A novel cascade structure for joint backward blind acoustic noise and echo cancellation systems

Rahima Henni^{1,2} · Mohamed Djendi¹

Received: 2 November 2019 / Accepted: 21 May 2020 / Published online: 2 June 2020
© Springer Nature Switzerland AG 2020

Abstract

This paper addresses joint acoustic echo and punctual noise cancellations framework in hands-free communication systems. A novel cascade structure that combines the backward blind source separation structure (BBSS) as an acoustic noise canceller (ANC), and a conventional acoustic echo canceller (AEC), intended to solve the problem of acoustic echo, is presented. It is proposed to use two adaptive filtering algorithms to update the filter coefficients of each process (ANC and AEC stages). In the first ANC stage, the BBSS structure is combined with the two-channel simplified fast transversal (SFTF) filter algorithm. In the second AEC stage, a single-channel SFTF algorithm is used. Experimental results show both superiority and efficiency of the new structure in comparison with the AEC based SFTF system (without a joint process) in terms of echo return loss enhancement and mean square error gain criteria.

Keywords AEC · ANC · Backward BSS · Single-channel SFTF · Two-channel SFTF

1 Introduction

Echo and noise related problems are very common in many applications involving speech communication, e.g., audio conference and hands-free telephony. In these systems, the speech that comes from the far-end speaker and echoes back with a time delay produces perception problems. In such condition, the perception is further impaired when the speaker is situated in a noisy environment [1]. In order to provide a better communication service, an acoustic echo canceller (AEC) is required to cancel the echo returned to the transmission room and also to allow uninterrupted communication between the rooms [2–4].

A class of promising state of the art techniques exists in the literature in which many one microphone [5, 6], and two microphones [7, 8], adaptive techniques have been proposed to resolve this critical issue, i.e. acoustic echo. However, the presence of the punctual noise impedes the

convergence of acoustic echo cancellation algorithms, which leads to poor echo cancellation. Wherefore, for hands-free mobile equipment, it is strongly required to employ an acoustic noise canceller besides the acoustic echo canceller. A survey of techniques that disjointedly reduce the noise are proposed, for example, in previous studies [9, 10], some adaptive techniques based on one [11, 12], and or two [13, 14] microphones have been proposed to correct speech distortion. In [15–17], the source separation structures principle (forward and backward) were used to improve the corrupted speech signals [18, 19]. Combinations between these two structures and adaptive filtering algorithm families have given new insight to acoustic noise cancellation field. In the situation where two acoustical disturbances are present in one time, i.e. acoustic noise and echo, and to improve the quality of the communication, the aforementioned noise reduction techniques should be concatenated by a second process,

✉ Mohamed Djendi, m_djendi@yahoo.fr; m_djendi@univ-Blida.dz; Rahima Henni, henni93rahima@gmail.com | ¹Laboratory of Signal Processing and Imaging (LATSI), University of Blida 1, Route de Soumaa, B.P. 270, 09000 Blida, Algeria. ²Detection, Information and Communication (DIC) Laboratory, University of Blida 1, Route de Soumaa, B.P. 270, 09000 Blida, Algeria.

i.e. acoustic echo cancellation. Therefore, a survey of techniques for combining acoustic noise and echo cancellation systems can be found in literature [20–24]. The order in which the processing blocks are applied is very important, in the case where the AEC process is firstly applied and then the ANC, and since most noise reduction techniques make use of multiple microphones, the acoustic echo canceller then obviously has to be repeated for each of the microphones, and these AEC systems need to be robust against the noise that is still present in their input signals. The advantage of the second case where the ANC precedes the AEC is that only one acoustic echo canceller is needed.

In this paper, we propose a novel cascade structure for joint backward blind acoustic noise and echo cancellation systems. In the new structure, we propose to use the BBSS structure with the two-channel SFTF algorithm [25] to suppress the punctual noise components from the far-end signal before to be processed with the AEC system based on the use of the single-channel SFTF algorithm [26–30].

The outline of this paper is as follows: Sect. 2 presents the development of the proposed approach, in which a detailed description of the acoustical environment and the new cascading structure for combining acoustic noise and echo cancellation systems are given. In Sect. 3, we present the experimental study, where we describe the used signals and we present the simulation and the evaluation results of the proposed approach. Finally, we conclude our work in Sect. 4.

2 Development of the proposed approach

In realistic hands-free communication applications, the observed speech signal is corrupted by the echo and noise components as well. As shown in Fig. 1, in addition to the punctual noise, the two microphones of the near-end pick up the amplified and broadcasted far-end signal.

