A New Two-Microphone Reduce Size SMFTF Algorithm for Speech Enhancement in New Telecommunication Systems

Zineddine Guernaz(✉) and Xuanli Wu

School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin, Heilongjiang, China
zinoudrai@outlook.com, xlwu2002@hit.edu.cn

Abstract. This paper considers the problem of speech enhancement and noise reduction in speech recognition systems and 5G mobile communication systems. The presence of these systems in a noisy environment reduces their effectiveness and makes degradation in their performance. here, we propose a new contribution to resolve noise reduction and speech enhancement problem in these systems by proposing a new algorithm. The proposed two microphones reduce size simplified fast transversal filter (TM-RSMFTF) algorithm is an outcome of the good combination between the well-known forward blind source separation structure and the adaptive algorithm reduce size simplified fast transversal filter properties which is a stable version of fast transversal filter (FTF) algorithms. The proposed algorithm has low computational complexity. The simulation results show a good performances and effectiveness of this new TM-RSMFTF algorithm in comparison with conventional TM-NLMS algorithm and almost similar performances with full-size TM-SFTF in terms of various objectives criteria such as Segmental SNR, System Mismatch, Segmental MSE.

Keywords: Noise reduction · Speech enhancement · TM-RSMFTF

1 Introduction

Nowadays, many applications such as 4G and 5G telecommunication systems, speech recognition systems, VoIP and teleconferencing systems, insist the presence of speech enhancement and acoustic noise reduction methods because these applications are designed for the quiet environment but in real life, the presence of different noise sources corrupt the speech signal causing degradation in the performances of these systems.

In the last decades, a lot of algorithms and techniques have been developed in the literature for speech enhancement with the first goal of raising speech quality such as spectral subtraction techniques [1], Wiener filter [2], minimum mean-square error estimator [3, 4]. We found also the most powerful speech enhancement techniques which are based on the adaptive filter [5]. The coefficients of the adaptive filter can be adjusted automatically by using various adaptive algorithms.

S. Han et al. (Eds.): AICON 2019, LNICST 286, pp. 186–202, 2019.
https://doi.org/10.1007/978-3-030-22968-9_17

Several adaptive algorithms have been proposed, the most popular one is normalized least mean square (NLMS) [5], and it is characterized by low computational complexity and low convergence speed. Another popular algorithms are recursive least square (RLS) algorithms [5], the main drawback of these algorithms is high computation complexity.

Many fast recursive least mean square algorithms have been proposed to resolve the high computational complexity problem such as the fast Kalman [6], fast transversal filter (FTF) [7], and the main drawback of these algorithms is the numerical instability. In the literature, several solutions have been proposed to solve this problem [8, 9].

To solve acoustic noise reduction and speech enhancement problem, a stable version of FTF algorithm, i.e., the simplified FTF algorithm (SMFTF) was proposed in [10]. In this algorithm, the adaptation gain is evaluated by using only the forward prediction variables and discarding the backward prediction and also by adding regularization constant and leakage factor. In the same work, reduce size simplified FTF (R-SMFTF) algorithm was proposed by reducing the size of forward predictor and calculating two likelihood variables, the first one was used to update the transversal filter and the forward predictor order P and the second one was used to update the forward prediction error variance, and this algorithm shows numerical stability and similar performance with FRLS algorithms with low computational complexity $(2L + 5P)$ where L is the length of the adaptive filter, and P is the size of forward predictor.

To reduce more computational complexity for SMFTF and R-SMFTF algorithms, the author in [11] proposed new algorithms similar to SMFTF and R-SMFTF with new relations to calculate the likelihood variables which reduce the computational complexity of simplified FTF algorithm to $6L$ and $2L + 4P$ for reduce size simplified FTF algorithm, respecively.

To reduce noise and enhance speech, several algorithms based on the combination of forward blind source separation structure and various adaptive algorithms have been proposed [12, 13]. In this work, we propose a new two-microphones reduce size simplified fast transversal filter (TM-RSMFTF) algorithm for speech enhancement and noise reduction application. This proposed algorithm obtained by the good combination between the reduce size simplified FTF algorithm and the forward blind sources separation (FBSS) structure.

The low computational complexity and the good performances shown in the simulation results of this new algorithm makes it a good choice for implementation in the 4G and 5G mobile communication and speech recognition systems.

This work is organized as follows. Section 2 presents the simplified model of a convolutive mixture that we used. In Sect. 3, FBSS is described. Section 4 discusses the proposed algorithm and in Sect. 5, we highlight the computational complexity. In Sect. 6, we present simulation results and comparative study of the proposed algorithm with two-microphone algorithms. Finally, we present the conclusion in Sect. 7.

2 The Simplified Model of a Convolutive Mixture

In our work, we consider the situation of two microphones and two uncorrelated signals which are speech signal $s(n)$ and noise signal $b(n)$, we consider that the speech signal is captured by two sensors, and the useful signal is near to the first sensor and the unwanted signal is near to the other sensor, this mixing model [14–17] is shown in Fig. 1.

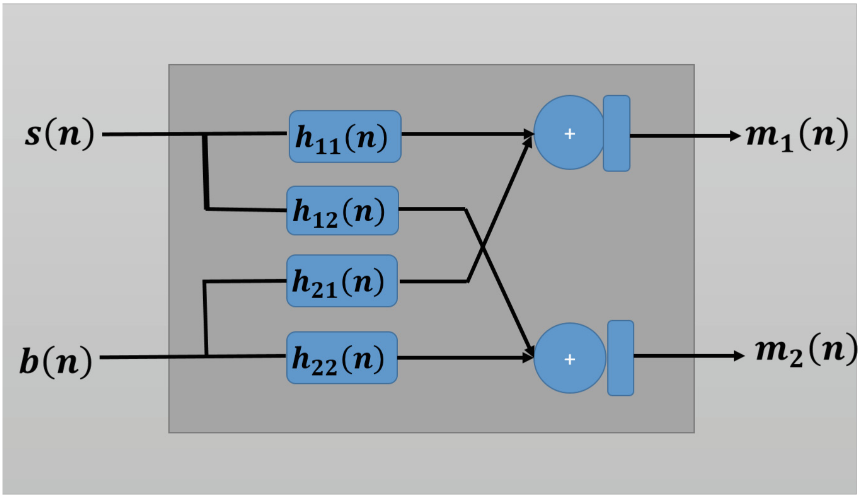

Fig. 1. The mixture model with two microphones

The two convolutive observations we obtained at the output of this mixing model are given by the following equations:

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

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

where, $s(n)$ is the source of speech and $b(n)$ is the source of noise, $h_{11}(\mathrm{n})$ and $h_{22}(n)$ represent the impulse responses of each direct channel respectively and they assumed to be identity $(h_{11}(n) = h_{22}(n) = \delta(n))$, $h_{21}(n)$ and $h_{12}(n)$ represent the cross-coupling effects between the 2 channels. The symbol * represents the convolution operation.

