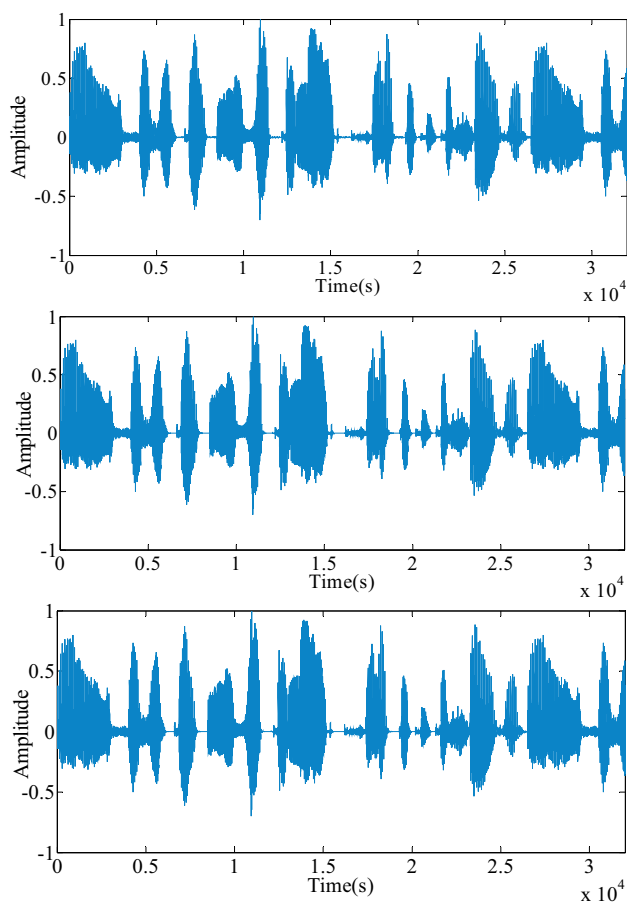(iii) The segmental signal to noise ratio (SegSNR),

Fig. 9 Time evolution of the output speech signal $u_1(n)$ obtained by the proposed switching algorithm with: (in top) 2 subbands, (in middle) 4 subbands, (in bottom) 8 subbands

(iv)   The Cepstral distance (CD).

4.3.1 System mismatch (SM) evaluation

To assess the speed convergence performance of the proposed switching algorithm in comparison with the TC-FNLMS and TC-SNLMS algorithms, we have used the SM criterion which is computed between the adaptive filter $w_{21}(n)$ and the real one $h_{21}(n)$ as follows [31]:

$$SM_{dB} = 20 log_{10}\left(\frac{\boldsymbol{h}_{21} - \boldsymbol{w}_{21}(n)}{\boldsymbol{h}_{21}} \right) \quad (29)$$

where the symbol ‖. ‖ represent the mathematical Euclidean norm operator. The simulation parameters of each algorithm (i.e. TC-FNLMS, TC-SNLMS and proposed switching algorithm) are summarized in Table 1. The obtained results of the SM comparison between TC-FNLMS, TC-SNLMS algorithms and proposed switching algorithm for three input *SNRs* i.e. -3 dB, 0 dB and 3 dB, and with different

subband configurations (2, 4 and 8 subbands) are reported on Figs. 10, 11 and 12. It can be obviously seen from these figures that the proposed algorithm with (2, 4 and 8 subbands) outperforms the other simulated algorithms in terms of speed convergence for all input SNR levels. This good performance of the proposed algorithm is achieved thanks to the fullband–subband switching procedure based on MSE estimation, where, the use of the TC-SNLMS algorithm when the MSE_e is superior than the MSE_{th}, and the alternative use of the TC-FNLMS algorithm in the opposite case allows getting a fast speed convergence in transient and steady phases.

4.3.2 Segmental mean square error (SegMSE) evaluation

In this subsection, we have evaluated the mean square error criterion for the TC-FNLMS, TC-SNLMS algorithms and the proposed one. The segmental mean square error (SegMSE) allows to quantify the speed convergence performance of each simulated algorithm. This SegMSE criterion is given by the following relation:

$$SegMSE_{dB} = \frac{10}{M}\sum_{m=0}^{M-1} \log_{10}\left(\frac{1}{N}\sum_{n=Nm}^{Nm+N-1} |s(n) - u_1(n)|^2 \right) \quad (30)$$

where N is the segment length of the original signal $s(n)$ and the enhanced one $u_1(n)$, and M represent the number of segments in silence periods. We note that the SegMSE criterion is evaluated only in the absence speech periods (Ghribi et al. 2016). The simulation parameters of each algorithm are summarized in Table 1. Considering Figs. 13, 14 and 15 which show the SegMSE evaluation of the TC-FNLMS, TC-SNLMS (with 2, 4, 8 subbands) and the proposed switching algorithm (with 2, 4, 8 subbands) for three input SNRs i.e. − 3 dB, 0 dB and 3 dB, we notice that the proposed algorithm has better speed convergence performance in comparison with the other ones. This remarks is observed with the entire test when $M = 2, 4$ and 8.