In order improve the AEC in the presence of punctual noise components, an acoustic noise canceller should be integrated to deal with noise components. In this section, we present a novel cascade structure for joint backward blind acoustic noise and echo cancellation systems, where the blind ANC stage is placed before the AEC one. In Fig. 2, we give a detailed scheme of the proposed structure, and then each step is described in the next sub-sections.

2.1 Modeling of the acoustical environment

In Fig. 2 [1st step], we give the mixing model that is physically coherent [17], where $s(n)$ is the far-end signal, $b(n)$ is the punctual noise, and the impulse response $h_{11}(n)$ models the direct acoustic path, the impulse responses

Fig. 1 Illustrative scheme of the acoustical environment

$h_{12}(n)$ and $h_{21}(n)$ model the cross acoustic paths between the source signals and the two microphones. In our work, it is assumed that the noise source is close to the second microphone, hence, the direct impulse response $h_{22}(n)$ is the Kronecker unit impulse $\delta(n)$ [13]. Both of $d_1(n)$ and $d_2(n)$ are the echo signals picked out by the near-end microphones. These echo signals are given by relations (1) and (2) respectively.

$$d_1(n) = s(n) * h_{11}(n) \quad (1)$$

$$d_2(n) = s(n) * h_{12}(n) \quad (2)$$

However, the observed and available signals are $p_1(n)$ and $p_2(n)$ and are given in a function notation as follow:

$$p_1(n) = d_1(n) + b(n) * h_{21}(n) \quad (3)$$

$$p_2(n) = d_2(n) + b(n) \quad (4)$$

where $*$ denotes the linear convolution operator.

Fig. 2 Detailed scheme of the novel cascading structure for joint backward blind acoustic noise and echo cancellation systems

2.2 Problem formulation of joint backward blind ANC and AEC cancellation systems

The proposed system is depicted in Fig. 2 [2nd and 3rd steps]. It consists of a backward blind ANC process, for dealing with the noise cancellation task, and an AEC process, which is designed for the echo cancellation task. The backward blind ANC process [See Fig. 2 (2nd step)] comprises a two channel noise cancellation system based on the BBSS structure [25]. This blind approach aims to recover the original sources estimates of $s(n)$ and $b(n)$ by

using only the noisy observations $p_1(n)$ and $p_2(n)$ that are generated by the model of Fig. 2 [1st step].

The outputs of the ANC process are given by the following relations:

$$u_1(n) = p_1(n) - u_2(n) * w_{21}(n) \quad (5)$$

$$u_2(n) = p_2(n) - u_1(n) * w_{12}(n) \quad (6)$$

By inserting (3) and (6) in (5), and (4) and (5) in (6), we get:

$$u_1(n) [\delta(n) - w_{12}(n) * w_{21}(n)] = [s(n) * (h_{11}(n) - h_{12}(n) * w_{21}(n)) + b(n) * (h_{21}(n) - w_{21}(n))] \quad (7)$$

$$u_2(n) [\delta(n) - w_{12}(n) * w_{21}(n)] = [s(n) * (h_{12}(n) - h_{11}(n) * w_{12}(n)) + b(n) * (\delta(n) - h_{21}(n) * w_{12}(n))] \quad (8)$$

by using the optimal solutions:

$$w_{21}(n) = h_{21}(n) \quad (9)$$

$$w_{12}(n) = h_{12}(n) * \frac{1}{h_{11}(n)} \quad (10)$$

It can recover the original signals, i.e. $s(n)$ and $b(n)$, as given in (11) and (12), with a distortion as given in (11).

$$u_1(n) = s(n) * h_{11}(n) \quad (11)$$

$$u_2(n) = b(n) \quad (12)$$

As depicted in Fig. 2 [3rd step], and since the echo signal is speech we employ the AEC system only at the first output, i.e. $u_1(n)$, we thus obtain the following output signal relation:

$$u_3(n) = u_1(n) - s(n) * w_{13}(n) \quad (13)$$

For more development, we insert (11) in (13). Then the AEC output signal formula can be written as:

$$u_3(n) = s(n) * (h_{11}(n) - w_{13}(n)) \quad (14)$$

To suppress the acoustic echo, the filter $w_{13}(n)$ must identify the impulse response $h_{11}(n)$, i.e. $w_{13}(n) = h_{11}(n)$.

