Journal of Information Systems and Telecommunication
Vol.12, No.2, April-June 2024, 105-116
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

An Acoustic Echo Canceller using Moving Window to Track Energy Variations of Double-Talk-Detector

Mouldi Makdir[1*], Mohamed Bouamar[1,2], Mourad Benziane[2]

[1] Department of Electronics, Faculty of Technology, University of M'sila, M'sila, 28000, Algeria
[2] Laboratory of Analysis of Signals and Systems, M'sila, 28000, Algeria

Received: 21 Nov 2022/ Revised: 07 Oct 2023/ Accepted: 06 Dec 2023

Abstract

As a fundamental device in acoustic echo cancellation (AEC) systems, the echo canceller based on adaptive filters relies on the adaptive approximation of the echo-path. However, the adaptive filter must face the risk of divergence during the double-talk periods when the near-end is present. To solve this problem, the double-talk-detector (DTD) is often used to detect the double-talk periods and prevent the echo canceller from being disturbed by the other end of the speaker's signal. In this paper, we propose a DTD based on a new method that can detect quickly and track accurately double-talk periods. It is based on the sum of energies of the estimated echo and the microphone signals which is continuously compared to the error energy. A window that moves with time and tracks energy variations of the different input signals of the DTD represents a fundamental feature of the proposed method compared to several other methods based on correlation. The goal is to outperform conventional normalized cross-correlation (NCC) methods which are well-known in terms of small steady-state misalignment and stability of decision variable. In this work, the normalized least mean squares (NLMS) algorithm is used to update the filter coefficients along speech signals which are taken from the NOIZEUS database. Efficiency of the proposed method is particularly compared to the conventional Geigel algorithm and normalized cross-correlation method (NCC) that depends on the cross-correlation between the microphone signal and the error signal of AEC. Performance evaluation is confirmed by computer simulation.

Keywords: AEC; DTD; NLMS; NCC; Moving Window.

1- Introduction

The technique of acoustic echo cancellation (AEC) known for its interest in various applications of signal processing plays an important role in the field of telecommunications. The use of "hands-free" terminals allows maintaining the speaker's freedom of movement and ensuring the comfort of conversations. When the acoustic echo is present in a troublesome way, specific treatment must imperatively be implemented to preserve the quality of communication. The object of such a treatment is to estimate the acoustic echo between the received signal (signal sent in the loudspeaker) and the output of the room (signal picked up by the microphone) then to subtract an estimate of this output's signal without affecting the local speech signal in the case of double-talk (DT) [1,2]. This processing is done by using adaptive filtering where a double-talk-detector (DTD) is used to sense when the echo signal is corrupted by near-end signal. The role of this main function is to freeze adaptation of the filter coefficients when the near-end speech is present in order to avoid divergence of the adaptive algorithm [3-5].

Other methods based on combined adaptive filtering without DTD retain the advantages of both fast convergence rate and small steady-state misalignment but suffer from the same problems encountered in this field, such as abrupt changes in the acoustic echo-path, surrounding noise, and tracking capability. Indeed, they are complex and consume more computing time [6, 7].

In this work, we propose an efficient DTD to solve the problem provoked by the acoustic echo with the capability to improve speech intelligibility during telephone calls. To do this, a simulation will be started to allow a comparative study with two other methods [8-11].

The paper is organized as follows, in Section 2, the acoustic echo canceller with the proposed DTD is presented. In Sections 3 and 4, the used methods and the proposed one are formulated. The computational complexity is illustrated in Section 5. Simulation results are discussed in Section 6. Finally, the conclusion is given in Section 7.

✉ **Mouldi Makdir**
mouldi.makdir@univ-msila.dz

2- Acoustic Echo Cancellation

The acoustic echo canceller is used to remove the echo created due to the loudspeaker-microphone environment. We present in Fig.1 the structure of the device based on the new DT-detection method. In this case, the proposed DTD is controlled with three input signals where the energies of the estimated echo $\hat{y}(n)$ and the microphone signal *d(n)* are continuously compared to the error energy of the signal *e(n)*.

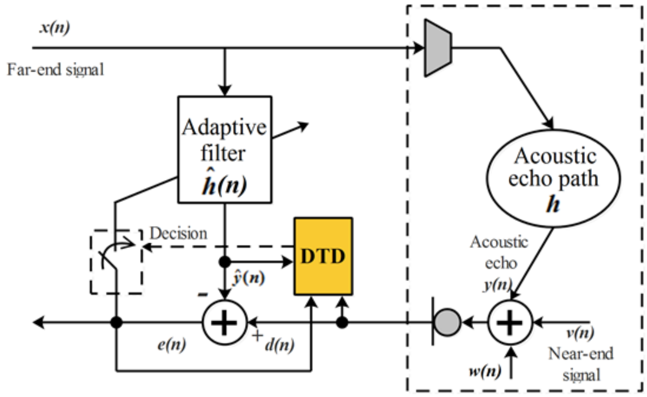


Fig. 1 Acoustic echo canceller with the proposed DTD

Note that the far-end vector signal $\mathbf{x}(n)$is filtered by the impulse response $\mathbf{h}$ modeling the room. At time n, the resulting signal (echo $y(n)$) is added to the near-end signal $v(n)$ and background noise *w(n)* to give the corrupted microphone signal *d(n)*.
We have:

$$d(n) = y(n) + v(n) + w(n) \tag{1}$$

$\mathbf{x}(n)$is filtered through the impulse response $\mathbf{h}$ to get the echo signal:

$$y(n) = \mathbf{h}^T \mathbf{x}(n) \tag{2}$$

where:

$$\mathbf{x}(n) = [x(n) \quad x(n-1) \quad \dots \quad x(n-\mathrm{L}+1)]^T$$
$$\mathbf{h} = [h_0 h_1 \quad \dots \quad h_{\mathrm{L}-1}]^T$$

We have assumed that the length L of the vector signal $\mathbf{x}(n)$ is the same as the effective length of the echo-path $\mathbf{h}$.
At time n, the estimated echo $\hat{y}(n)$ is created by the convolution of the coefficients vector of adaptive filter $\hat{\mathbf{h}}(n)$ with the received input vector signal $\mathbf{x}(n)$.

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n-1)\, \mathbf{x}(n) \tag{3}$$

Where $\hat{\mathbf{h}}(n) = [\hat{h}_0(n) \quad \hat{h}_1(n) \quad \dots \quad \hat{h}_{\mathrm{L}-1}(n)]^T$

The estimated echo signal is subtracted from the microphone signal, and the error signal is therefore given by:

$$e(n) = d(n) - \hat{y}(n) \tag{4}$$

Error signal which represents the error of the impulse response estimation is used in the adaptive algorithm to adapt the L coefficients of the filter $\hat{\mathbf{h}}$.
Several algorithms have been used to update the adaptive filter coefficients to converge to the optimal solution such as least mean squares (LMS), normalized least mean squares (NLMS), recursive least squares (RLS), and affine projection algorithms [4,12-15]. As an adaptive filtering algorithm that allows updating the filter coefficients, we use NLMS [16] to validate the proposed method. This is one, of the most used adaptive filtering algorithms which is defined by:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{\beta}{c + \mathbf{x}^T(n)\mathbf{x}(n)} e(n)\mathbf{x}(n) \tag{5}$$

Where $\hat{\mathbf{h}}(n+1)$ it is the next tap weight value, and $\hat{\mathbf{h}}(n)$ the current tap weight value of the adaptive filter. A constant β $(0 < \beta < 2)$ controlling convergence is considered as a stabilization factor and a step size parameter used in updating the weight vector. The regularization parameter is a constant $c > 0$ that prevents division by zero [3, 12].
When the near-end signal is not present with any adaptive algorithm, the filter $\hat{\mathbf{h}}$ will quickly converge to an estimate of the echo-path $\mathbf{h}$ and this is the best way to cancel the echo. When the far-end signal is not present, or very small, the adaptation is stopped by the nature of the adaptive algorithm. When both signals are present, the near-end signal could disrupt the adaptation of filter $\hat{\mathbf{h}}$ and cause divergence. An effective DT-detection algorithm is used to stop adaptation of filter $\hat{\mathbf{h}}$ as fast as possible when the level of the near-end signal becomes appreciable in relation to the level of the far-end signal and to keep the adaptation going when the level of near-end signal is negligible. This is the case where it is important to use efficient DTD.