3 The Forward Blind Sources Separation (FBSS) Structure

The widely known forward blind source is effectively combined with many different algorithms to reduce noise and enhance the performance of speech. The forward BSS structure is shown in Fig. 2, and the aim of two channels forward BSS is to restore $s(n)$ and $b(n)$ only from the two convolutive observations [16]. For identification of the two real impulse response $h_{12}(n)$ and $h_{21}(n)$, we use the two adaptive filters of the two channel forward BSS structure.

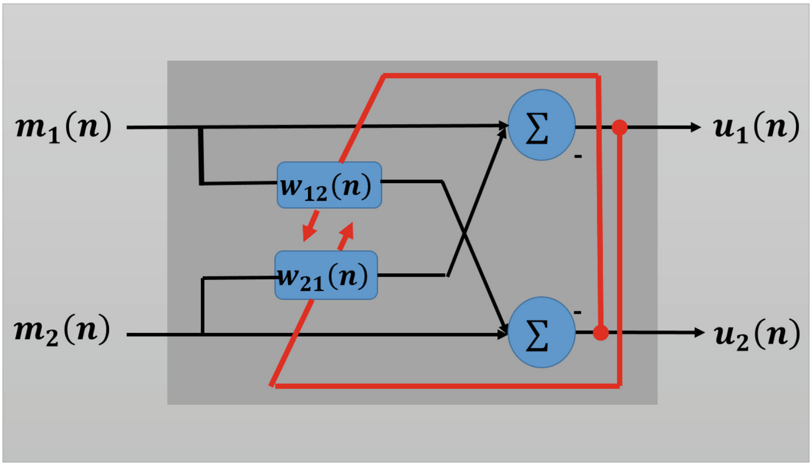

Fig. 2. The forward blind sources separation (FBSS) structure

The two outputs of this structure are given by:

$$u_1(n) = m_1(n) - m_2(n) * w_{21}(n), \tag{3}$$

$$u_2(n) = m_2(n) - m_1(n) * w_{12}(n), \tag{4}$$

where, $(*)$ represents the convolution operation. By inserting (1) and (2) into (3) and (4), we can obtain the two outputs equations $u_1(n)$ and $u_2(n)$ as follows:

$$u_1(n) = b(n) * [h_{21}(n) - w_{21}(n)] + s(n) * [\delta(n) - h_{12}(n) * w_{21}(n)], \tag{5}$$

$$u_2(n) = s(n) * [h_{12}(\mathrm{n}) - w_{12}(n)] + b(n) * [\delta(n) - h_{21}(n) * w_{12}(n)]. \tag{6}$$

If we use an optimal assumption for the two adaptive filters $(h_{21}(n) = w_{21}(n))$ and $(h_{12}(\mathrm{n}) = w_{12}(n))$, in this situation the two outputs become:

$$u_1(n) = s(n) * [\delta(n) - h_{12}(n) * w_{21}(n)], \tag{7}$$

$$u_2(n) = b(n) * [\delta(n) - h_{21}(n) * w_{12}(n)]. \tag{8}$$

where, $\varphi_1(n)$ and $\varphi_2(n)$ are post filters which are given respectively by:

$$\varphi_1(n) = \delta(n) - h_{12}(n) * w_{21}(n), \tag{9}$$

$$\varphi_2(n) = \delta(n) - h_{21}(n) * w_{12}(n). \tag{10}$$

From (7) and (8), the post filter distorts the output signal, and if the two sensors are closely spaced, the effect of this post filter is critical [18]. To escape the effect of this post filter, we take the situation where two sensors are loosely spaced which has a minimal distortion because of this post filter [13].

4 Proposed Algorithm

In this work, we propose a new approach for speech enhancement and noise reduction by combining the FBSS structure with reduce size simplified FTF algorithm. The R-SMFTF is a stable version of the FTF algorithms, and has been proposed in [10] and [11]. The combination of FBSS structure with reduce size simplified FTF affords to restore two source signals from two mixtures observations. The scheme of the proposed TM-RSMFTF algorithm is shown in Fig. 3.

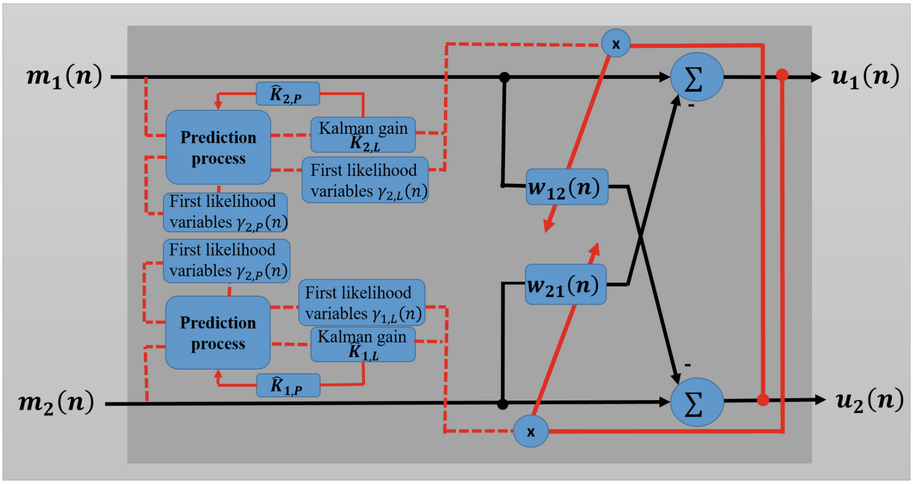

Fig. 3. The proposed TM-RSMFTF algorithm scheme

The two outputs of the TM-RSMFTF algorithm $u_1(n)$ and $u_2(n)$ are given as:

$$u_1(n) = m_1(n) - \boldsymbol{w}_{21}^{T}(n)\boldsymbol{M}_2(n), \tag{11}$$

$$u_2(n) = m_2(n) - \boldsymbol{w}_{12}^{T}(n)\boldsymbol{M}_1(n). \tag{12}$$

where, $\boldsymbol{M}_1(n) = [m_1(n), m_1(n-1), \ldots, m_1(n-L+1)]$ and $\boldsymbol{M}_2(n) = [m_2(n), m_2(n-1), \ldots, m_2(n-L+1)]$ are two vectors represent respectively the 2 mixture inputs $m_1(\mathrm{n})$ and $m_2(\mathrm{n})$. The updated equations of the two cross filters tap-weights w_{21} and w_{12} are controlled by the TM-RSMFTF algorithm and they are given as follows, respectively:

$$\boldsymbol{w}_{21}(n) = \boldsymbol{w}_{21}(n-1) + \boldsymbol{K}_1(n)u_1(n), \tag{13}$$

$$\boldsymbol{w}_{12}(n) = \boldsymbol{w}_{12}(n-1) + \boldsymbol{K}_2(n)u_2(n). \tag{14}$$

The two vectors $\boldsymbol{K_1}(n)$ and $\boldsymbol{K_2}(n)$ represent the adaptation gain which are given by:

$$\boldsymbol{K_1}(n) = \gamma_{1,L}(n)\widehat{\boldsymbol{K_1}}(n), \tag{15}$$

$$\boldsymbol{K_2}(n) = \gamma_{2,L}(n)\widehat{\boldsymbol{K_2}}(n). \tag{16}$$

where, the variables $\gamma_{1,L}(n)$ and $\gamma_{2,L}(n)$ which represent the first likelihood variables are used to update the two cross filters tap-weights that will give bellow and $\widehat{\boldsymbol{K_1}}(n)$, $\widehat{\boldsymbol{K_2}}(n)$ are normalized Kalman gain vectors. From [11, 19], the normalized Kalman gain vectors can be calculated by using only the forward predictor $\boldsymbol{a}(n)$, and they are given by:

$$\begin{bmatrix} \widehat{\boldsymbol{K_1}}(n) \\ c_1(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \widehat{\boldsymbol{K_1}}(n-1) \end{bmatrix} + \varphi_1(n) \begin{bmatrix} 1 \\ -\boldsymbol{a_{1,P}} \\ 0_{L-P} \end{bmatrix}, \tag{17}$$

$$\begin{bmatrix} \widehat{\boldsymbol{K_2}}(n) \\ c_2(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \widehat{\boldsymbol{K_2}}(n-1) \end{bmatrix} + \varphi_2(n) \begin{bmatrix} 1 \\ -\boldsymbol{a_{2,P}} \\ 0_{L-P} \end{bmatrix}. \tag{18}$$

where, $c_1(n)$ and $c_2(n)$ are the last unused components of the normalized Kalman gain vectors, and the variables $\varphi_1(n)$ and $\varphi_2(n)$ are given respectively by:

$$\varphi_1(n) = \frac{e_1(n)}{\lambda\alpha_1(\mathrm{n}) + c_a}, \tag{19}$$

$$\varphi_2(n) = \frac{e_2(n)}{\lambda\alpha_2(\mathrm{n}) + c_a}. \tag{20}$$

where, $0 > \lambda > 1$ is forgetting factor and c_a is small regularization constant use to avoid performing numerical divisions by very small values in the absence of the input signal (silence period). The parameters α_1 and α_2 are the forward prediction errors variance, and they are given by:

$$\alpha_1(n) = \lambda\alpha_1(n-1) + \gamma_{1,P}e_1^2(n), \tag{21}$$

$$\alpha_2(n) = \lambda\alpha_2(n-1) + \gamma_{2,P}e_2^2(n). \tag{22}$$

where, $\gamma_{1,P}$ and $\gamma_{2,P}$ are second likelihood variables that will also give bellow, and they are used to update the forward prediction errors variance. The prediction error $e_1(n)$ and $e_2(n)$ can be be calculated by using the same algorithm and are given by:

$$e_1(n) = m_2(n) - \boldsymbol{a_{1,P}}(n)\boldsymbol{M_{2,P}}(n-1), \tag{23}$$

$$e_2(n) = m_1(n) - \boldsymbol{a_{2,P}}(n)\boldsymbol{M_{1,P}}(n-1). \tag{24}$$

where, $\boldsymbol{M_{1,P}}(n)$ and $\boldsymbol{M_{2,P}}(n)$ are two vectors that represent the P last simples of the two mixture signals $m_1(n)$ and $m_2(n)$, respectively. $\boldsymbol{a_{1,P}}(n)$ and $\boldsymbol{a_{2,P}}(n)$ are the forward predictors of order P that are calculated by minimizing the criteria $E\left[e_1^2(n)\right]$ and $E\left[e_2^2(n)\right]$,

$$\boldsymbol{a_{1,P}}(n) = \eta(\boldsymbol{a_{1,P}}(n-1) - \gamma_{1,L}(n)\widehat{\boldsymbol{K}}_{\boldsymbol{1,P}}(n-1)e_1(n), \tag{25}$$

$$\boldsymbol{a_{2,P}}(n) = \eta(\boldsymbol{a_{2,P}}(n-1) - \gamma_{2,L}(n)\widehat{\boldsymbol{K}}_{\boldsymbol{2,P}}(n-1)e_2(n). \tag{26}$$