2.3 Algorithms outline

In this analysis, we consider two algorithms to update the filters coefficients of the proposed system: the first one is two-channel SFTF algorithm proposed in our previous work [25], it is used in the blind ANC block, and the single-channel SFTF algorithm [26] which is used in the AEC block.

The output signals $u_1(n)$ and $u_2(n)$ of the blind ANC block are given in a vector notation as follow:

$$u_1(n) = p_1(n) - \mathbf{w}_{21}^T(n) \mathbf{u}_2(n) \quad (15)$$

$$u_2(n) = p_2(n) - \mathbf{w}_{12}^T(n) \mathbf{u}_1(n) \quad (16)$$

where $\mathbf{u}_1(n) = [u_1(n), \dots, u_1(n-L+1)]^T$, and $\mathbf{u}_2(n) = [u_2(n), \dots, u_2(n-L+1)]^T$ are the two error vectors of the BBSS structure of Fig. 2 [2nd step]. $\mathbf{w}_{21}(n) = [w_{21}(n), \dots, w_{21}(n-L+1)]^T$ and $\mathbf{w}_{12}(n) = [w_{12}(n), \dots, w_{12}(n-L+1)]^T$ are the two filters updated by two-channel SFTF algorithm and given as:

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) - u_1(n) \gamma_1(n) \mathbf{k}_1(n) \quad (17)$$

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) - u_2(n) \gamma_2(n) \mathbf{k}_2(n) \quad (18)$$

The two-channel SFTF algorithm is obtained by eliminating the backward prediction process, thus only the forward predictor is used to compute the dual Kalman gain vectors, i.e. $\mathbf{k}_1(n)$ and $\mathbf{k}_2(n)$, these two vectors can be written as:

$$\begin{bmatrix} \mathbf{k}_1(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}_1(n-1) \end{bmatrix} - \frac{\varepsilon_1(n)}{\lambda \alpha_1(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_1(n-1) \end{bmatrix} \quad (19)$$

$$\begin{bmatrix} \mathbf{k}_2(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}_2(n-1) \end{bmatrix} - \frac{\varepsilon_2(n)}{\lambda \alpha_2(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_2(n-1) \end{bmatrix} \quad (20)$$

where α_1 and α_2 are the forward prediction errors variances parameters presented by the following relations:

$$\alpha_1(n) = \lambda \alpha_1(n-1) + \gamma_1(n-1) \varepsilon_1^2(n) \quad (21)$$

$$\alpha_2(n) = \lambda \alpha_2(n-1) + \gamma_2(n-1) \varepsilon_2^2(n) \quad (22)$$

We get the forward prediction coefficients vectors, i.e. $\mathbf{a}_1(n)$ and $\mathbf{a}_2(n)$, by minimizing the functions $E[u_1(n)]$, and $E[u_2(n)]$, respectively. The update formulas of the forward predictors $\mathbf{a}_1(n)$ and $\mathbf{a}_2(n)$ are given by:

$$\mathbf{a}_1(n) = \rho [\mathbf{a}_1(n-1) - \varepsilon_1(n) \gamma_1(n) \mathbf{k}_1(n-1)] \quad (23)$$

$$\mathbf{a}_2(n) = \rho [\mathbf{a}_2(n-1) - \varepsilon_2(n) \gamma_2(n) \mathbf{k}_2(n-1)] \quad (24)$$

The prediction errors $\varepsilon_1(n)$ and $\varepsilon_2(n)$ can be estimated as follows:

$$\varepsilon_1(n) = u_2(n) - \mathbf{a}_1(n) \mathbf{u}_2(n-1) \quad (25)$$

$$\varepsilon_2(n) = u_1(n) - \mathbf{a}_2(n) \mathbf{u}_1(n-1) \quad (26)$$

And we apply the following definitions to calculate the likelihood variables $\gamma_1(n)$, $\gamma_2(n)$:

$$\gamma_1(n) = \frac{1}{1 - \mathbf{k}_1^T(n) \mathbf{u}_2(n)} \quad (27)$$

$$\gamma_2(n) = \frac{1}{1 - \mathbf{k}_2^T(n) \mathbf{u}_1(n)} \quad (28)$$

Now, we describe the mathematical derivation of the single-channel SFTF algorithm that we use it in the AEC stage. The a priori filtering error $u_3(n)$ and the adaptive filter $w_{13}(n)$ equations of this algorithm are given by:

$$u_3(n) = u_1(n) - \mathbf{w}_{13}^T(n) \mathbf{s}(n) \quad (29)$$

$$\mathbf{w}_{13}(n+1) = \mathbf{w}_{13}(n) - u_3(n) \gamma_3(n) \mathbf{k}_3(n) \quad (30)$$