3- Double-Talk-Detection

DT-detection is used with an echo canceller to sense when echo signal is corrupted by near-end signal. Its role is to freeze the adaptation of the filter $\hat{\mathbf{h}}$ when near-end signal is present in order to avoid divergence of the adaptive algorithm. Typically, the DT-detection algorithm calculates a decision variable $\xi(n)$ and the DT is declared when $\xi(n)$ it is lower or upper than a threshold level T[10,11,17].
Methods based on DT-detection can be classified into two major categories, namely signal energy based and signal correlation based. Several methods such as cross-correlation (CC) [18-22], coherence, voice activity

detection, and fundamental frequency estimation have been proposed in the literature [23-26]. Methods based on cross-correlation between the far-end and error signals are then proposed. Moreover, approximate versions, such as a normalized cross-correlation (NCC), are developed but with a different combination of DTD input signals. Therefore, we will discuss in this study two of the most prominent methods in order to demonstrate their underlying ideas.

3-1- Geigel Method

A simple but elegant DT-detection algorithm was proposed by Geigel which is widely used for its easy implementation [8]. It is usually limited to network echo application where the echo level is typically 6 dB below that of far-end signal. It performs an amplitude level comparison between the maximum of a length L_G observation of $\mathbf{x}(n)$ and the microphone signal $d(n)$, where the decision variable is defined as :

$$\xi_G(n) = \frac{\max\left\{|x(n)|,|x(n-1)|,....,|x(n-L_G+1)|\right\}}{|d(n)|} \tag{6}$$

L_G it is a constant that determines the number of past samples of the far-end signal that are used for the DT-detection. Decision is made by comparing $\xi_G(n)$ with a suitable threshold level T_G [19].

3-2- Cross-Correlation Method

The first method based on the cross-correlation between the far-end signal and the error signal is proposed by Hua Ye and Bo-Xiu Wu [9]. Some approximate versions as NCC are appeared in different articles where each method differs from the others in the DTD input signals [10,11,27]. Among these, we find one that depends on the cross-correlation between the microphone signal $d(n)$and the error signal $e(n)$ which we will use in this paper with the mentioned Geigel algorithm to compare them with the proposed method. Note that the performance of the proposed method in [11] is exactly similar to the best-known cross-correlation based DTD [10].
A statistical decision ξ_{NCC} of the NCC method is given by [11]:

$$\xi_{NCC}(n) = 1 - \frac{\hat{r}_{ed}(n)}{\hat{\sigma}_d^2(n)} \tag{7}$$

Where r_{ed} is the cross-correlation between $e(n)$ and $d(n)$, and σ_d^2 the variance of $d(n)$.
$\xi_{NCC}(n)$, it is based on estimates $\hat{r}_{ed}(n)$ and $\hat{\sigma}_d{}^2(n)$ which are found by using exponential weighting recursive estimation form [28, 29]:

$$\hat{r}_{ed}(n) = \lambda \hat{r}_{ed}(n-1) + (1-\lambda) e(n) d(n) \tag{8}$$

$$\hat{\sigma}_d^2(n) = \lambda \hat{\sigma}_d^2(n-1) + (1-\lambda) d(n) d(n) \tag{9}$$

Where λ is the exponential weighting factor ($0.9 < \lambda < 1$).
It should be noted that this method based on recursive estimation has a remarkable performance. However, it is significantly simpler and computationally very efficient. In addition, its main advantage is that only the maximum value of cross-correlation needs to be computed instead of computing the entire cross-correlation vector required by the other algorithms [11].
In this work, we propose to compare particularly the NCC method with the proposed one which is based on a moving temporal window used to track energy variations of three vector signals: error vector signal $\mathbf{e}(n)$, microphone vector signal $\mathbf{d}(n)$ and estimated vector signal $\hat{\mathbf{y}}(n)$.

4- Proposed Method

A fundamental feature of the proposed method compared to other ones is based on a window that moves with time to track energy variations of each input signal of the DTD. Three input signals to control the DTD are used where the sum of energies of the estimated echo and the microphone signals is continuously compared to the error signal energy.

We get the three input vector signals of the proposed DTD at time n as:

$$\mathbf{e}(n) = [e(n) \quad e(n\text{-}1) \quad \dots \quad e(n\text{-}N\text{+}1)]^T \tag{10}$$

$$\mathbf{d}(n) = [d(n) \quad d(n\text{-}1) \quad \dots \quad d(n\text{-}N\text{+}1)]^T \tag{11}$$

$$\hat{\mathbf{y}}(n) = [\hat{y}(n) \quad \hat{y}(n\text{-}1) \quad ... \quad \hat{y}(n\text{-}N+1)]^T \tag{12}$$

Where N it is a constant length of the temporal window chosen to compute initial energy. It determines the number of past samples for each input vector signal of the DTD. From equation 4, we define the energy of the error vector signal as:

$$\|\mathbf{e}(n)\|^2 = \|\mathbf{d}(n) \text{ - } \hat{\mathbf{y}}(n)\|^2 \tag{13}$$

$$\|\mathbf{e}(n)\|^2 = \|\mathbf{d}(n)\|^2 + \|\hat{\mathbf{y}}(n)\|^2 - 2\mathbf{d}^T(n)\hat{\mathbf{y}}(n) \tag{14}$$

With: $\|.\|$, the Euclidian norm of a vector.
Equation 15 can be defined as the decision variable ξ_{EE} (n) of the DTD.

$$\xi_{EE}(n) = \frac{\|\mathbf{e}(n)\|^2}{\|\mathbf{d}(n)\|^2 + \|\hat{\mathbf{y}}(n)\|^2} \tag{15}$$

With:

$0 < \xi_{EE}(n) < 1$ if $\{ \mathbf{d}^T(n)\ \hat{\mathbf{y}}(n) \} > 0$

$\xi_{EE}(n) > 1$ if $\{ \mathbf{d}^T(n)\ \hat{\mathbf{y}}(n) \} < 0$

- If $\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$ are independents, $\xi_{EE}(n)$ =1, it is the case of orthogonality when DT is present.
- If $\mathbf{d}(n) = \hat{\mathbf{y}}(n)$, $\xi_{EE}(n) = 0$, it is the theoretical case of similarity when DT is not present.

Variations values of $\xi_{EE}(n) \geq 0$ reflect or not the presence of DT-situations. In Fig. 2, we show the variation range of the threshold levels (zones Z_0 and Z_1) where it will be judicious that a constant threshold level (T_{EE}), will be set initially in zone Z_0 to control the adaptive filter $\hat{\mathbf{h}}$.

The binary decision is then calculated as follows:

- if $\xi_{EE} > T_{EE}$, DT detected, the binary decision = 1, then no adaptation of the filter $\hat{\mathbf{h}}$;
- if $\xi_{EE} \leq T_{EE}$, DT not detected, the binary decision = 0, then adaptation of the filter $\hat{\mathbf{h}}$.