Where $\widehat{\boldsymbol{K}}_{\boldsymbol{1,P}}(n)$ and $\widehat{\boldsymbol{K}}_{\boldsymbol{2,P}}(n)$ are the first P elements of $\widehat{\boldsymbol{K_1}}(n)$ and $\widehat{\boldsymbol{K_2}}(n)$ respectively, and the constant η, which is close to one, is called the leakage factor that allows the forward predictors to return back to zero [10].By saving $\varphi_1(n), \varphi_2(n), e_1(n)$ and $e_2(n)$ in vectors of length $(L+1)$, we can obtain: $\boldsymbol{\varphi_1(n)} = [\varphi_1(n), \varphi_1(n-1), \ldots, \varphi_1(n-L)]^{\boldsymbol{T}}$, $\boldsymbol{\varphi_2(n)} = [\varphi_2(n), \varphi_2(n-1), \ldots, \varphi_2(n-L)]^{\boldsymbol{T}}$, $\boldsymbol{E_1(n)} = [e_1(n), e_1(n-1), \ldots, e_1(n-L)]^{\boldsymbol{T}}$ and $\boldsymbol{E_2(n)} = [e_2(n), e_2(n-1), \ldots, e_2(n-L)]^{\boldsymbol{T}}$.

The first likelihood variables $\gamma_{1,L}(n)$ and $\gamma_{2,L}(n)$ of the Eqs. (15), (16), (25) and (26) are given by [11]:

$$\gamma_{1,L}(n) = \frac{1}{1-\psi_1(n)}, \text{ at each period of } N \text{ simples we do: } \gamma_{1,L}(n) = \eta\,\gamma_{1,L}(n), \tag{27}$$

$$\gamma_{2,L}(n) = \frac{1}{1-\psi_2(n)}, \text{ at each period of } N \text{ simples we do: } \gamma_{L,2}(n) = \eta\,\gamma_{2,L}(n). \tag{28}$$

where, $\psi_1(n)$ and $\psi_2(n)$ are given by:

$$\psi_1(n) = \psi_1(n-1) - \varphi_1(n)e_1(n) + \varphi_1(n-L)e_1(n-L), \tag{29}$$

$$\psi_2(n) = \psi_2(n-1) - \varphi_2(n)e_2(n) + \varphi_2(n-L)e_2(n-L). \tag{30}$$

where, $\varphi_i(n)$ and $\varphi_i(n-L)$ are the first and $(L+1)^{th}$ components of vector $\boldsymbol{\varphi_i(n)}$, respectively. The elements $e_i(n)$ and $e_i(n-L)$ are the first and $(L+1)^{th}$ components of vector $\boldsymbol{E_i(n)}$, respectively $i \in \{1,2\}$.

The second likelihood variables $\gamma_{1,P}(n)$ and $\gamma_{2,P}(n)$ of the Eqs. (21) and (22) are given by [11]:

$$\gamma_{1,P}(n) = \frac{1}{1-\phi_1(n)}, \text{ at each period of } N \text{ simples we do: } \gamma_{1,P}(n) = \eta\,\gamma_{1,P}(n), \tag{31}$$

$$\gamma_{2,P}(n) = \frac{1}{1-\phi_2(n)}, \text{ at each period of } N \text{ simples we do: } \gamma_{2,P}(n) = \eta\,\gamma_{2,P}(n). \tag{32}$$

Where $\phi_1(n)$ and $\phi_2(n)$ are given by:

$$\phi_1(n) = \phi_1(n-1) - \varphi_1(n)e_1(n) + \varphi_1(n-L+P)e_1(n-L+P), \tag{33}$$

$$\phi_2(n) = \phi_2(n-1) - \varphi_2(n)e_2(n) + \varphi_2(n-L+P)e_2(n-L+P). \tag{34}$$

Where $\varphi_i(n-L+P)$ and $e_i(n-L+P)$ are the $(L-P)^{th}$ components of vectors $\boldsymbol{\varphi_i}(n)$ and $\boldsymbol{E_i}(n)$, respectively.$i \in \{1,2\}$. We recall here that the forgetting factor λ is calculating by [10]:

$$\frac{1+\sqrt{1+(P+2)\left(\frac{1}{\eta^2}-1\right)}}{P+2} \geq \lambda > 1. \tag{35}$$

5 Computational Complexity

In this Section, we quantify the computational complexity of the proposed TM-RSMFTF algorithm in comparison with TM-SFTF algorithm and TM-FNLMS algorithm. To highlight the computational complexity of each algorithm, we have calculated the number of multiplications and additions per iteration. We have reported in Table 1 the computational complexity of the proposed TM-RSMFTF algorithm for various size predictor (P = 1, 40 and 100), TM-SFTF algorithm, TM-FNLMS algorithm. The length of the real adaptive filter is L = 128 and 512.

Table 1. Computational complexity of the proposed algorithm, TM-SFTF, and TM-FNLMS algorithm.

Algorithms		Computational complexity		L = 128	L = 512
		Multiplications per iteration	Additions per iteration		
TM-RSMFTF	P = 1	$6L + 6P + 22$	$6L + 4P + 18$	1586	6194
	P = 40			1976	6584
	P = 100			2576	7184
TM-SFTF		$12L + 18$	$10L + 14$	2848	11296
TM-FNLMS		$6L + 4$	$6L + 4$	1544	6152

From Table 1, we can observe that the proposed algorithm is less complex than the full-size version (TM-SFTF) by about $4L - 4P$ additions and $6L - 6P$ multiplications and is more complex than the TM-FNLMS algorithm by about $4P$ additions and $6P$ multiplications and it has the same computational complexity of the TM-FNLMS algorithm if P = 1 and lower than the TM-SFTF algorithm for any size $P < L$.

6 Simulation Results

In this Section, we present the simulation results of the proposed algorithm described previously, many experiments were performed to evaluate the performance and to show the behavior of the proposed algorithm TM-RSMFTF in comparison with the TM-FNLMS and TM-SFTF algorithm. To evaluate the performance of the proposed algorithm, the following objective criteria are used: Time evolution of the output speech signal $u_1(n)$ with original speech signal, the system mismatch (SM) and segmental mean square error (SegMSE) to evaluate convergence speed and we also used the segmental signal to noise ratio (segSNR) to evaluate the noise reduction performance and Cepstral distance (CD) to evaluate the distortion caused by these techniques in the output speech signal.