where $s(n) = [s(n), s(n-1), \dots, s(n-L+1)]^T$ is the coefficients vector of the far-end signal $k_3(n)$ and $\gamma_3(n)$ are the Kalman gain vector and the likelihood variable, respectively, and they are calculated as follow:

$$\begin{bmatrix} k_3(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ k_3(n-1) \end{bmatrix} - \frac{\varepsilon_3(n)}{\lambda\alpha_3(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_3(n-1) \end{bmatrix} \quad (31)$$

$$\gamma_3(n) = \frac{1}{1 - \mathbf{k}_3^T(n)\mathbf{s}(n)} \quad (32)$$

The forward prediction coefficients vector $\mathbf{a}_3(n)$ is calculated by minimizing the cost function $E[\varepsilon_3^2(n)]$. The update formula of the forward predictor is given by the following relation:

$$\mathbf{a}_3(n) = \rho[\mathbf{a}_3(n-1) - \varepsilon_3(n)\gamma_3(n)\mathbf{k}_3(n-1)] \quad (33)$$

The forward prediction error variance $\alpha_3(n)$ is given by:

$$\alpha_3(n) = \lambda\alpha_3(n-1) + \gamma_3(n-1)\varepsilon_3^2(n) \quad (34)$$

and the prediction error $\varepsilon_3^2(n)$, is calculated by the following relation:

$$\varepsilon_1(n) = s(n) - \mathbf{a}_1(n)\mathbf{s}(n-1) \quad (35)$$

The asterisk "*" represents the last unused element of the dual Kalman gain vector. λ is a smoothing factor, δ is a small positive constant used to avoid division by very small values in silence periods. The parameter ρ allows better robustness against numerical propagation errors.

In Tables 1 and 2, the proposed cascade structure is summarized.

Table 1 1st proposed processing algorithm: two-channel SFTF algorithm

Initialization	$\alpha_1(0) = \alpha_2(0) = 0, \mathbf{p}_1(0) = [0]^T, \mathbf{p}_2(0) = [0]^T, \mathbf{a}_1(0) = [0]^T, \mathbf{a}_2(0) = [0]^T, \mathbf{k}_1(0) = [0]^T, \mathbf{k}_2(0) = [0]^T,$ $\mathbf{w}_{21}(0) = [0]^T, \mathbf{w}_{12}(0) = [0]^T, \mathbf{u}_1(0) = [0]^T, \mathbf{u}_2(0) = [0]^T$
Prediction part	<p>Kalman gains</p> $\begin{bmatrix} \mathbf{k}_1(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}_1(n-1) \end{bmatrix} - \frac{\varepsilon_1(n)}{\lambda\alpha_1(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_1(n-1) \end{bmatrix}$ $\begin{bmatrix} \mathbf{k}_2(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}_2(n-1) \end{bmatrix} - \frac{\varepsilon_2(n)}{\lambda\alpha_2(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_2(n-1) \end{bmatrix}$ <p>Prediction errors</p> $\varepsilon_1(n) = u_2(n) - \mathbf{a}_1(n)\mathbf{u}_2(n-1)$ $\varepsilon_2(n) = u_1(n) - \mathbf{a}_1(n)\mathbf{u}_1(n-1)$ <p>Forward prediction coefficients</p> $\mathbf{a}_1(n) = \rho[\mathbf{a}_1(n-1) - \varepsilon_1(n)\gamma_1(n)\mathbf{k}_1(n-1)]$ $\mathbf{a}_2(n) = \rho[\mathbf{a}_2(n-1) - \varepsilon_2(n)\gamma_2(n)\mathbf{k}_2(n-1)]$ <p>Forward prediction errors variances</p> $\alpha_1(n) = \lambda\alpha_1(n-1) + \gamma_1(n-1)\varepsilon_1^2(n)$ $\alpha_2(n) = \lambda\alpha_2(n-1) + \gamma_2(n-1)\varepsilon_2^2(n)$ <p>Likelihood variables</p> $\gamma_1(n) = \frac{1}{1 - \mathbf{k}_1^T(n)\mathbf{u}_2(n)}$ $\gamma_2(n) = \frac{1}{1 - \mathbf{k}_2^T(n)\mathbf{u}_1(n)}$
Filtering part	<p>A priori filtering errors</p> $u_1(n) = p_1(n) - \mathbf{w}_{21}^T(n)\mathbf{u}_2(n)$ $u_2(n) = p_2(n) - \mathbf{w}_{12}^T(n)\mathbf{u}_1(n)$ <p>Adaptive filters updates</p> $\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) - u_1(n)\gamma_1(n)\mathbf{k}_1(n)$ $\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) - u_2(n)\gamma_2(n)\mathbf{k}_2(n)$