In practical cases or under hostile environments, the choice of a fixed threshold level with other methods will no longer be valid and must imperatively be replaced by an adaptive threshold [30]. However, the proposed method presents a nice property based on its ability to initially set one and only one fixed threshold level with $T_{EE} \approx 0$.When the far-end signal is present, the relation between the vector signals $\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$ swings between two states : slightly correlated (when $\mathbf{v}(n) \neq 0$) and strongly correlated (when $\mathbf{v}(n) = 0$). It is considered that if the value of the threshold level T_{EE} is fixed in the zone Z_0, the better the correlation between the two vector signals ($\mathbf{d}(n)$ and $\hat{\mathbf{y}}(n)$) and the adaptation of the filter $\hat{\mathbf{h}}$ will be initiated.

Initial energies of the different input vector signals of the DTD are computed with a constant length N of the temporal window. The energy evolution of each vector signal is based on the preliminary calculation of an initial quantity of energy with a small number N of samples.

We have:

$$\|\mathbf{e}(j)\|^2 = \sum_{i=j}^{N+j-1} e^2(i) \tag{16}$$

$$\|\mathbf{d}(j)\|^2 = \sum_{i=j}^{N+j-1} d^2(i) \tag{17}$$

$$\|\hat{\mathbf{y}}(j)\|^2 = \sum_{i=j}^{N+j-1} \hat{y}^2(i) \tag{18}$$

Initially, the decision variable is calculated as:

$$\xi_{EE}(0) = \frac{\|\mathbf{e}(0)\|^2}{\|\mathbf{d}(0)\|^2 + \|\hat{\mathbf{y}}(0)\|^2} \tag{19}$$

When error signal $e(n)$ evolves with time, we get the following:

at time n=1

$$\|\mathbf{e}(1)\|^2 = \sum_{i=1}^{N} e^2(i) = \sum_{i=0}^{N-1} e^2(i) - e^2(0) + e^2(N)$$
$$= \|\mathbf{e}(0)\|^2 - e^2(0) + e^2(N) \tag{20}$$

at time n=2

$$\|\mathbf{e}(2)\|^2 = \sum_{i=2}^{N+1} e^2(i) = \sum_{i=1}^{N} e^2(i) - e^2(1) + e^2(N+1)$$
$$= \|\mathbf{e}(1)\|^2 - e^2(1) + e^2(N+1) \tag{21}$$

$$\vdots \qquad\qquad \vdots$$

at time n=k

$$\|\mathbf{e}(k)\|^2 = \sum_{i=k}^{N+k-1} e^2(i) = \sum_{i=k-1}^{N+k-2} e^2(i) - e^2(k-1) + e^2(N+k-1)$$
$$= \|\mathbf{e}(k-1)\|^2 - e^2(k-1) + e^2(N+k-1) \tag{22}$$

Idem, we get with $\mathbf{d}(k)$ and $\hat{\mathbf{y}}(k)$:

$$\|\mathbf{d}(k)\|^2 = \|\mathbf{d}(k-1)\|^2 - d^2(k-1) + d^2(N+k-1) \tag{23}$$

$$\|\hat{\mathbf{y}}(k)\|^2 = \|\hat{\mathbf{y}}(k-1)\|^2 - \hat{y}^2(k-1) + \hat{y}^2(N+k-1) \tag{24}$$

Decision variable that evolves continuously with time is given at k ($k > 0$) by:

$$\xi_{EE}(k) = \frac{\|\mathbf{e}(k-1)\|^2 - e^2(k-1) + e^2(N+k-1)}{\|\mathbf{d}(k-1)\|^2 - d^2(k-1) + d^2(N+k-1) + \|\hat{\mathbf{y}}(k-1)\|^2 - \hat{y}^2(k-1) + \hat{y}^2(N+k-1)} \tag{25}$$

5- Computational Complexity

As previously reported, energy evaluation of each input vector signal of the DTD is computed by using a temporal window initialized at the beginning with a constant length N. The calculation moves with time sample by sample and the decision variable is then evaluated on each time. In Fig. 3, we show an example of a moving temporal window which tracks energy variations of a signal.

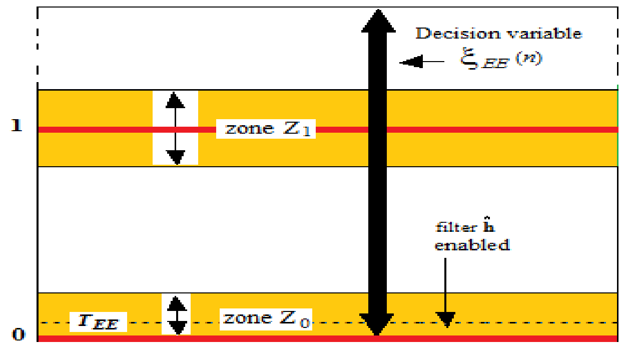


Fig. 2 Variation range of the threshold levels

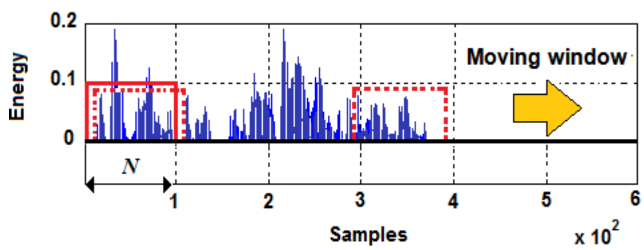


Fig. 3 Moving temporal window

It should be noted that the initial calculation of the energy performed on N samples for each input vector signal is done only at the beginning of the process. Thus, and as the evolution of these signals with time, the squared calculation will be done only on the new sample which replacing the oldest. Accordingly, by moving the window sample by sample, the total energy of each input signal will evolve continuously.

After N iterations, a first move of the window is thus achieved and the process works like a FIFO memory. The Fig. 4 shows an example of the first move of the initial window. At each time and for each input vector signal, we have one and only one squared sample computed. We require per iteration: 1 addition, 1 division and for each signal vector, 1 multiplication, 1 addition and 1 subtraction to compute the decision variable (i.e. 11 operations). A comparison between the previous and proposed method for the total number of computations per iteration is given in Table 1.

Table 1: Computational complexity per iteration

Method	Add	Sub	Mul	Div	Comp
Geigel	0	0	0	1	L_G-1
NCC	2	1	6	1	0
Proposed	4	3	3	1	0

The comparison indicates that Geigel method has a higher computational complexity. The algorithm depends directly on the tap-length L_G of the window used to calculate the maximum of $\mathbf{x}$(n) samples. On the contrary, the proposed and NCC methods are independent regardless of this parameter. Furthermore, the proposed method remains faster than NCC with only three operations of multiplication per iteration. It appears that the proposed method can be considered more efficient for optimizing computation time.