4.3.3 Segmental signal-to-noise-ratio (SegSNR) evaluation

In order to analyze the noise reduction performance of the proposed algorithm in comparison with the TC-FNLMS and TC-SNLMS algorithms, we have used the SegSNR criterion, which is evaluated for each algorithm as follows (Deller et al. 1993; Sayed 2003):

$$SegSNR_{dB} = \frac{10}{M}\sum_{m=0}^{M-1} \log_{10}\left(\frac{\sum_{n=Nm}^{Nm+N-1} |s(n)|^2}{\sum_{n=Nm}^{Nm+N-1} |s(n) - u_1(n)|^2} \right) \quad (31)$$

where $s(n)$ and $u_1(n)$ are the original and the enhanced speech signals, respectively. The parameters M and N are the

Table 1 Simulation parameters of the simulated algorithms i.e. TC-FNLMS algorithm (Van Gerven and Van Compernolle 1992), TC-SNLMS algorithm (Djendi and Bendoumia 2013), proposed switching algorithm (In this paper)

| Simulation parameters	
Input signals	Original speech signal: a French sentence phonetically equilibrated Noise signal: USASI noise Sampling frequency rate: $fs = 8$ kHz Input $SNRs$: $SNR_1 = SNR_2 = -3$ dB, 0 dB and 3 dB
TC-FNLMS algorithm (Van Gerven and Van Compernolle 1992)	Adaptive filters length: $L = 128$ Fixed step-sizes: $\mu_{12} = \mu_{21} = 0.9$ Positive constant: $\alpha = 10^{-6}$
TC-SNLMS algorithm (Djendi and Bendoumia 2013)	Adaptive filters length: $L = 128$ Subband filters length for $M = 2$, $M = 4$ and $M = 8$ are respectively: $l = 16$, $l = 32$ and $l = 64$ Fixed step-sizes: $\mu_1 = \mu_2 = 0.9$ Positive constant: $\varepsilon = 10^{-6}$
Proposed switching algorithm (In this paper)	Switching mechanism
	MSE threshold: $MSE_{th} = -45$ dB Smoothing factor: $\beta = 0.05$
	TC-SNLMS algorithm
	Adaptive filters length: $L = 128$ Subband filters length for $M = 2$, $M = 4$ and $M = 8$ are respectively: $l = 16$, $l = 32$ and $l = 64$ Fixed step-sizes: $\mu_1 = \mu_2 = 0.9$ Positive constant: $\varepsilon = 10^{-6}$
	TC-FNLMS algorithm
	Adaptive filters length: $L = 128$ Fixed step-sizes: $\mu_{12} = \mu_{21} = 0.9$ Positive constant: $\alpha = 10^{-6}$
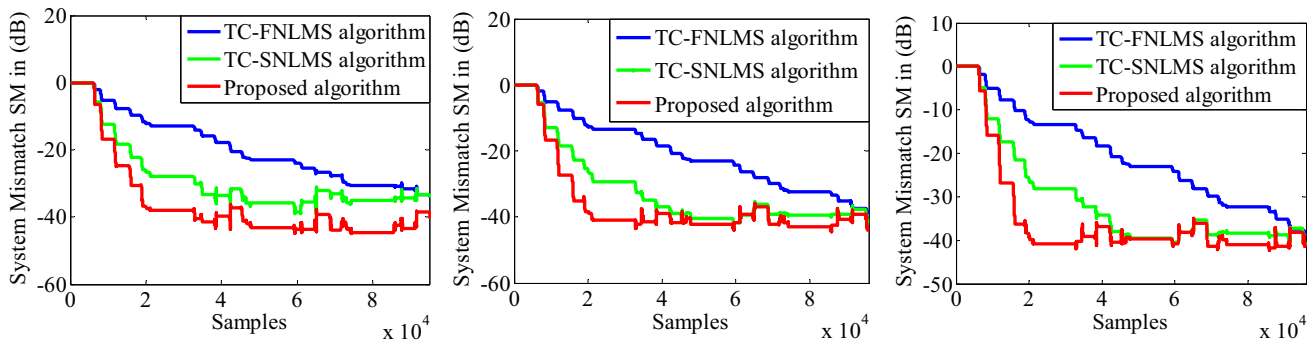
Fig. 10 SM evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$, (in right) $M = 8$. Input $SNR_1 = SNR_2 = -3$ dB