Table 2 2nd proposed processing algorithm: single-channel SFTF algorithm

Initialization	$\alpha_3(0) = 0, \mathbf{s}(0) = [0]^T, \mathbf{a}_3(0) = [0]^T, \mathbf{k}_3(0) = [0]^T, \mathbf{w}_{13}(0) = [0]^T, \mathbf{u}_3(0) = [0]^T$
Prediction part	$\text{Kalman gain} \begin{bmatrix} \mathbf{k}_3(n) \\ * \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}_3(n-1) \end{bmatrix} - \frac{\varepsilon_3(n)}{\lambda \alpha_3(n-1) + \delta} \begin{bmatrix} 1 \\ -\mathbf{a}_3(n-1) \end{bmatrix}$ $\text{Prediction error } \varepsilon_1(n) = s(n) - \mathbf{a}_1(n)\mathbf{s}(n-1)$ $\text{Forward prediction coefficient } \mathbf{a}_3(n) = \rho[\mathbf{a}_3(n-1) - \varepsilon_3(n)\gamma_3(n)\mathbf{k}_3(n-1)]$ $\text{Forward prediction error variance } \alpha_3(n) = \lambda \alpha_3(n-1) + \gamma_3(n-1)\varepsilon_3^2(n)$ $\text{Likelihood variable } \gamma_3(n) = \frac{1}{1 - \mathbf{k}_3^T(n)\mathbf{s}(n)}$
Filtering part	$\text{A priori filtering error } u_3(n) = u_1(n) - \mathbf{w}_{13}^T(n)\mathbf{s}(n)$ $\text{Adaptive filter update } \mathbf{w}_{13}(n+1) = \mathbf{w}_{13}(n) - u_3(n)\gamma_3(n)\mathbf{k}_3(n)$

Fig. 3 Time evolution of the signals used in simulation. **a** The far-end signal $s(n)$. **b** The punctual noise signal $b(n)$. The microphones signals $p_1(n)$ and $p_2(n)$ are presented respectively in **(c)** and **(d)**

3 Experimental results

3.1 Description of the used signals

The observed signals $p_1(n)$ and $p_2(n)$, are generated by the model of Fig. 2 [1st step], these signals contains two statistically independent source signals: (1) the speech signal $s(n)$, is a short French sentence of about 4 s, pronounced by a male speaker and is phonetically balanced, and (2) a punctual noise $b(n)$, that is USASI noise [United states of America Standard Institute now (ANSI)]. The impulse responses (IR) $h_{12}(n)$ and $h_{21}(n)$ are generated by the model of [9] and have a length $L = 256$ [Samples of

Fig. 4 Simulated impulse responses with $L = 256$. **a**, **b** and **c** represent respectively the impulse responses $h_{11}(n)$, $h_{12}(n)$ and $h_{21}(n)$ **Fig. 5** Time evolution of: [In top]: the output signal obtained by the AEC based SFTF system. [In bottom]: The output signal obtained by the proposed cascading structure

Fig. 6 ERLE evaluation of the proposed approach and the AEC based SFTF system, with USASI noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

Fig. 7 ERLE evaluation of the proposed approach and the AEC based SFTF system, with white noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

Fig. 8 ERLE evaluation of the proposed approach and the AEC based SFTF system, with babble noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

these IRs are given by Fig. 4]. These IRs are used to generate the observations, i.e. $p_1(n)$ and $p_2(n)$, a samples of all these signals are given by Fig. 3. The sampling frequency is set to $f_s = 8$ kHz and the input SNR is chosen equal to $SNR = 15$ dB. All experimental signals and IRs are presented below on Figs. 3 and 4, respectively.

The black stepped curve represented with the original speech signal depicted on Fig. 3a represents the manual

voice activity detector (Manual VAD). We use this system to extract the speech signal at the first output $u_1(n)$ of the ANC process. It is recalled here that all the obtained results are mean averaged.

In order to evaluate the backward blind ANC process on the AEC stage, we compare in Fig. 5 the output signal obtained by the proposed structure and that obtained by the conventional AEC system. The conventional AEC

system considered in our comparison is based on the use of the adaptive single-channel SFTF algorithm. According to Fig. 5, we see the good behavior of the proposed structure in comparison with the conventional AEC system, where both disturbances have been canceled.