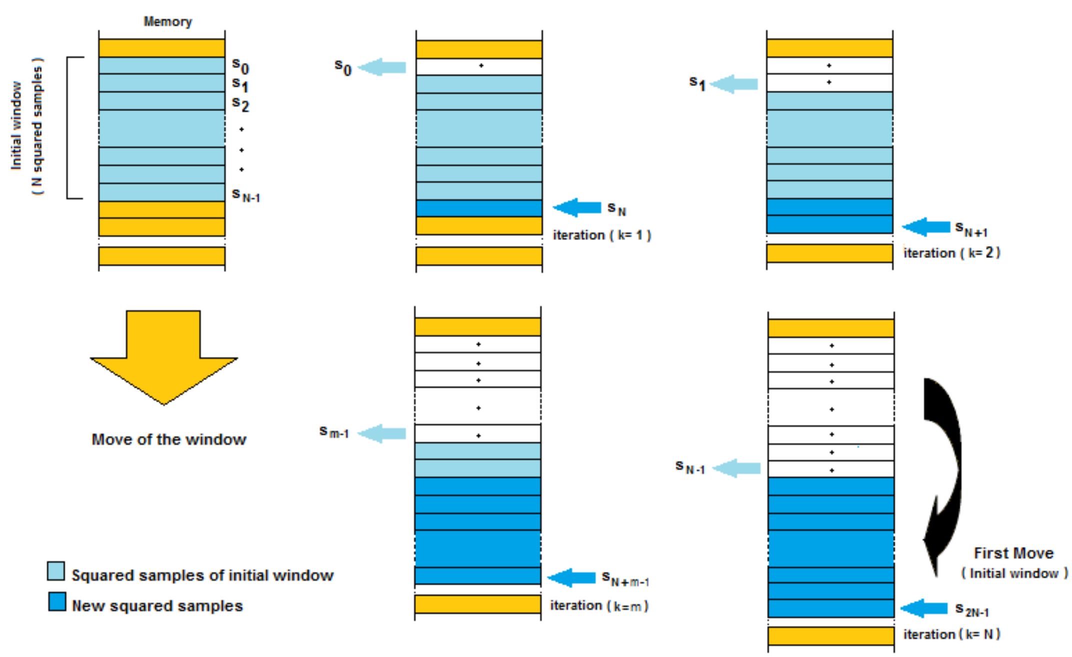


Fig. 4 First Move of the initial window based on FIFO technique

6- Simulation results

In this section, we evaluate the performances of the proposed method compared with Geigel and NCC using three different scenarios of speech signals (Sc1, Sc2, and Sc3) which are sampled at 8 kHz and issued from the NOIZEUS database. The echo model is based on the real impulse response with the length of echo-path L = 128 [31,32]. The three scenarios are presented in Fig. 5.

Three criteria for evaluating the performance of the proposed method are used: Misalignment, Echo Return Loss Enhancement (ERLE) and the probability of miss detection (P_m) [33,34].

The criteria are given as follows:

$$Misalignment(dB) = 10\log_{10}\left[\frac{\left\|\hat{\mathbf{h}}(n)-\mathbf{h}\right\|^2}{\left\|\mathbf{h}\right\|^2}\right] \tag{26}$$

$$ERLE(dB) = 10\log_{10}\left(\frac{E\left\{\left|d(n)^2\right|\right\}}{E\left\{\left|e(n)^2\right|\right\}}\right) \tag{27}$$

$$P_m = 1 - \frac{\sum_{n=1}^{M}\bar{x}(n)\bar{v}(n)\phi(n)}{\sum_{n=1}^{M}\bar{x}(n)\bar{v}(n)} \tag{28}$$

where:

P_m is defined as the probability of detection failure when DT is present.

$\bar{x}(n)$ is the voice activity detection of far-end signal *x(n)*.

$\bar{v}(n)$ is the voice activity detection of near-end signal *v(n)*.

$\phi(n)$ is the binary decision of the DTD.

In the first step tests, a comparison of the different methods is performed with the background noise *w(n)=0*.Parameters used to update the adaptive filter $\hat{\mathbf{h}}$ are summarized in Table 2. Geigel, NCC and the proposed method have been performed respectively with the best parameter values selected for the scenario Sc1 and indicated in Table 3.

Table 2: Parameter values of AEC adaptive filtering

Parameter	Value
β	0.3
C	5.10^{-6}
L	128

Table 3: Parameter values selected for the different methods

Method	Parameter	Value
Geigel	T_G	0.8
	L_G	128
NCC	T_{NCC}	0.982
	λ	0.95
Proposed	T_{EE}	0.001
	N	40

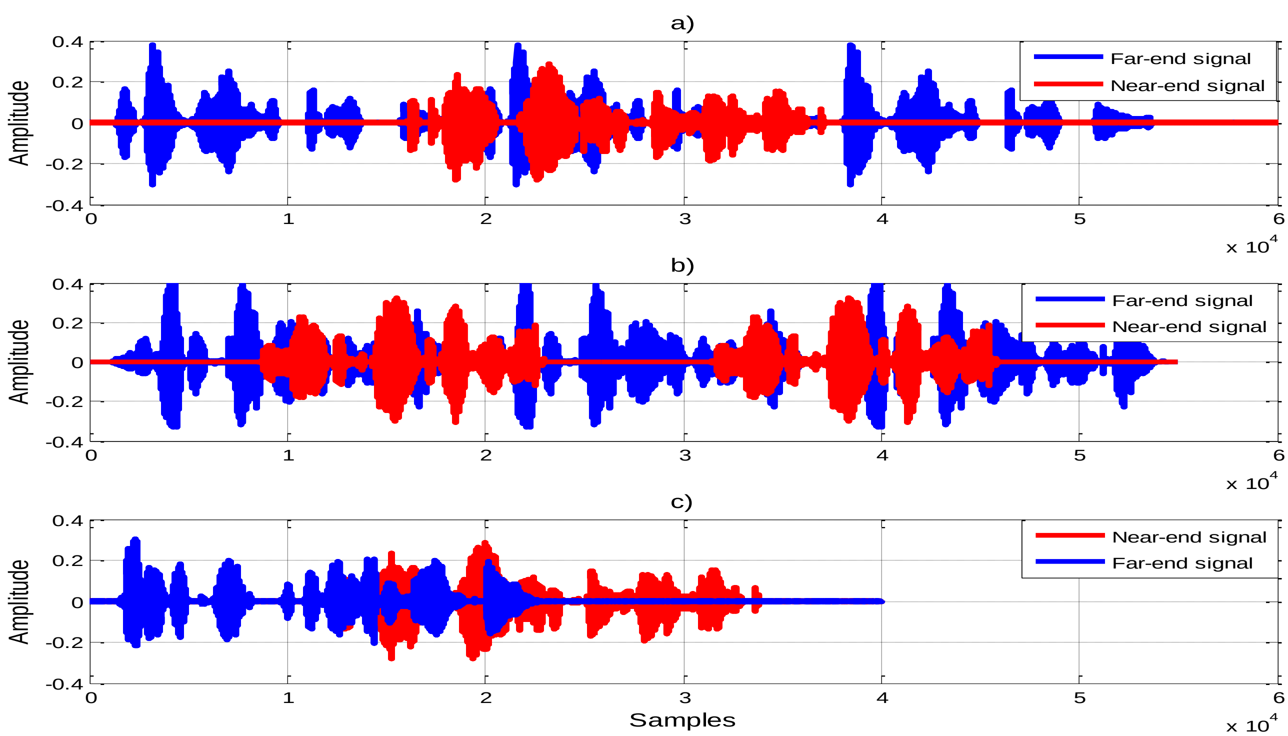


Fig. 5 Speech signals of the three Scenarios, a) Scenario Sc1, b) Scenario Sc2, c) Scenario Sc3.

Table 4: Parameter values of ERLE of the different methods with the three Scenarios.

Scenario	Geigel			NCC			Proposed		
	Peak	Average	Min	Peak	Average	Min	Peak	Average	Min
Sc1	46,49	17,14	-0,77	51,38	17,77	-0,33	52,19	19,96	-0,65
Sc2	60,50	19,69	-1,34	62,80	14,60	-0,47	67,96	23,97	-0,51
Sc3	47,98	11,26	-0,32	45,80	10,63	-0,96	45,85	11,47	-0,98

The ERLE criterion is considered to be one of the most used criteria in performance measurements of AEC algorithms. Recommendation G.131 of the International Telecommunications Union (ITU) requires an attenuation of more than 40 dB in the absence of double-talk [35]. The obtained results with the above scenarios presented in Table 4 and Fig. 6, confirm the superiority of the proposed method with peak values of an echo attenuation more than (52 dB for Sc1, 67 dB for Sc2, and 45 dB for Sc3).