number of segments and the segment length, respectively. We note that at the output, we get M values of the Seg-SNR criterion, each one is mean averaged on 'N' samples. The symbol $|\cdot|$ represents the absolute operator. We recall here that all the 'M' segments correspond to only speech signal presence periods. The symbol $log10$ is the base 10 logarithm. Figures 16, 17 and 18 show the SegSNR evaluation of the proposed algorithm in comparison with the TC-FNLMS and TC-SNLMS ones for three global input $SNRs$

i.e. -3 dB, 0 dB and 3 dB. For each algorithm, we use the same parameters given in Table 1.

According to the obtained results, we can say that the proposed algorithm behaves more efficiently than the other competitive algorithms (TC-FNLMS and TC-SNLMS) for different input SNR levels i.e. -3 dB, 0 dB and 3 dB. We have also noted that the output SegSNR values of the TC-SNLMS algorithm decrease in the steady state regime when the number of subbands is selected high (4,
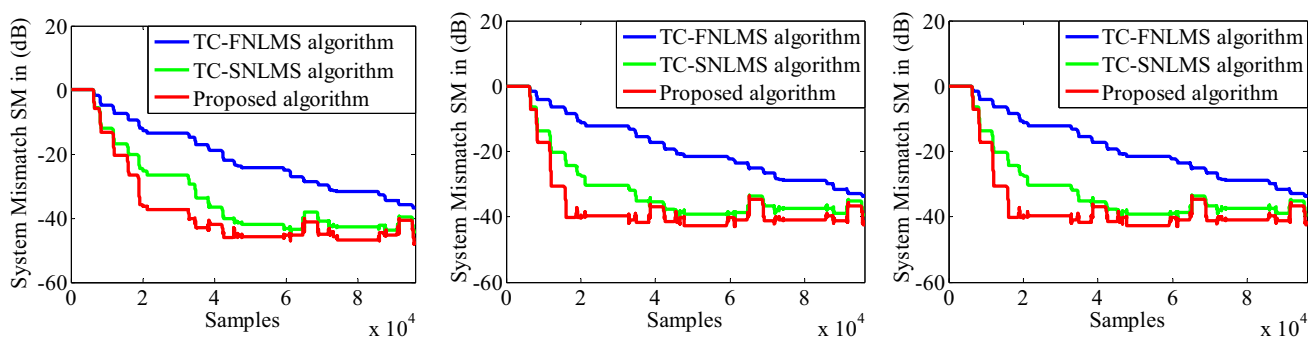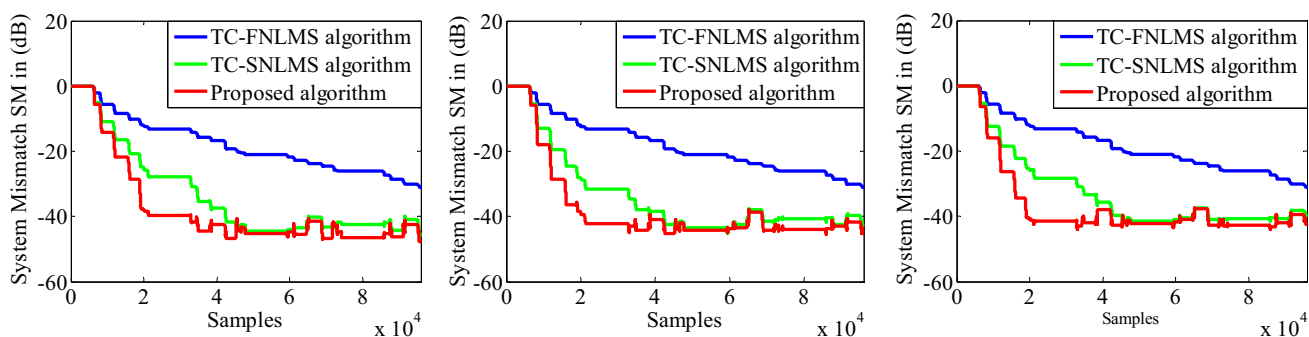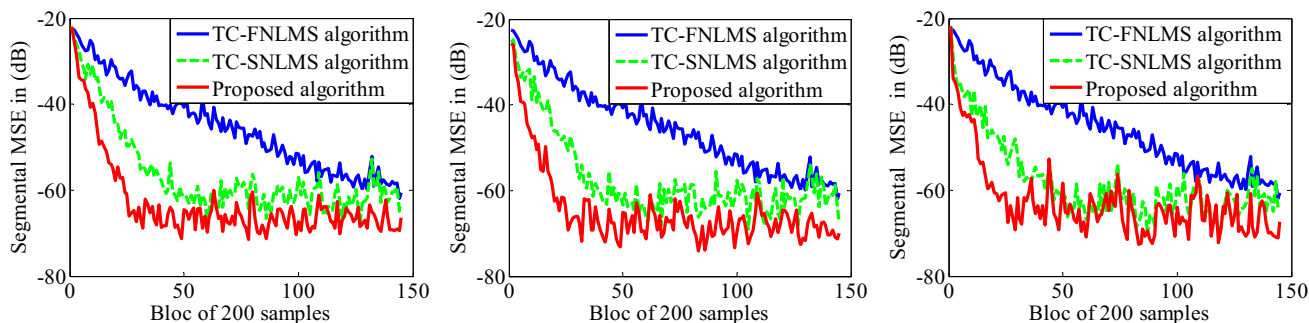
Fig. 11 SM evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different sub-band configurations, (in left) $M = 2$, (in middle) $M = 24$, (in right) $M = 28$. Input $SNR_1 = SNR_2 = 0$ dB