3.2 Objective criteria

In order to compare the proposed algorithm with the AEC based SFTF system, several experiments in different conditions were performed. We consider the echo return loss enhancement (ERLE) criterion given by [25]:

$$ERLE_{dB} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left(\frac{\sum_{n=Nm}^{Nm+N-1} |s(n)|^2}{\sum_{n=Nm}^{Nm+N-1} |s(n) - u_3(n)|^2} \right) \quad (36)$$

and the mean square error (MSE) gain, computed as follows [25]:

$$MSE_{dB} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left(\frac{1}{N} \sum_{n=Nm}^{Nm+N-1} |s(n) - u_3(n)|^2 \right) \quad (37)$$

where $s(n)$ and $u_3(n)$ are respectively, the original speech signal and the output signal obtained by the proposed

Table 3 Summary of the simulation parameters

AEC based SFTF system

Processing algorithm: single-channel SFTF algorithm

$L = 256, \lambda = 0.999, \rho = 0.96, \gamma_1(0) = 1, \delta = 0.00001$, initialization constant $E_0 = 1$

Proposed approach

ANC processing algorithm: two-channel SFTF algorithm

$L = 256, \lambda = 0.988, \rho = 0.6, \gamma_1(0), \gamma_2(0) = 1, \delta = 0.00001$, initialization constant $E_0 = 0.7$

AEC processing algorithm: Single-channel SFTF algorithm

$L = 768, \lambda = 0.999, \rho = 0.96, \gamma_3(0) = 1, \delta = 0.00001$, initialization constant $E_0 = 1$

Fig. 9 MSE gain evaluation of the proposed approach and the AEC based SFTF system, with USASI noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

Fig. 10 MSE gain evaluation of the proposed approach and the AEC based SFTF system, with white noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

Fig. 11 MSE gain evaluation of the proposed approach and the AEC based SFTF system, with babble noise, and for three input SNRs: [left]: 5 dB, [middle]: 15 dB and [right]: 25 dB

Table 4 Parameter values of the following algorithms: (1) AEC based NLMS [1], (2) joint system [31], and [1] the proposed approach [in this paper]

	Parameters
Input signals	Speech signal: a male speaking in French, Noises: USASI noise, Sampling frequency rate: $f_s = 8$ kHz
Mixing signals	Real filters length: 256
AEC NLMS [2]	Adaptive filter length w_1 : 256, Fixed step-sizes:
Joint AEC ANC system [31]	Adaptive filter length of w_1, w_2, w_3 : 256, Fixed step-sizes: $\mu = 0.9$
Proposed approach [In this paper]	Adaptive filter length of w_1, w_2, w_3 : 256, Fixed step-sizes: $\mu = 0.9$. Exponential forgetting factor: $\lambda = 0.9982$. Forgetting factor: $\lambda_a = 0.9988$. Positive constant: $c_0 = 1, c_a = 0.0035$ Initialization constant: $E_0 = 3$

structure. The parameters M and N are the number of segments and the segment length, respectively.

We have evaluated the performances of the proposed joint backward blind acoustic noise and echo cancellation systems in comparison with the conventional AEC system based on the single-channel SFTF algorithm and we have reported the obtained results on Figs. 6, 7 and 8. The simulation parameters of the proposed algorithms are summarized in Table 3 (selected by simulation), and the obtained results are given for different conditions test and noisy observations levels (high and low input SNR). In these experiments, three input SNRs, i.e. 5 dB, 20 dB and 25 dB are used. Furthermore, we have used three types of noise, i.e. USASI, white, and babble. The white noise is used to test the stability performance of the used algorithms, for the evaluation of the convergence speed performance, we use the USASI noise and a real babble noise is used to evaluate the ability of the processing algorithms to track the non-stationarity of the input signal.

Fig. 12 ERLE comparison of the proposed approach with the AEC based NLMS [2], and the joint system [31], USASI noise, and for input SNR: 15 dB

Fig. 13 MSE gain comparison of the proposed approach with: the AEC based NLMS [2], and the joint system [31], USASI noise, and for input SNR: 15 dB

According to Figs. 6, 7 and 8, we see that the proposed system gives better results in comparison with the AEC based SFTF system in terms of ERLE. This is due to the noise reduction improvement provided by the integrated backward blind ANC process that allows a better estimate of the acoustic echo. We can see that the proposed algorithm performs well even in high punctual noise (input $SNR = 5$ dB), unlike the classical AEC system, where the algorithm is disturbed by the punctual noise present in the input signal.