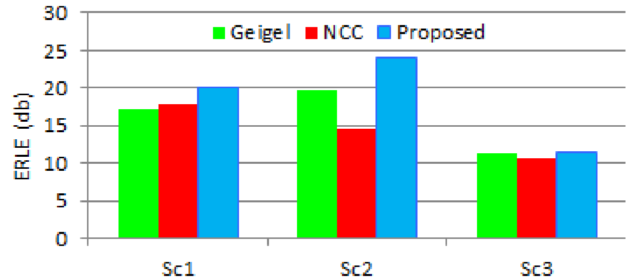
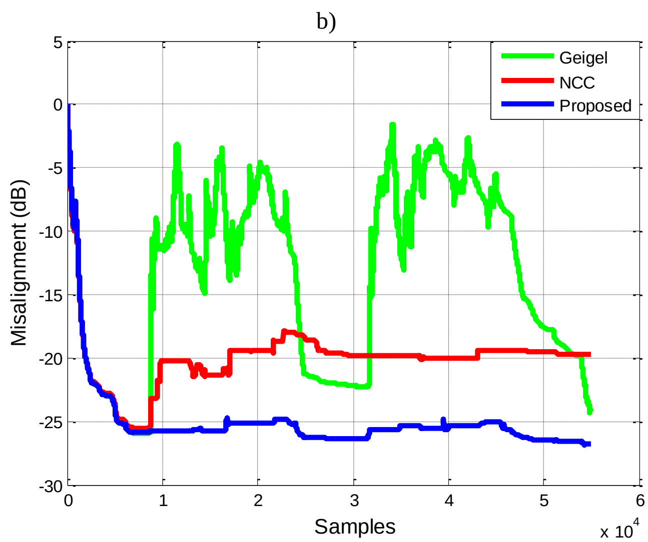


Fig. 6 Evolution of ERLE average of the different methods with the three scenarios

In Fig. 7, we compare the performance of the different methods in terms of misalignment. We remark that in the single-talk and before the apparition of the DT-period, the filter $\hat{\mathbf{h}}$ converges. Indeed, the proposed method maintained the constancy of the filter coefficients as soon as a DT-period occurred, whereas the NCC does false detection with a relative divergence. The Geigel method has detected too late the occurrence of DT-period with more divergence of the filter $\hat{\mathbf{h}}$. Therefore, the proposed method shows its superiority in terms of small steady-state misalignment and stability of decision variable.
In order to validate the proposed method concerning the choice of the parameter value of T_{EE} indicated in Table 3, we propose to illustrate in Fig. 8 the evolution of the decision variable $\xi_{EE}(n)$ obtained with the above scenarios.



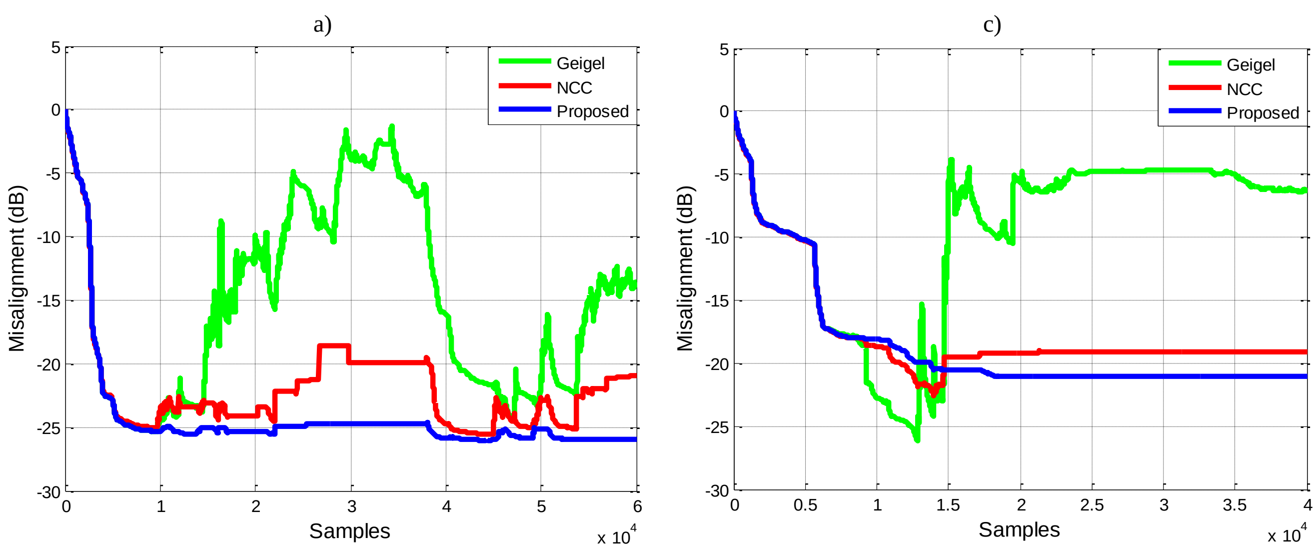


Fig. 7 Misalignment evaluation of the different methods with the three scenarios: a) Sc1, b) Sc2, c) Sc3.

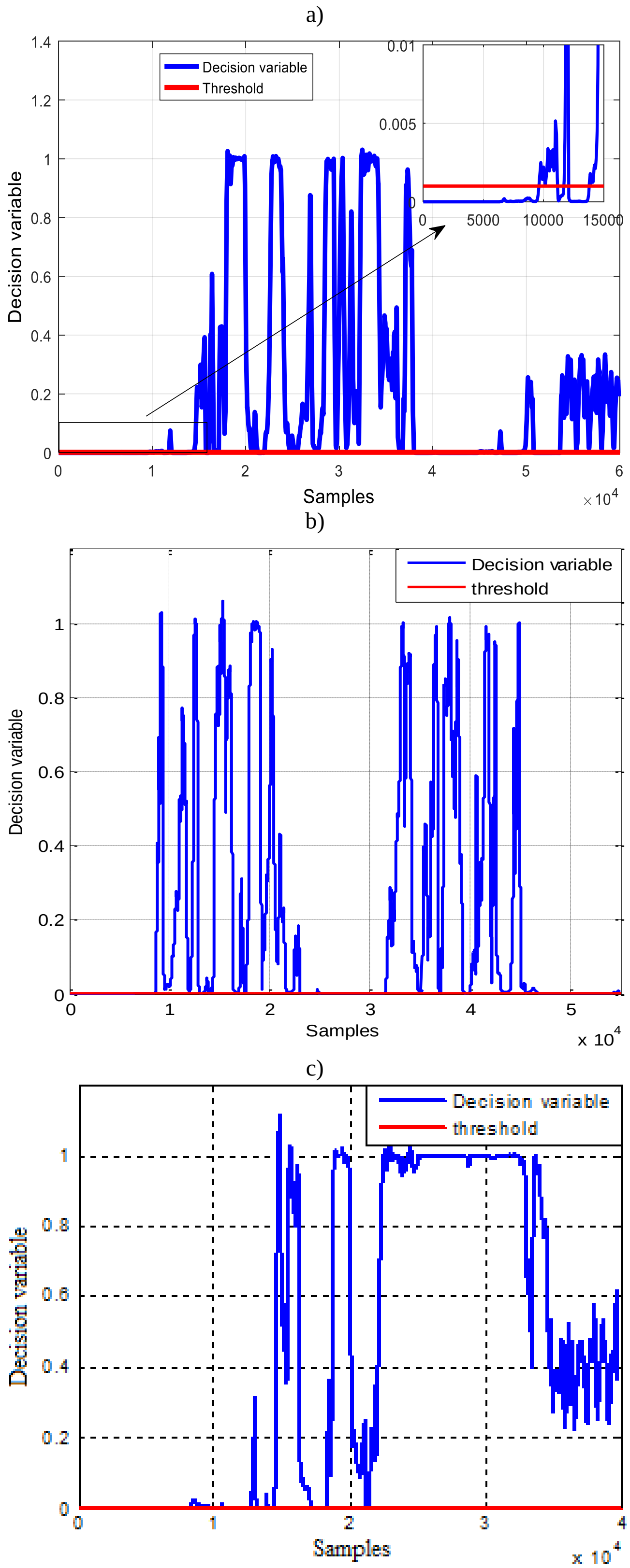


Fig. 8 Evolution of the decision variable of the proposed method with the three scenarios: a) Sc1, b) Sc2, c) Sc3.