Fig. 12 SM evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$ (in right) $M = 8$. Input $SNR_1 = SNR_2 = 3$ dB

Fig. 13 SegMSE evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$, (in right) $M = 8$. Input $SNR_1 = SNR_2 = -3$ dB

8 subbands), however the proposed switching algorithm with (2,4 and 8 subbands) has given the higher values of the SegSNR in transient and steady phases. This is the main benefit of the proposed switching algorithm that aim to combine the good convergence performance of the TC-SNLMS algorithm when the number of subbands is selected high with the good final values of the TC-FNLMS algorithm.
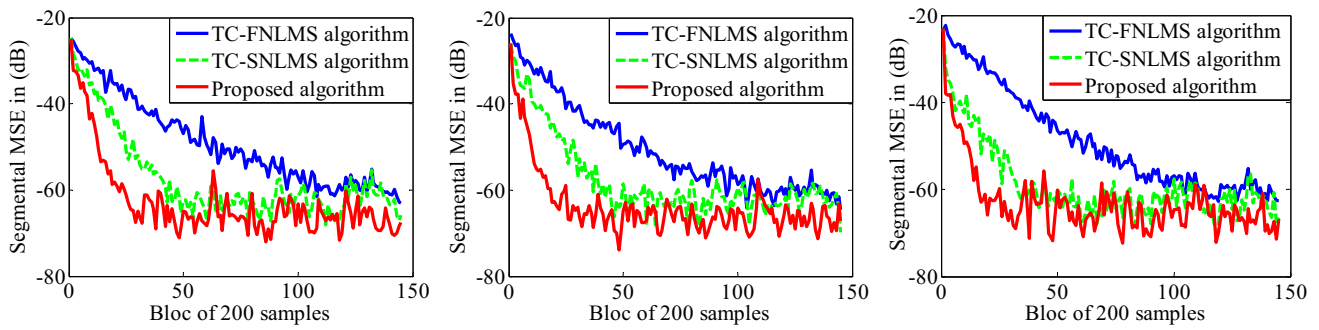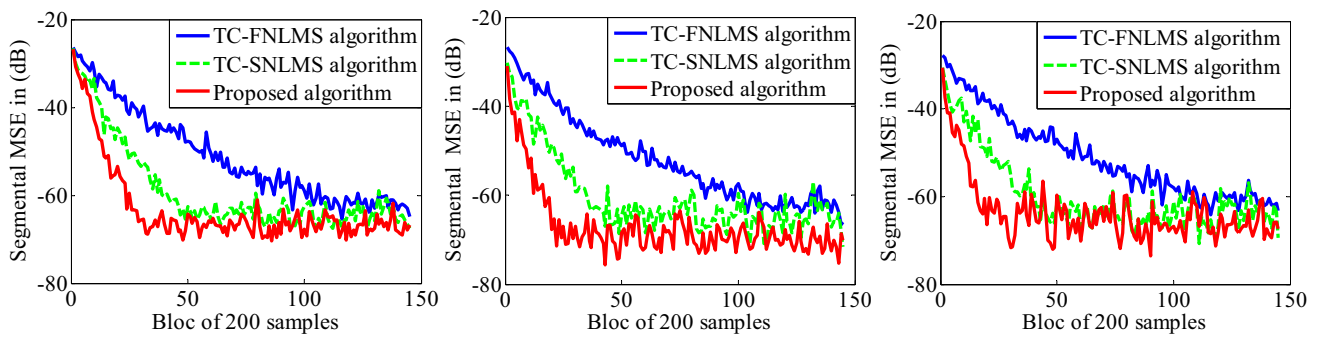
Fig. 14 SegMSE evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$, (in right) $M = 8$. Input $SNR_1 = SNR_2 = 0$ dB