To support the previous results, we have performed several other experiments in terms of MSE criterion and have selected one of them to evaluate the AEC performance of the proposed cascade structure in comparison with AEC based SFTF system. The simulation parameters are summarized in Table 3, and the obtained results for three input SNRs, i.e. 5 dB, 15 dB and 25 dB, and three noise types, i.e. USASI, white and babble, are reported on Figs. 9, 10 and 11.

From the obtained results (that are mean averaged), it can be noted the superiority of the proposed approach (cascade structure) over the AEC based SFTF system in transient and steady-state regime. This superiority is more attractive when the input SNR is selected low, this is due to the backward blind ANC stage that makes the AEC process of the proposed cascade structure more robust against acoustic echo components, whereas, the AEC based SFTF system is highly penalized by the noise components presence. Finally, it can be concluded that the proposed cascade structure allows better cancellation of acoustic echo and noise components in the same time.

3.2.1 A comparative study between the proposed approach, AEC NLMS [2], and joint AEC and ANC system [31]

In this section, the performance of the proposed approach is compared with the following algorithms: (1) AEC based NLMS [2], and (2) a joint AEC and ANC system that uses both two-channel NLMS and single-channel NLMS in their process [31]. In this comparative study, the AEC NLMS algorithm [2] is taken as a reference algorithm. In this experiment, the real and adaptive filters are equal to $L = 256$. The impulse responses is of a room and composed about 256 points, this means an exact modelization of the room impulse response by the adaptive filters of the same length (equal to 256). The parameters of each algorithm are summarized in Table 4.

Parameters in Table 4 are selected to get the best performance of each algorithm. It is evaluated the ERLE and MSE criteria for different input SNRs and different noise type. However, we have selected some results of ERLE and MSE and reported in Figs. 12 and 13, respectively. From these results, it can be conclude that the proposed approach behaves more efficiently than the AEC based NLMS and [2] and the joint AEC and ANC system of [31]. The superiority of our approach is got from the good behavior of the one and two-channel SFTF algorithms that are integrated in the proposed joint AEC and ANC structure.

4 Conclusion

In this paper, we have proposed a novel algorithm for joint backward blind acoustic noise and echo cancellation systems. The proposed algorithm is based on two cascade stages for cancelling punctual noise and then acoustic echo signal. In the ANC stage, the BBSS structure is used to cancell punctual noise, then an AEC system is used to suppress the echo signal. Both stages use efficient SFTF algorithm to take advantage of the adaptive behaviour of this algorithms when combined with the proposed cascade system.

To evaluate the proposed system, we have done intensive tests in terms of ERLE and MSE criteria, in various conditions of noisy observations (highly and slightly noisy observations). In these experiments, our proposed approach is compared with an AEC based SFTF system. The obtained results have shown the superiority of the proposed algorithm in terms of objective criteria even under low input SNR condition (more punctual noise is present). Finally, it can be concluded that the proposed system can be a good alternative for AEC techniques in the presence of strong punctual noise components where the classical techniques fail.

Acknowledgements This study was carried out without funding. Authors are grateful to the Blida University, Algeria, and Professor Abderreak Guessoum, Professor at Blida University, Algeria, and Director of Signal Processing and imaging Laboratory (LATSI) for providing the infrastructure to achieve this work.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that they have no conflict of interest.