It appears that with the scenarios (Sc1, Sc2 and Sc3) the decision variable $\xi_{EE}(n)$ displays a very close value to zero during single-talk periods and confirms the choice of a constant threshold level fixed in the zone Z_0.

To assess the impact of the fixed threshold level on the performance of the above methods, we show in Fig.9 the misalignments obtained with different threshold levels for the two scenarios (Sc2 and Sc3). It can be seen that the proposed method shows for an appropriate threshold level (T_{EE} = 0.001) initially set in zone Z_0 with scenario Sc1, leads to a result without degradation of the misalignment performance in scenarios Sc2 and Sc3. On the other hand, with Geigel and NCC methods, it can be seen that the threshold levels chosen with scenario Sc1 have been replaced by other more adequate threshold levels thus maintaining the performance of the corresponding misalignments. Therefore, we consider that for the proposed method, the T_{EE} threshold level initially set for a given scenario will also be valid with any other scenarios. Rather, Geigel and NCC methods will require an adaptive threshold level to maintain misalignment performance.

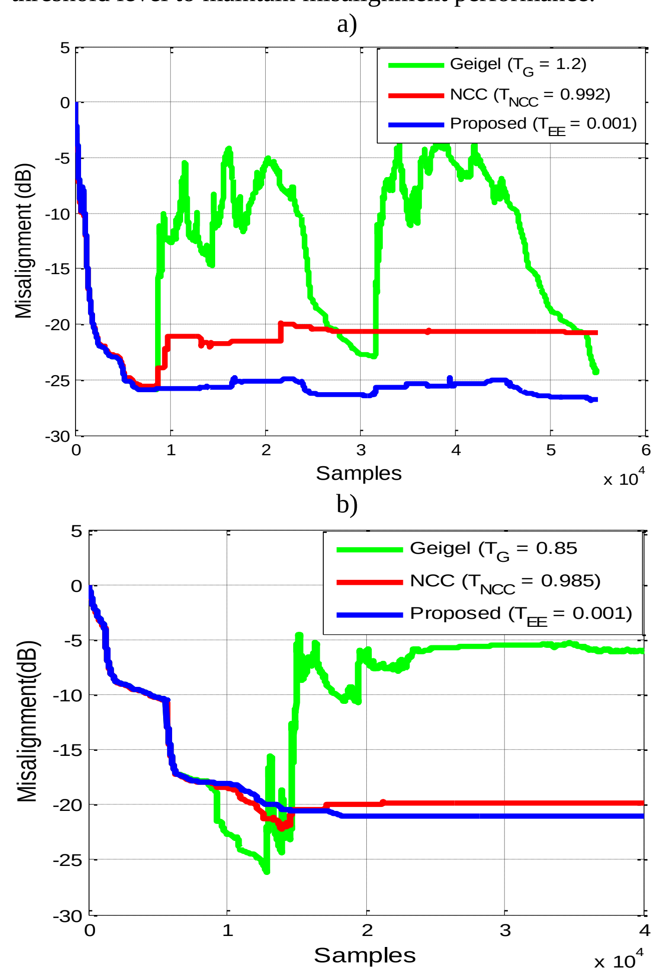


Fig. 9 Misalignment evaluation of the different methods

with variable threshold, a) Sc2, b) Sc3.

In Fig. 10, we propose to evaluate with scenario Sc1 the impact of the length N on the misalignment of the proposed method. The results show misalignments with different values of N which demonstrate that a better performance is obtained with an appropriate set of N (N<100). We confirm that the preliminary calculation of energies requires a small number of samples and a reduced length N of the moving temporal window justifies a good tracking capability.

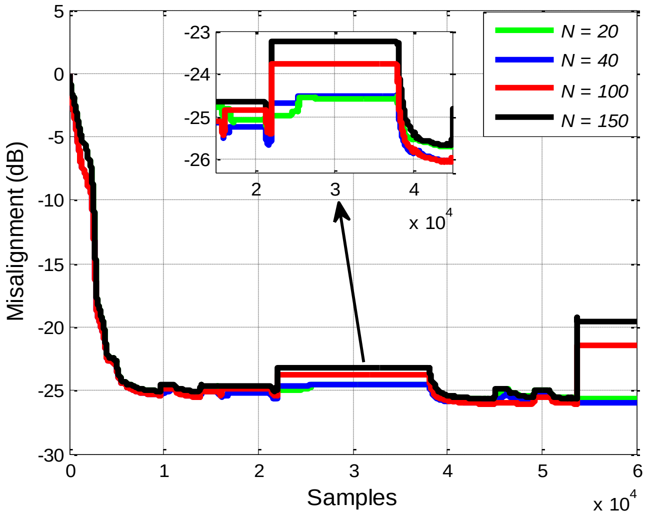


Fig. 10 Misalignment evaluation of the proposed method with different values of the length N

In order to evaluate ERLE and misalignment in a noisy environment ($w(n) \neq 0$),an independent white Gaussian noise is added to the echo signal of the scenario Sc1 with different signal-to-noise ratio (SNR) in the period between 6400 and 60000samples. Note that a constant noise with SNR = 50 dB is added only to the first 6400 input samples.

The SNR is defined as:

$$SNR(dB) = 10\log_{10}\left\{\frac{E\left[|y(n)|^2\right]}{E\left[|w(n)|^2\right]}\right\} \tag{29}$$

Near-end and Far-end signals are used with different levels of near-end-to-far-end ratio (NFR), which is calculated as:

$$NFR(dB) = 10\log_{10}\left\{\frac{E\left[|v(n)|^2\right]}{E\left[|x(n)|^2\right]}\right\} \tag{30}$$

We show in Table 5 parameter values of ERLE in a noisy environment obtained from the different methods with scenario Sc1. The results demonstrate a better and an appropriate ERLE values performed by the proposed method compared to Geigel and NCC.

Fig.11, illustrates clearly the superiority of ERLE average values obtained by the proposed method in a noisy environment.

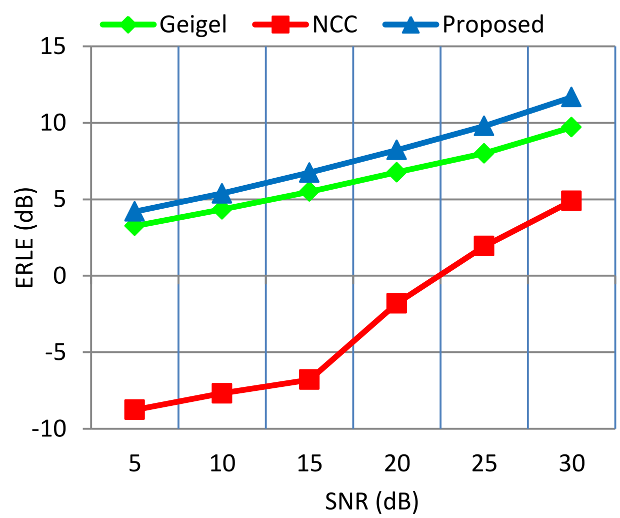


Fig. 11 Evolution of ERLE average for the different methods in a noisy environment

Table 5: Parameter values of ERLE of the different methods in a noisy environment.