Fig. 15 SegMSE evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching [in this paper] algorithms, with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$, (in right) $M = 28$. Input $SNR_1 = SNR_2 = 3$ dB
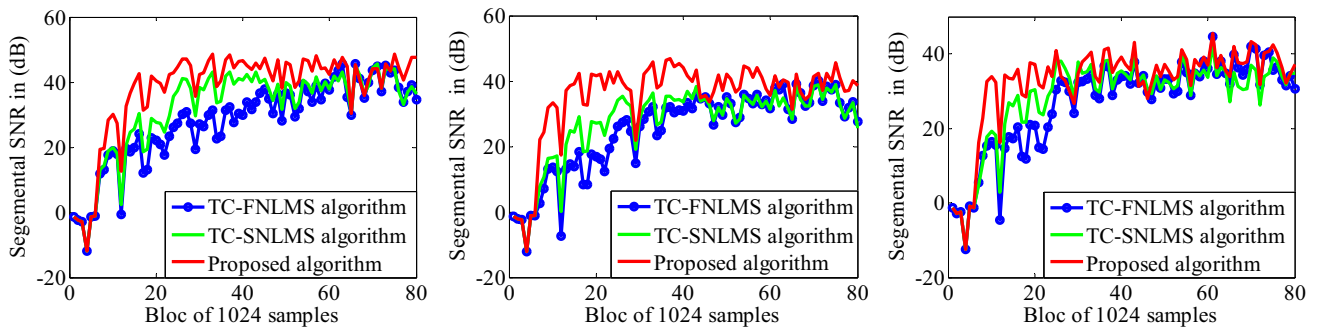

Fig. 16 SegSNR evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper), algorithms with different subband configurations, (in left) $M = 2$, (in middle) $M = 4$, (in right) $M = 8$. Input $SNR_1 = SNR_2 = -3$ dB

4.3.4 Cepstral distance (CD) evaluation

In order to quantify the distortion amount introduced in the output speech signal obtained by the proposed switching algorithm in comparison with the TC-FNLMS and TC-SNLMS ones, we have used the cepstral distance criterion which is estimated by the following relation (Hu and Loizou 2008; Rabiner and Juang 1993):

$$CD_{dB} = \sum_{\lambda=0}^{T-1} IFFT \left[\log \left(S|(\lambda, \omega)| \right) - \log \left(U_1|(\lambda, \omega)|VAD_\lambda \right) \right]^2$$
(32)

where $S(\lambda, \omega)$ and $U_1(\lambda, \omega)$ represent the short Fourier transform of the original speech signal $s(n)$ and the enhanced one $u_1(n)$ respectively at each frame λ, and T is the mean
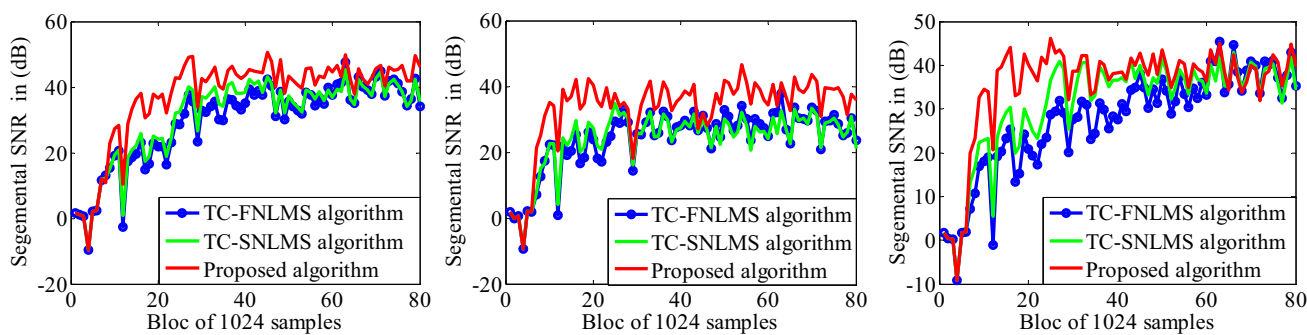
Fig. 17 SegSNR evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M=2$, (in middle) $M=4$, (in right) $M=28$. Input $SNR_1 = SNR_2 = 0$ dB
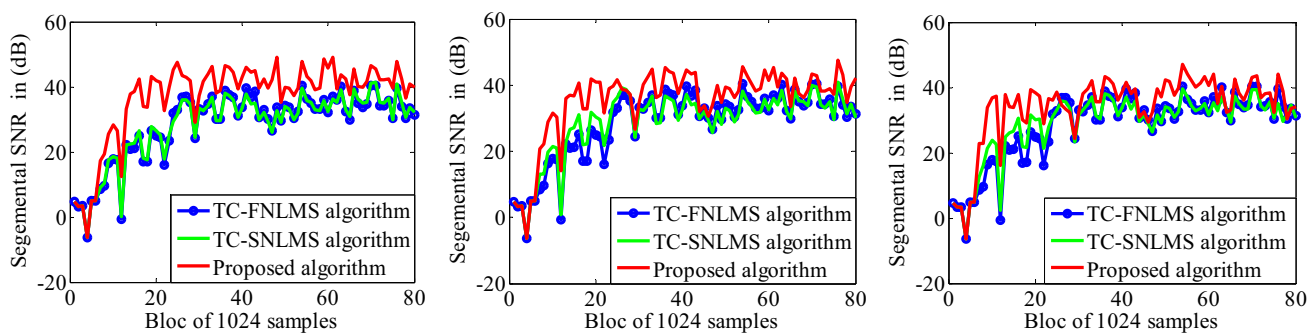