References

- Puder H, Dreiseitel P (2000) Implementation of a hands-free car phone with echo cancellation and noise-dependent loss control. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, vol 6, pp 3622–3625
- Widrow B, Stearns SD (1985) Adaptive signal processing. Prentice Hall, Upper Saddle River
- Sondhi M, Kellermann W (1992) Adaptive echo cancellation for speech signal. In: Furui S, Sondhi M (eds) Chapter 11 in Advances in speech signal processing. Dekker, New York
- Hänsler E, Schmidt GU (2004) Acoustic echo and noise control: a practical approach. Wiley, New York
- Sakai Y, Tahir Akhtar M (2013) The performance of the acoustic echo cancellation using blind source separation to reduce double-talk interference. In: International symposium on intelligent signal processing and communication systems, Naha, Japan
- Mader A, Puder H, Schmidt GU (2000) Step-size control for acoustic echo cancellation filters—an overview. Signal Process 80:1697–1719
- Paleologu C, Benesty J, Ciocină S (2014) Widely linear general Kalman filter for stereophonic acoustic echo cancellation. Signal Process 94:570–575
- Benesty J, Gänsler T, Morgan DR, Sondhi MM, Gay SL (2013) Advances in network and acoustic echo cancellation. Springer, Berlin
- 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 IEEE ICASSP, vol 3, pp 744–747
- Gupta D, Gupta VK, Mishra AN (2017) Article: Pseudo affine projection algorithm based noise minimization from speech signals. In: IJCA proceedings on national conference on recent trends in electronics and electrical engineering NCRTEEE, no 1, pp 1–5
- Gabrea M (2003) Double affine projection algorithm-based speech enhancement algorithm. In: Proceedings of the IEEE ICASSP Montréal, Canada, vol 2, pp 904–907
- Al Kindi MJ, Dunlop J (1989) Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel. Signal Process 17(3):241–250
- Van Gerven S, Van Compernelle D (1995) Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness. IEEE Trans Signal Process 43(3):1602–1612
- Gupta VK, Gupta DK, Chandra M (2014) Real time noise canceller using modified sigmoid function RLS algorithm. In: Proceedings of international conference on computational vision and robotics (ICCVIR-2014), 9th–10th, August, 2014 Bhubaneswar, India
- Davis GM (2002) Noise reduction in speech applications. CRC Press, Boca Raton. ISBN 0-8493-0949-2
- Buchner H, Aichner R, Kellermann W (2005) A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics. IEEE Trans Speech Audio Process 13(1):120–134
- Djendi M, Zoulikha M (2014) New automatic forward and backward blind sources separation algorithms for noise reduction and speech enhancement. Comput Electr Eng 40:2072–2088
- Rupp M (2011) Pseudo affine projection algorithms robustness and stability analysis. IEEE Trans Signal Process 59:2017–2023
- Gonzalez A, Ferrer M, Albu F, de Diego M (2012) Affine projection algorithm: evolution to smart and fast algorithms and application. In: 20th European signal processing conference, EURASIP, 2012. ISSN 2076-1465
- Jeannès RLB, Faucon G (1997) How to reduce the noise influence in a joint system developed for echo and noise cancellation. IEEE Signal Process Lett 4(10):280–282
- Kuo SM, Gan WS, Asthana P (2005) Integrated noise reduction and acoustic echo cancellation in hands-free systems. In: Proceedings of 2005 international symposium on intelligent signal processing and communication systems, Hong Kong
- Hanshi SM, Chong YW, Ramadass S, Naeem AN, Ooi KC (2014) Efficient acoustic echo cancellation joint with noise reduction framework. In: International conference on computer, communication, and control technology (I4CT 2014), September 2–4, 2014, Langkawi, Kedah, Malaysia
- Park SJ, Cho CG, Lee C, Youn DH (2002) Integrated echo and noise canceller for hands-free applications. IEEE Trans Circuits Syst II 49(3):186–195
- Jeannes J, Scalart P, Faucon G, Beaugent C (2001) Combined noise and echo reduction in handsfree systems: a survey. IEEE Trans Speech Audio Process 9(8):808–820
- Henni R, Djendi M, Djebbari M (2019) A new efficient two-channel fast transversal adaptive filtering algorithm for blind speech enhancement and acoustic noise reduction. Comput Electr Eng 73:349–368
- Djendi M (2015) New efficient adaptive fast transversal filtering (FTF)-type algorithms for mono and stereophonic acoustic echo cancellation. Int J Adapt Control Signal Process 29:273–301
- Sundaradhas SN, Panchama moorthy SP, Ramapackiyam SSK (2020) Upgraded NLMS algorithm for speech enhancement with sparse and dispersive impulse responses. Indian J Phys. <https://doi.org/10.1007/s12648-020-01688-5>
- Li H, Feng J, Wang Y, Zhang X (2019) A joint optimized robust acoustic echo cancellation algorithm. Int J Pattern Recognit Artif Intell 33(09):1959029
- Sayoud A, Djendi M, Guessoum A (2019) A new speech enhancement adaptive algorithm based on fullband–subband MSE switching. Int J Speech Technol 22:993–1005
- Hassani I, Arezki M, Benallal A (2020) A novel set membership fast NLMS algorithm for acoustic echo cancellation. Appl Acoust 163:107210
- Henni R, Djendi M, Djebbari M (2019) A new adaptive solution based on joint acoustic noise and echo cancellation for hands-free systems. Int J Speech Technol 22:407–420

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.