6.1 Description of the Used Signals and the Experimental

In this sub-Section, the speech signal that we have used is selected from AURORA database and given in Fig. 4, and the sources noise is stationary noise USASI (United States of America Standard Institute now ANSI) which is correlated signal noise. All these signals are sampled at sampling frequency $F_s = 8$ kHz and coded on 16 bits and given in Fig. 4. For the spaced microphones configuration, we have constructed the impulse responses $h_{21}(n)$ and $h_{12}(n)$ which are given in Fig. 5 according to the model presented in [16]. We note here that for generating the two mixture signals $m_1(n)$ and $m_2(n)$, we have used (1) and (2), and they are shown in Fig. 6. The real filters length is $L = 128$, and the input SNR at the first input $m_1(n)$ is equal to 0 dB and at the second input $m_2(n)$ is equal to −3 dB. Finally, Table 2 summarizes the controlling parameters for each algorithm.

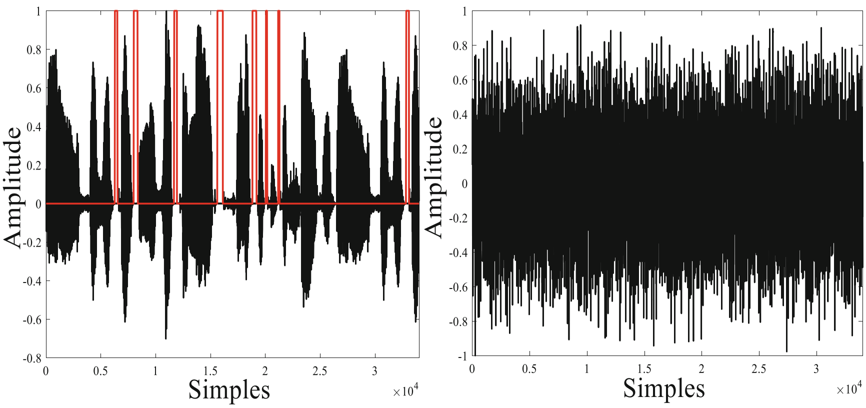

Fig. 4. Original speech signal (in left) with its segmentation (in red) and USASI noise (in right) (Color figure online)

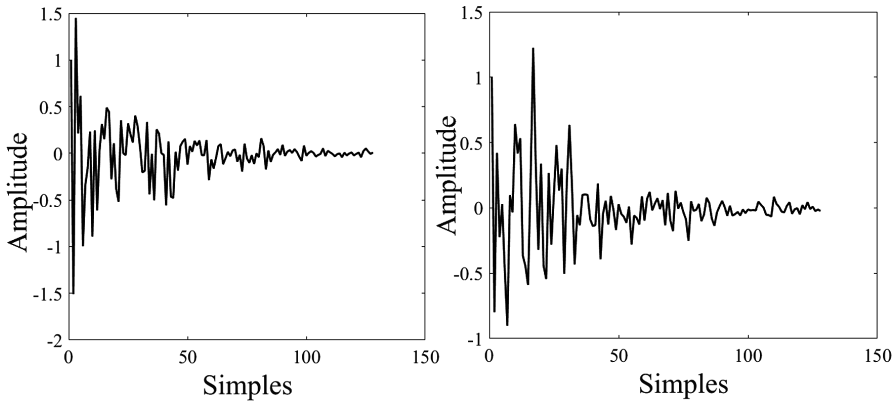

Fig. 5. Examples of simulated impulse responses h_{21}, h_{12}

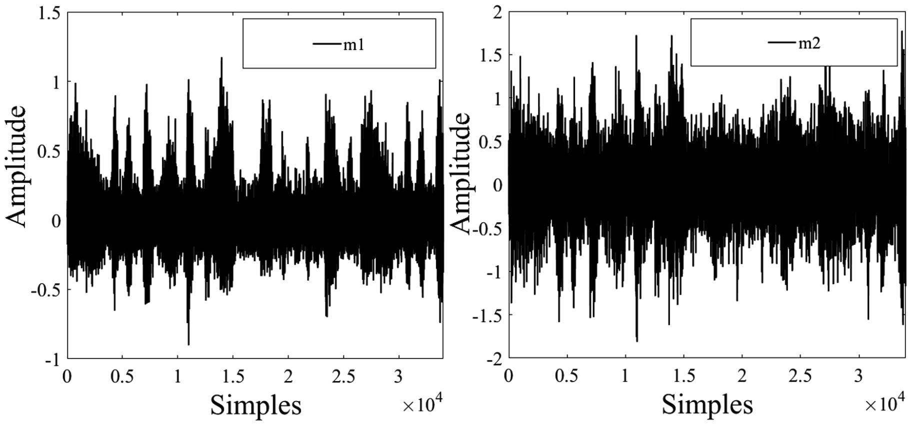

Fig. 6. The mixture signal $m_1(n)$ (in left) and the mixture signal $m_2(n)$ (in right).

Table 2. Controlling parameters for the proposed algorithm, TM-FNLMS and TM-SFTF

TM-FNLMS	TM-SFTF	Proposed TM-RSMFTF algorithm		
$\mu_{21} = 0.7$	$\eta = 0.995$	$P = 1$	$P = 40$	$P = 100$
$\mu_{12} = 0.7$		$\eta = 0.995$	$\eta = 0.999$	$\eta = 0.991$
$C = 0.000001$	$\lambda = 0.5$	$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.5$
	$c_a = 1$	$c_a = 0.1$	$c_a = 0.5$	$c_a = 1$

6.2 Time Evolution of the Output Speech Signal

In order to evaluate the performance of the proposed algorithm TM-RSMFTF, we have given in Fig. 7, the estimated speech signals $u_1(n)$ that are obtained by the proposed

TM-RSMFTF algorithm with different size of predictors (P = 1, 40 and 100) and the results obtained from TM-SFTF algorithm.