Fig. 18 SegSNR evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M=2$, (in middle) $M=4$, (in right) $M=82$. Input $SNR_1 = SNR_2 = 3$ dB

Fig. 19 CD evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M=2$, (in middle) $M=4$, (in right) $M=8$. Input $SNR_1 = SNR_2 = -3$ dB
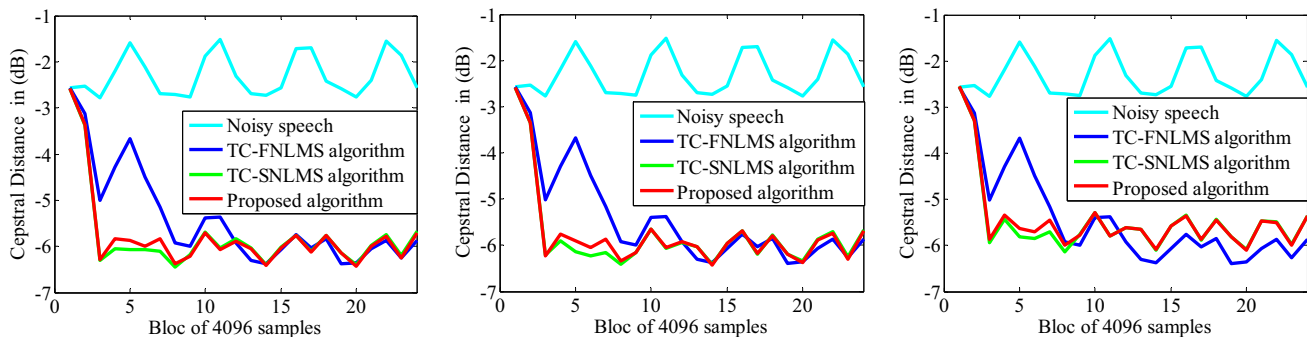
averaging value of the *CD* criterion, and *VAD* parameter is a voice activity detector. We have reported on Figs. 19, 20 and 21 the CD evaluation results obtained by the three algorithms for three input SNRs i.e. −3 dB, 0 dB and 3 dB and with different subband configurations (2,4 and 8 subbands). We recall that the simulation parameters of each algorithm are the same as given by Table 1. These results show that the TC-FNLMS algorithm outperforms the other algorithms (i.e. TC-SNLMS and proposed algorithm) in terms of steady-state CD values. Also a close behavior of the proposed algorithm with the TC-SNLMS one is noted. It is worth noting that the CD values of the proposed algorithm for all subband configurations and in divers situations are
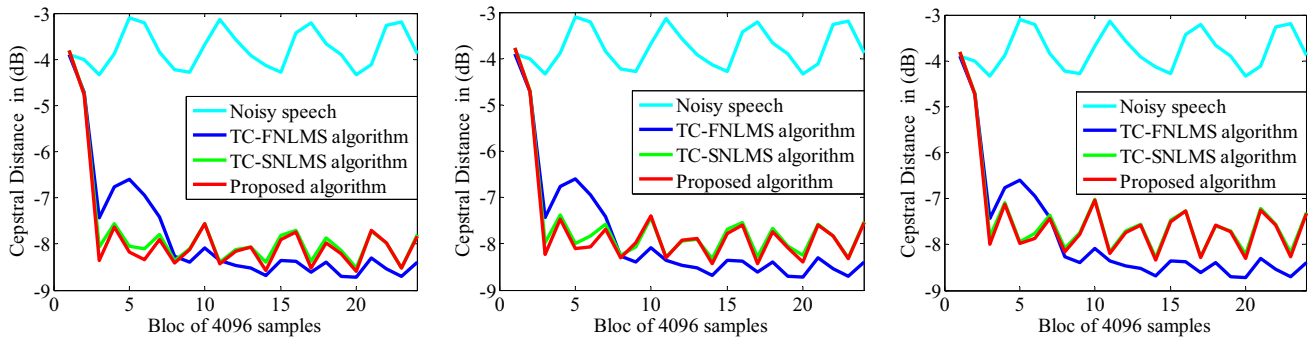
Fig. 20 CD evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M=2$, (in middle) $M=24$, (in right) $M=8$. Input $SNR_1 = SNR_2 = 0$ dB