SNR (dB)	Geigel			NCC			Proposed		
	Peak	Average	Min	Peak	Average	Min	Peak	Average	Min
5	38,46	3,26	-1,36	38,46	-8,75	-30,54	38,46	4,20	-0,90
10	38,46	4,34	-1,19	38,46	-7,67	-32,05	38,46	5,38	-0,72
15	38,46	5,50	-1,41	38,46	-6,79	-31,90	38,46	6,74	-0,55
20	38,46	6,76	-1,10	38,46	-1,79	-23,13	38,46	8,22	-0,50
25	38,46	8,00	-1,18	38,46	1,95	-9,94	38,46	9,79	-0,57
30	38,46	9,71	-1,16	38,46	4,89	-7,92	40,80	11,67	-0,59

Misalignment evaluation in a noisy environment with scenario Sc1 is illustrated in Fig. 12. The results show that the proposed method presents good performances in terms of misalignment and minimizing false detection in the DT-situation. Robustness against additive noise of the proposed method is clearly appeared compared to other ones.

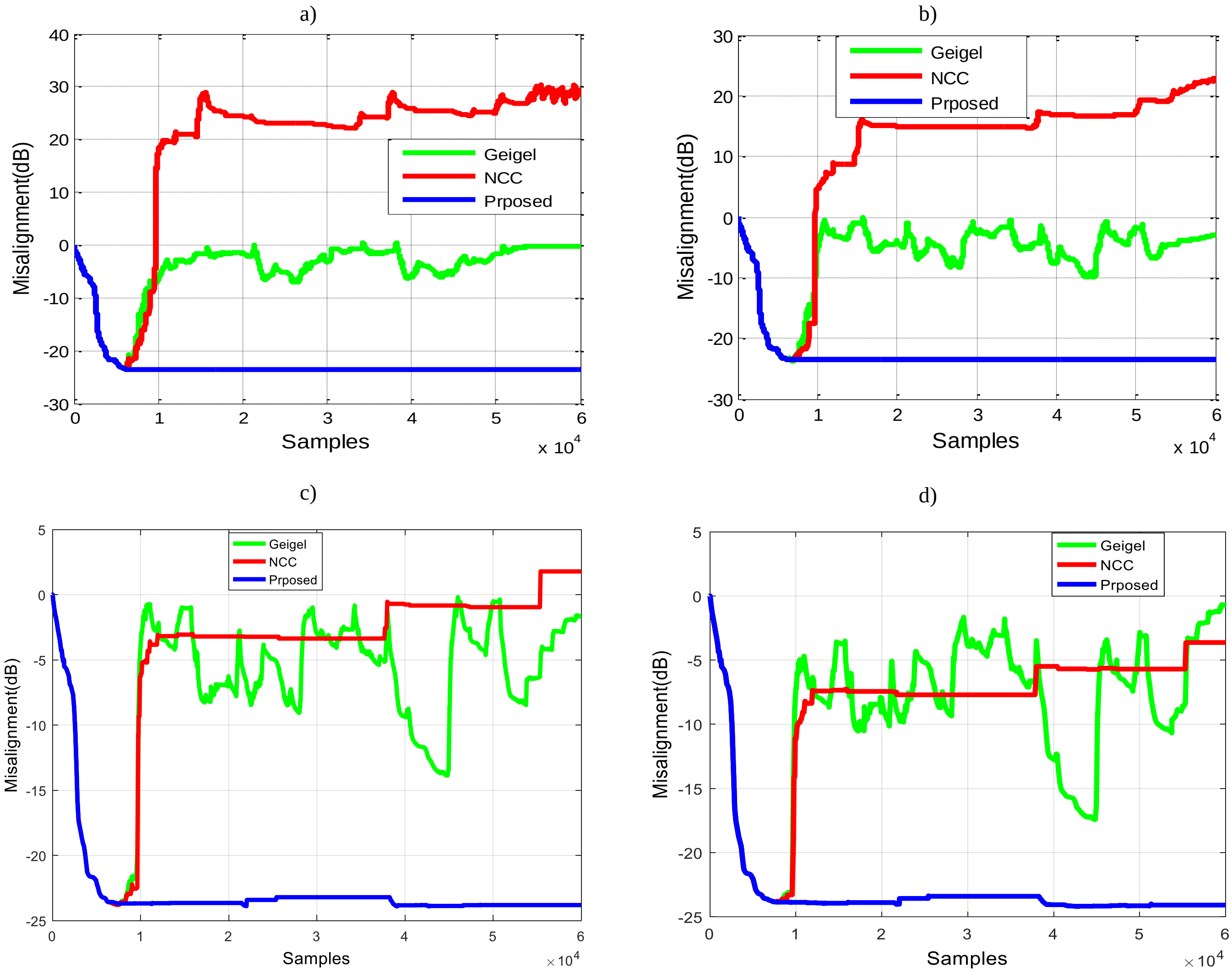


Fig. 12 Misalignment evaluation of the different methods in a noisy environment, a) SNR=5 dB, b) SNR=15 dB, c) SNR= 25 dB, d) SNR= 30 dB

To simulate the change in the echo-path, we increase the gain of the acoustic channel by 10 at sample 31000. The obtained results with scenario Sc1 are shown in Fig. 13. They demonstrate a good tracking capability by the proposed method which can distinguish between the near-end signal and an abrupt change of the acoustic channel.

Objective performance evaluation based on the probability of missed detection P_m is presented in Fig. 14. It is calculated with SNR = 20 dB as a function of NFR values varying between -10 dB and 20 dB. The used threshold for each method is chosen to give a probability of false detection $P_f = 0.2$ which is defined as the probability of declaring detection when DT does not exist. It is calculated without the near-end signal as:

$$P_f = \frac{1}{M} \sum_{n=1}^{M} \bar{x}(n)\phi(n) \tag{31}$$

The obtained result demonstrates that the proposed method is better than Geigel and NCC in terms of the probability of missed detection when NFR varies more than -10 dB.