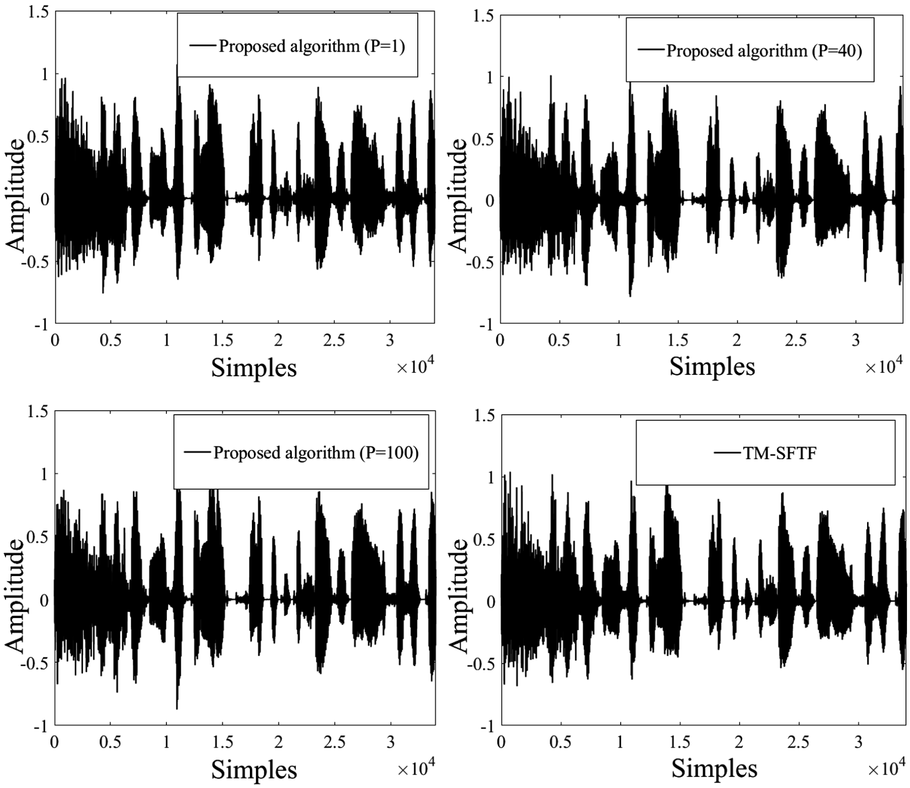

Fig. 7. Estimated speech signals obtained by the proposed TM-RSMFTF algorithm (P = 1, 40 and 100), and The TM-SFTF algorithm.

Based on Fig. 7, it can be observed that the acoustic noise almost entirely reduced from the output speech signals by the proposed algorithm TM-RSMFTF with various size predictor and TM-SFTF algorithm, and from these results we can conclude that the proposed algorithm has good performance to reduce acoustic noise and enhance speech but we cannot judge whom of the 4 algorithms is the best in term of the quality of the estimated output speech and convergence speed that's why we have used different objective criteria such as the system mismatch(SM), mean square error(MSE), segmental signal to noise ratio(segSNR) and Cepstral distance(CD).

6.3 System Mismatch (SM)

In order to evaluate the convergence speed and tracking ability of the proposed algorithm, we used system mismatch criteria (SM). This objective criterion will facilitate quantifying the convergence and tracking ability of any methods. As we are interested in noise reduction and speech enhancement in our TM-FBSS structure so our interest is directed to the estimated speech output $u_1(n)$; and therefore, we concentrate only on the adaptive filter $w_{21}(n)$. The SM of $w_{21}(n)$ is computed by the following equation [11]:

$$SM_{dB} = 20log_{10}\left(\frac{\|h_{21} - w_{21}(n)\|^2}{\|h_{21}\|^2}\right). \tag{36}$$

where, $w_{21}(n)$ and h_{21} are simulated and real impulse responses. We have carried out many experiments of the proposed algorithm TM-RSMFTF with different sizes predictors P, TM-SFTF and conventional TM-NLMS algorithm. The source noise is USASI noise and input SNR at the first input is equal to 0 dB and at the second input is equal to −3 dB.

We have given in Fig. 8 the SM criterion evaluation obtained by the proposed algorithm TM-RSMFTF for different size predictor (P = 1, 40 and 100), TM-SFTF algorithm and TM-FNLMS algorithm. According to Fig. 8, we noticed that for different size predictor, the proposed algorithm have almost the same performance with the full-size version(TM-SFTF), and this result is obtained by the good choice of the controlling parameters (λ, η, c_a) for each algorithm. We can conclude that the good convergence speed of the proposed algorithm allows us to say that this algorithm is a good choice for noise reduction and speech enhancement applications.

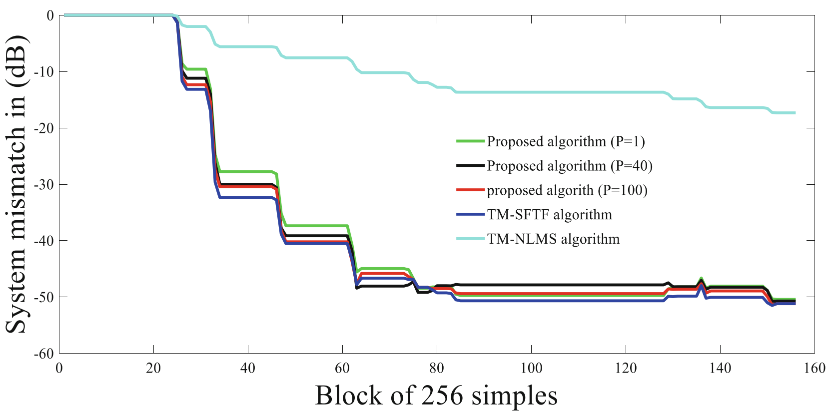

Fig. 8. SM comparison between proposed TM-RSMFTF for size predictor P = 1(in green), P = 40(in black), P = 100(in red), TM-SFTF (in blue) and TM-FNLMS algorithm (in cyan). (Color figure online)

6.4 Signal to Noise Ratio (SNR)

In order to quantify the noise reduction performance of the proposed TM-FRSMFTF algorithm, we have calculated the output SNR of the estimated speech output. The SegSNR criterion is given by the following expression:

$$SNR_{dB} = 10log_{10}\left(\frac{\sum_{k=1}^{Q-1}|s(k)|^2}{\sum_{k=1}^{Q-1}|s(k) - u_1(k)|^2} VAD_K\right). \tag{37}$$

where, $s(n)$ and $u_1(n)$ are respectively the original speech signal and the enhanced signal. The parameter Q is the mean averaging value of the output SNR. The presence of the term *VAD* (voice activity detector) in (37) means that this objective criterion calculated only during speech signal periods. We have reported in Fig. 9 that the results of SegSNR obtained by the proposed algorithm with different size predictor (P = 1, 40 and 100), TM-SFTF algorithm and the TM-FNLMS algorithm.