Fig. 21 CD evaluation of the TC-FNMLS (Van Gerven and Van Compernolle 1992), TS-SNLMS (Djendi and Bendoumia 2013) and the proposed switching (in this paper) algorithms, with different subband configurations, (in left) $M=2$, (in middle) $M=4$, (in right) $M=8$. Input $SNR_1 = SNR_2 = 3$ dB
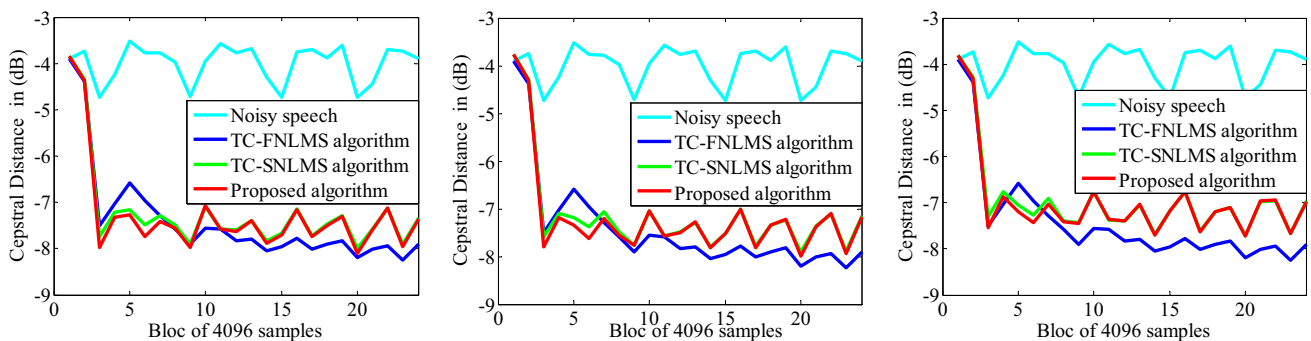
below − 5 dB. This indicates the good intelligibility of the output speech signal.

5 Conclusion

In this paper, we have proposed a new switching adaptive speech enhancement algorithm, wherein, the two-channel fullband NLMS (TC-FNLMS) algorithm and the two-channel subband NLMS (TC-SNLMS) algorithm are switched alternatey according to the estimated MSE. To validate the performance of the proposed switching algorithm in comparison with TC-FNLMS and TC-SNLMS algorithms, intensive experiments have been performed using several objective criteria. The obtained results of the SM and the segmental MSE have confirmed the superiority of the proposed algorithm in terms of convergence speed, this good performance is obtained thanks to the proposed fullband–subband switching technique. The SegSNR evaluation has also proved the efficiency of the proposed algorithm on reducing the acoustic noise at the

processing output. Unfortunately, for the CD evaluation we have noted a slight degradation on the performance of the proposed algorithm in the steady state regime. In a future work, we aim to address the issue and improve the proposed switching algorithm in the situation where it fails.

References

Benesty, J., & Cohen, I. (2018) Single-channel speech enhancement in the time domain. In: *Canonical correlation analysis in speech enhancement. Springer Briefs in Electrical and Computer Engineering*. Cham: Springer.

Combescure, P. (1981). 20 listes de dix phrases phonétiquement équilibrées. *Revue d'Acoustique, 56,* 34–38.

Deller, J., Proakis, J., & Hansen, J. (1993). *Discrete time processing of speech signals*. New York: MacMillan Publishing.

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. (2018). An efficient wavelet-based adaptive filtering algorithm for automatic blind speech enhancement. *International Journal of Speech Technology, 21,* 355–367.