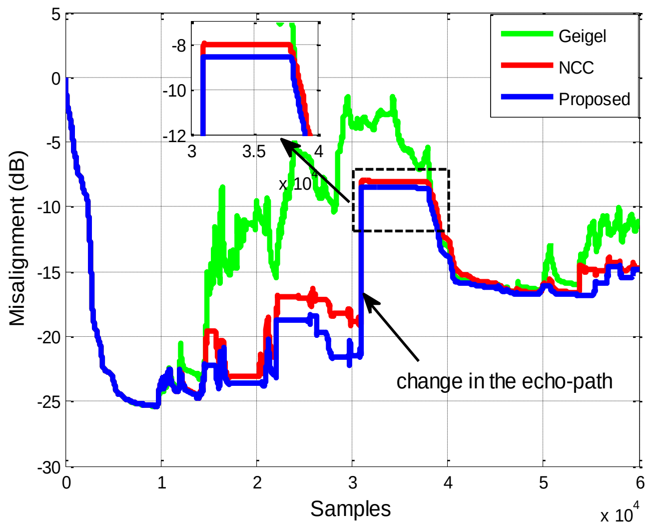


Fig. 13 Misalignment evaluation with a change in the echo-path

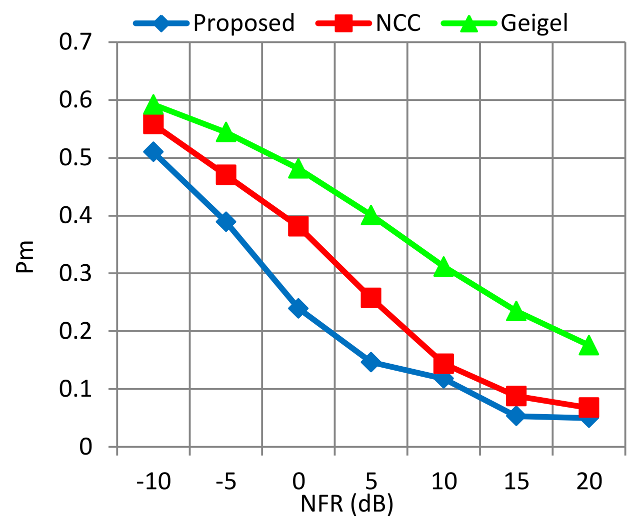


Fig. 14 Probability of missed detection

7- Conclusion

In this paper, we have presented a new and efficient method used for AEC systems where the main purpose is to halt quickly and accurately the update filter coefficients during the DT-periods. The method is based on a moving temporal window that tracks variations of the error energy compared to the sum of energies of the estimated echo and the microphone signals. We consider that the decision variable based on a window that moves with time to track variations of the error energy improves the distinguishing capability between far-end and near-end speech signals. Computer simulation has demonstrated the superiority of the proposed method in terms of small steady-state misalignment, high ERLE, and robustness against the additive white noise and abrupt change in the echo-path. It has also presented improvement in terms of minimizing the number of miss detection and false alarm with no variable threshold level. As an algorithm performed with FIFO technique, the proposed method can be considered also efficient for optimizing computation time. It is significantly simpler and has the capability to outperform conventional NCC methods. Further work remains necessary to compare it with other recent methods.

References

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, S. L. Gay, “Advances in network and acoustic echo cancellation. Digital Signal Processing,” Springer, Berlin, Heidelberg, 2001.

[2] M. M. Sondhi, “An adaptive echo canceler,” The Bell Syst, Technical journal, Vol. 46, No. 3, 1967, pp. 497–511.

[3] J. M. Gil-Cacho, “Adaptive filtering algorithms for Acoustic Echo Cancellation and Acoustic feedback control in speech communication applications,” PhD. Thesis, University of Belgium Ku Leuven, 2013.

[4] S. Haykin, “Adaptive filter Theory,” Prentice-Hall, Inc, Upper Saddle River, NJ, USA, 1996.

[5] J. Benesty, T. Gänsler, “Audio signal processing for next generation multimedia communication systems,” Kluwer Academic Publishers, 2004.

[6] F. Huang, J. Zhang, S. Zhang, “Combined-step size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments,” IEEE Transactions on Circuits and Systems II: Express Briefs, Vol. 63, No. 5, 2015, pp. 493-497.

[7] Y. R. Chien, J. Li-You, “Convex combined adaptive filtering algorithm for acoustic echo cancellation in hostile environments,” IEEE Access, Vol. 6, 2018, pp. 16138-16148.

[8] D. Duttweiler, “A twelve-channel digital echo canceler,” IEEE Transactions on Communications, Vol. 26, No. 5, 1978, pp. 647-653.

[9] H. Ye, B. X. Wu, “A new double-talk detection algorithm based on the orthogonality theorem,” IEEE Transactions on Communications, Vol. 39, No. 39, 1991, pp. 1542-1545.

[10] J. Benesty, D. R. Morgan, J. H. Cho, “A new class of double-talk detectors based on cross-correlation,” IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 2, 2000, pp. 168-172.

[11] M. A. Iqbal, J. W. Stokes, S. L. Grant, “Normalized double-talk detection based on microphone and AEC error cross- correlation,” IEEE International Conference on Multimedia and Expo, 2007, pp. 360-363.

[12] P. S. R. Diniz, “Adaptive Filtering Algorithms and Practical Implementation,” Springer, 2013.

[13] M. Hajiabadi, “Acoustic Noise Cancellation Using an Adaptive Algorithm Based on Correntropy Criterion and Zero Norm Regularization,” JIST Journal of Information

Systems and Telecommunication, Vol. 3, No. 3, 2015, pp. 150-156.

[14] Hun, Choi.Hyeon-Deok, Bae, "Subband Affine Projection Algorithm for Acoustic Echo Cancellation System," EURASIP Journal on Advances in Signal Processing, 2007, pp. 1-12.

[15] B. H. Yang, "An adaptive filtering algorithm for non-Gaussian signals in alpha-stable distribution," Traitement du Signal, Vol. 37, No. 1, 2020, pp.69-75.

[16] S. Hannah, D. Samiappan , R. Kumar , A. Anand, A. Kar, "Variable tap-length non-parametric variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation," Applied Acoustics, Vol. 159, 2020.

[17] M. Hamidia, A. Amrouche, "A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation," Digital Signal Processing, Vol. 60, 2017, pp. 99-112.

[18] V. Thien-An, H. Ding, M. Bouchard, "A survey of double-talk detection schemes for echo cancellation applications," Canadian Acoustics, Vol. 32, No. 3, 2004, pp. 144-145.

[19] M. Benziane, M. Bouamar, M. Makdir, "Doubletalk detection based on enhanced Geigel algorithm for acoustic echo cancellation," In 2018 6th International Conference on Control Engineering & Information Technology (CEIT), 2018, pp. 1-5.

[20] T. Gänsler, J. Benesty, "A frequency-domain double-talk detector based on a normalized cross-correlation vector," Signal Processing, Vol. 81, No 8, 2001, pp. 1783–1787.

[21] J. Benesty, T. Gänsler, "A multichannel acoustic echo canceler double-talk detector based on a normalized cross-correlation matrix*," European Transactions on Telecommunications, Vol. 13, No 2, 2002, pp. 95–101.

[22] T. Gänsler, J. Benesty, "The fast normalized cross-correlation double-talk detector," Signal Process, Vol. 86, No. 6, 2006, pp. 1124–1139.

[23] T. Gansler, M. Hansson, C.J.Ivarsson, G. Salomonsson, "A double-talk detector based on coherence,"IEEE Transactions on Communications, Vol. 44, No. 11, 1996, pp. 1421-1427.

[24] H. Bao, Y. Yang, J. Liu, X. Ba, Q. Yuan, "A robust algorithm of double talk detection based on voice activity detection," Proc. Inter. conf. on Audio Language and Image Processing, 2010, pp. 12–15.

[25] S. Cecchi, L. Romoli, F. Piazza, "Multichannel Double-Talk Detector based on Fundamental Frequency Estimation," IEEE Signal Processing Letters, Vol. 23, No. 1, 2016, pp. 94-97.

[26] Y. Zhenhai, F. Yang, J. Yang, "Optimum step-size control for a variable step-size stereo acoustic echo canceller in the frequency domain," Speech Communication, Vol. 124, 2020, pp. 21–27.

[27] S. J. Park, C. G. Cho, C. Lee, D. H. Youn, S. H. Park, "Integrated echo and noise canceller for hands free applications," IEEE Transactions on circuits and systems, Part II, Analog and Digital Signal Processing, Vol. 49, No. 3, 2002, pp. 188-195.

[28] Y. Hua, "Adaptive filter theory and applications," PhD. Thesis, South-East university, China, 1989

[29] Honig, M.L., Messerschmitt, D.G., "Adaptive Filters," Kluwer, 1984.

[30] M. Benziane, M. Bouamar, M. Makdir, "Simple and Efficient Double-Talk-Detector for Acoustic Echo Cancellation," Traitement du signal, Vol. 37, No. 4, 2020, pp. 585-592.

[31] ITU-T. "Digital Network Echo Cancellers," Recommendation G.168, International Telecommunication Union; Geneva, 2007.

[32] Y. Hu, P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Communication , Vol. 49, No. 7, 2007, pp. 588-601.

[33] H. Wonchul, K. Taehwan, B. Keunsung, "Robust double-talk detection in the acoustic echo canceller using normalized error signal power," Proc. ISSPA'07.UAE, 2007, pp. 1-4

[34] J.H. Cho, D.R. Morgan, J. Benesty., "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 6, 1999, pp. 718–724.

[35] ITU-T. "Digital Network Echo Cancellers," Recommendation G.131, International Telecommunication Union; Geneva, 2003.