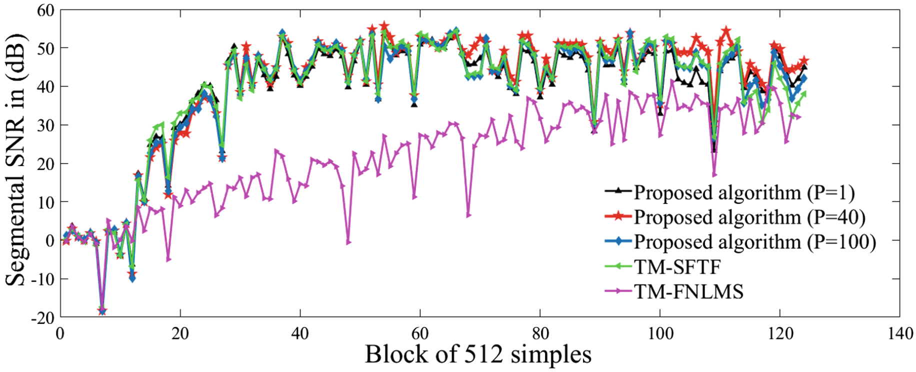

Fig. 9. SegSNR comparison between proposed TM-RSMFTF for size predictor P = 1(in black), P = 40 (in red), P = 100(in blue), TM-SFTF (in green) and TM-FNLMS (in magenta). (Color figure online)

From Fig. 9, we observed that the proposed TM-RSMFTF algorithm with different size predictor P has almost the same behavior in reducing noise in comparison with the TM-SFTF algorithm, and even for $P \ll L$, no degradation has been found in the output SegSNR values. From the same figure, we can also observe that the proposed TM-RSMFTF with predictor order one has good behavior in comparison with the TM-FNLMS algorithm and almost identical to TM-SFTF algorithm. From these results, we can conclude that the proposed TM-RSMFTF algorithm has good behavior in reducing noise.

6.5 Cepstral Distance (CD)

The CD criterion allows evaluating the distortion of the TM-FNLMS algorithm, TM-FRLS, and the proposed TM-FRSMFTF algorithm. This objective criterion is estimated by using the log-spectrum distance between the original speech signal $sp(n)$ and $u_1(n)$, where $u_1(n)$ is the estimated speech signal obtained at the output of the FBSS structure and is given by the following equation [11]:

$$CD_{dB} = \sum\nolimits_{\mathrm{k}=1}^{Q-1} IFFT[((log_{10}(|S(k, \omega)| - log_{10}(|U_1(k, \omega)|)VAD_K)]^2 \quad (38)$$

where, $S(\omega)$ and $U_1(\omega)$ are respectively the short Fourier transform of the original speech signal $s(n)$ and the enhanced one $u_1(n)$ at each frame k, and Q is the mean averaging value of the CD criterion. The presence of the parameter VAD (voice activity detector) in (38) means that we calculate this criterion only during the presence of speech signal periods, and in these periods, VAD takes 1 and when the speech signal is absent VAD takes 0.

We have done our simulation with the same speech signal of Fig. 4 and source noise is USASI with the same controlling parameters in Table 2 and the input SNR at the first input is 0 dB and at the second input is −3 dB. From Fig. 10, we have seen the superiority in less distortion of the proposed algorithm in comparison with the TM-FNLMS algorithm.

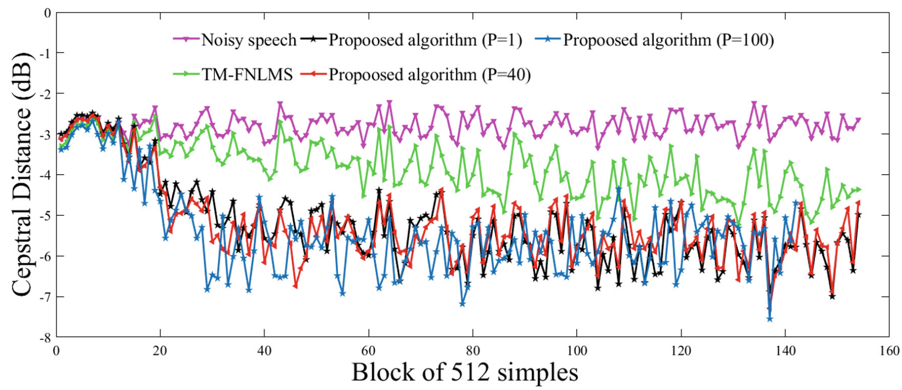

Fig. 10. Cepstral distance evaluation of the proposed algorithm (P = 1 in black), (P = 40 in red), (P = 100 in blue) TM-FNLMS algorithm (in green) and noisy speech (in magenta). (Color figure online)

6.6 Mean Square Error

In order to quantify the convergence of the algorithms i.e. the proposed TM-RSMFTF, TM-FNLMS, and TM-SFTF, we use another objective criterion which calls SegMSE. This SegMSE criterion is given by the following equation:

$$MSE(db) = 20log_{10}\left(\sum_{k=1}^{Q-1} u_1(k)VAD_K\right). \quad (39)$$

where, $u_1(n)$ is the enhanced speech signal, and the parameter Q is the mean averaging value of SegMSE. The VAD is voice activity detector used to estimate the SegMSE in absence speech periods.

From Fig. 11, we have confirmed the good speed convergence of the proposed TM-RSMFTF algorithm with different size of predictor (P = 1, 40 and 100). We have also confirmed that no retrogression has been seen in SegMSE values in the steady-state regime even for small size predictor. The good choice of the controlling parameters (λ, η, c_a) for each algorithm lead to the same convergence speed performance with the TM-SFTF algorithm. Finally, we confirmed that the good convergence speed of the proposed TM-RSMFTF algorithm allows us to say that this algorithm is a good choice for noise reduction and speech enhancement applications.