Djendi, M., & Bendoumia, R. (2013). A new adaptive filtering sub-band algorithm for two channel acoustic noise reduction and speech enhancement. *Computers & Electrical Engineering, 39*(8), 2531–2550.

Djendi, M., & Bendoumia, R. (2016). Improved subband-forward algorithm for acoustic noise reduction and speech quality enhancement. *Applied Soft Computing, 42,* 132–143.

Djendi, M., Gilloire, A., & Scalart, P. (2006). Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms. *IEEE Int. Conf. ICASSP, Toulouse, France, 14–19 May 2006 (Vol. 3, pp. 744–748).*

Djendi, M., Gilloire, A., & Scalart, P. (2007). New frequency domain post-filters for noise cancellation using two closely spaced microphones. *Proc. EUSIPCO, Poznan, 3–8 Sep. (Vol. 1, pp. 218–221).*

Djendi, M., Henni, R., & Sayoud, A. (2016). A new dual forward BSS based RLS algorithm for speech enhancement. *International Conference on Engineering and MIS, ICEMIS, Agadir, Morooco.*

Djendi, M., & Zoulikha, M. (2018). A new efficient backward BSS crosstalk-resistant algorithm for automatic blind speech quality enhancement. *International Journal of Speech Technology, 21,* 809–823.

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.

Habets, E. A. P., & Benesty, J. (2013). Multi-microphone noise reduction based on orthogonal noise signal decompositions. *IEEE Transactions on Audio, Speech and Language Processing, 21*(6), 1123–1133.

Henni, R., Djendi, M., & Djebari, M. (2019). A new efficient two-channel fast transversal adaptive filtering algorithm for blind speech enhancement and acoustic noise reduction. *Computer and Electrical Engineering, 73,* 349–368.

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.

Kim, L. H., & Hasegawa-Johnson, M. (2012). *Optimal multi-microphone speech enhancement in cars. Digital signal processing for in-vehicle systems and safety* (pp. 195–204). Berlin: Springer.

Kuo, S. M., & Peng, W. M. (1999). Asymmetric crosstalk-resistant adaptive noise canceller. *Proc. IEEE workshop on Signal Processing System*, pp. 605–614.

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*. New York: CRC Press.

Loizou, P. C. (2017). *Speech enhancement: theory and practice* (2nd ed.). Boca Raton: CRC Press, Taylor & Francis Group.

Man Kima, S., & Kook Kim, H. (2014). Noise variance estimation based on dual-channel phase difference for speech enhancement. *Digital Signal Processing, 26,* 169–182.

Nabi, W., Aloui, N., & Cherif, A. (2016). Speech enhancement in dual-microphone mobile phones using Kalman filter. *Applied Acoustics, 109,* 1–4.

Nabi, W., Aloui, N., & Cherif, A. (2017). An improved speech enhancement algorithm for dual-channel mobile phones using wavelet and genetic algorithm. *Computers & Electrical Engineering, 62,* 692–705.

Nabi, W., BenNasr, M., Aloui, N., & Cherif, A. (2018). A dual-channel noise reduction algorithm based on the coherence function and the bionic wavelet. *Applied Acoustics, 131,* 186–191.

Ni, J., & Li, F. (2010). A variable step-size matrix normalized subband adaptive filter. *IEEE Transactions on Audio, Speech and Language Processing, 18*(6), 1290–1299.

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Englewood Cliffs: Prentice-Hall.

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

Sayoud, A., Djendi, M., Medahi, S., & Guessoum, A. (2018). A dual fast NLMS adaptive filtering algorithm for blind speech quality enhancement. *Applied Acoustics, 135,* 101–110.

Seo, J. H., & Park, P. G. (2014). Variable individual step-size subband adaptive filtering algorithm. *Electronics Letters, 50*(3), 177–178.

Syskind Pedersen, M., Larsen, J., Kjems, U., & Parra, L. C. (2007). *A survey of convolutive blind source separation methods. Springer Handbook on speech processing and speech communication.* Cham: Springer.

Thumchirdchupong, H., & Tangsangiumvisai, N. (2013). A two-microphone noise reduction scheme for hands-free telephony in a car environment. *10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology.*

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

Van Gerven, S,. & Van Compernolle, D. (1992). Feedforward and Feedback in a symmetric adaptive noise canceller: Stability analysis in a simplified case. *Proc. IEEE. EUSIPCO, Belgium, Brussels (Vol. 24–27, pp. 1081–1084).*

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.

Yu, Y., Zhao, H., & Chen, B. (2016). A new normalized subband adaptive filter algorithm with individual variable step sizes. *Circuits, Systems and Signal Process, 35*(4), 1407–1418.