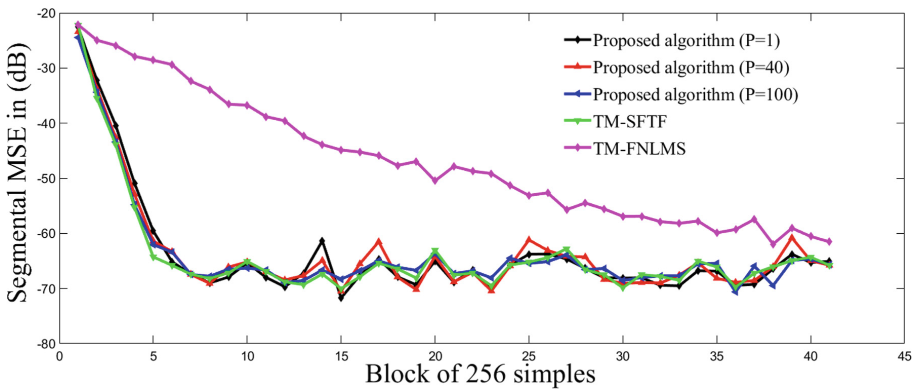

Fig. 11. SegMSE comparison between proposed TM-RSMFTF for size predictor P = 1(in black), P = 40(in red), P = 100(in blue), TM-SFTF algorithm (in green) and TM-FNLMS (in magenta). (Color figure online)

7 Conclusion

In this paper, we have proposed a new robust algorithm applied for speech quality improvement and noise reduction in new telecommunication systems and speech recognition systems. The proposed algorithm is an outcome of the combination of two microphones forward blind source separation structure and the reduce size simplified fast transversal filter which is a stable and robust version of the FTF algorithms.

The proposed TM-RSMFTF algorithm has low computational complexity, which is less than TM-SFTF algorithm, and similar to the TM-NLMS algorithm if P = 1. Reducing the size predictor P leads to the reduction of the computation complexity for this algorithm without degradation in its performance even for small size predictor $(P \ll L)$ due to the presence of regularization constant and leakage factor and also to the two likelihood variables.

The simulation results confirm this conclusion and also show that the proposed algorithm has almost similar performance with the TM-SFTF algorithm and better than the TM-FNLMS algorithm in term of less distortion on the enhanced speech signal, speed convergence and different objective criteria such as segmental signal to noise ratio (SegSNR), Segmental mean square error (SegMSE) and the system mismatch (SM).

Acknowledgment. This work was supported in part by the National Basis Research Program of China 973 Program under Grant 2013CB329003 and in part by the National Natural Science Foundation of China under Grants 61671179.

References

1. Boll, S.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Sig. Process. **27**(2), 113–120 (1979)
2. Lim, J.S., Oppenheim, A.V.: Enhancement and bandwidth compression of noisy speech. Proc. IEEE **67**(12), 1586–1604 (1979)
3. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Sig. Process. **33**(2), 443–445 (1985)
4. Zou, X., Zhang, X.: Speech enhancement using an MMSE short time DCT coefficients estimator with supergaussian speech modeling. J. Electron. (China) **24**(3), 332–337 (2007)
5. Haykin, S.: Adaptive Filter Theory, 4th edn. Prentice-Hall, Englewood Cliffs (2002)
6. Ljung, L., Morf, M., Falconer, D.: Fast calculation of gain matrices for recursive estimation schemes. Int. J. Control **27**(1), 1–19 (1978)
7. Cioffi, J., Kailath, T.: Fast, recursive-least-squares transversal filters for adaptive filtering. IEEE Trans. Acoust. Speech Sig. Process. **32**(2), 304–337 (1984)
8. Benallal, A., Gilloire, A.: A new method to stabilize fast RLS algorithms based on a first-order of the propagation of numerical errors. In: International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1988, pp. 1373–1376. IEEE (1988)
9. Slock, D.T.M., Kailath, T.: Numerically stable fast transversal filters for recursive least-squares adaptive filtering. In: Golub, G.H., Van Dooren, P. (eds.) Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, vol. 70, pp. 605–615. Springer, Heidelberg (1991). https://doi.org/10.1007/978-3-642-75536-1_49
10. Benallal, A., Benkrid, A.: A simplified FTF-type algorithm for adaptive filtering. Sig. Process. **87**(5), 904–917 (2007)
11. Djendi, M.: New efficient adaptive fast transversal filtering (FTF)-type algorithms for mono and stereophonic acoustic echo cancelation. Int. J. Adapt. Control Sig. Process. **29**(3), 273–301 (2015)
12. Rahima, H., Djebari, M., Mohamed, D.: Blind speech enhancement and acoustic noise reduction by SFTF adaptive algorithm. In: 2017 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B), pp. 1–4. IEEE (2017)
13. Djendi, M.: A new two-microphone Gauss-Seidel pseudo affine projection algorithm for speech quality enhancement. Int. J. Adapt. Control Sig. Process. **31**(8), 1162–1183 (2017)
14. Van Gerven, S., Van Compernolle, D.: Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness. IEEE Trans. Sig. Process. **43**(7), 1602–1612 (1995)

15. Thi, H.L.N., Jutten, C.: Blind source separation for convolutive mixtures. Sig. Process. **45** (2), 209–229 (1995)
16. Djendi, M., Gilloire, A., Scalart, P.: Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms. In: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 3, p. III. IEEE (2006)
17. Weinstein, E., Feder, M., Oppenheim, A.V.: Multi-channel signal separation by decorrelation. IEEE Trans. Speech Audio Process. **1**(4), 405–413 (1993)
18. Djendi, M., Scalart, P., Gilloire, A.: Analysis of two-sensors forward BSS structure with post-filters in the presence of coherent and incoherent noise. Speech Commun. **55**(10), 975–987 (2013)
19. Arezki, M., Benallal, A., Meyrueis, P., Berkani, D.: A new algorithm with low complexity for adaptive filtering. Eng. Lett. **18**(3), 205 (2010)