IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 2, April, 2023, pp. 701-707
Journal homepage: http://iieta.org/journals/ts

A New Fast Double-Talk Detector Based on the Error Variance for Acoustic Echo Cancellation

Mahfoud Hamidia*, Abderrahmane Amrouche

Speech Communication and Signal Processing Laboratory (LCPTS), Electrical Engineering Faculty, University of Science and Technology Houari Boumediene (USTHB), Algiers 16111, Algeria

Corresponding Author Email: mhamidia@usthb.dz

https://doi.org/10.18280/ts.400229

Received: 4 August 2022
Accepted: 29 January 2023

Keywords:
acoustic echo cancellation, double-talk detection, normalized least mean squares, Geigel, normalized cross-correlation, zero-crossing rate, error signal variance

ABSTRACT

In order to improve the speech quality in communication systems, acoustic echo cancellation techniques are commonly used to mitigate the deleterious effect of acoustic feedback. In fact, double-talk situations hinder the performance of acoustic echo cancellation when the two speakers in the two ends talk simultaneously. For this reason, double-talk detection is included to control the echo canceler system. In this paper we proposed a new method of double-talk detection based on the error signal variance. Opposed to the previous works where the most of the existing methods are based on a comparison between the received far-end and the microphone observation signals, we accurately account for the variation of the error signal. To evaluate the proposed method, we used acoustic echo cancellation based on the normalized least mean square algorithm. Simulation results indicate the good performance of the proposed double-talk detector.

1. INTRODUCTION

Acoustic echo cancellation (AEC) is a technique used to remove acoustic echo in communication systems when the presence of the undesirable feedback impairs significantly the voice quality especially with hands-free terminals. A major source of this echo is the acoustic coupling between the loudspeaker and the microphone at each end [1]. Since the 1960s, acoustic echo cancellation problem has attracted attention of many researchers in several applications, e.g., teleconferencing, car hands-free telephones, mobile phones, voice over internet protocol (VoIP), etc. [2]. In fact, AEC eventually has been understood as an application of adaptive filtering. Herein, the adaptive cancellation is done by estimating the acoustic echo path impulse response using an adaptive finite impulse response (FIR) filter and subtracting the echo estimate from the microphone signal.

In addition, adaptive algorithms are used for updating the FIR filter coefficients where the most frequently used are the gradient based normalized least mean squares (NLMS), the affine projection algorithm (APA), the recursive least squares (RLS) and frequency domain adaptive filters (FDAF) [3].

Undoubtedly, divergence risks of the adaptive filter coefficients may arise during the so-called double-talk (DT) periods, in which both near-end and far-end speakers talk at the same time [4, 5]. In another words, the presence of near-end signal makes a change in the desired signal (echo signal) that produce disturbance in the AEC process.

The main role of double-talk detection (DTD) is to prevent the divergence of the adaptive filter coefficients by halting the update during double-talk periods. In this manner, these coefficients maintain the convergence state in DT periods. On the other hand, pending the absence of the near-end signal the filter coefficients continue the convergence towards their optimal values. Many studies have been conducted in DTD with the goal of improving the AEC process, particularly in DT scenarios. Most of the proposed methods have focused on determining a similarity measure between the near-end and far-end speech signals. For instance, amplitude comparison is based in the well-known Geigel algorithm [6], also its generalized version is considered in the Holder inequality method [7]. Furthermore, cross-correlation (CC) [8, 9], signal envelope [10], coherence [11], and voice activity detection (VAD) [12] similarities are investigated for DTD. The main problem of these methods is the sensitivity to the background noise signal. In addition, frequency-domain methods have been used for DTD, based on a Gaussian mixture models (GMM) [13], spectral analysis [14], spectral slit [15] and time-frequency presentation by Stockwell transform [16]. Also, an audio watermarking technique is investigated in the study of Szwoch et al. [17], multimodal information (sound and image) is exploited in the study of Urakami and Kajikawa [18] and auxiliary adaptive filter is used in the study of Hamidia and Amrouche [19]. Nevertheless, most of these methods of DTD require a high computational complexity which cannot be apply in the practical applications.

Differently to previous works which based on similarities comparison between the near-end and the microphone signals, we propose in this paper a new method of DTD using only the variance of the error signal that considers DTD in this case as a change detection problem.

The rest of this paper is organized as follow. Section 2 introduces the AEC problem with the theoretical formulation. In Section 3, the proposed method of DTD is described. Simulation results are presented in Section 4. Finally, Section 5 concludes the paper.

2. ACOUSTIC ECHO CANCELLATION

2.1 Single-talk scenario

In communication systems, acoustic echo occurs when the loudspeakers are picked-up by the microphone in the terminal device. Acoustic echo cancellation is use to remove this undesirable feedback signal; this issue is classified as a system identification problem. The basic principle of AEC is to generate a replica $\hat{y}(n)$ of the actual echo signal $y(n)$ and subtract it from the microphone signal $d(n)$, as is illustrated in Figure 1. In another words, an adaptive filter is used to identify an unknown system, i.e., the acoustic echo path **h** which is modeled by the impulse response of the loudspeaker-enclosure-microphones (LEM). In the first case, we consider the single-talk scenario, when the near-end speaker is in the silence-state.

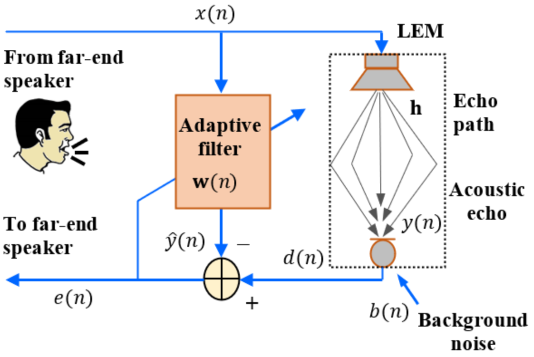

Figure 1. Illustration of AEC system

The acoustic echo signal $y(n)$ at the discrete-time index n is the filter resulting of the near-end signal $x(n)$ through the room impulse response **h** as is modeled by the following equation:

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

where, $\mathbf{h}=[h_0, h_1, \ldots, h_{L-1}]^T$ is the echo path of the length L, the superscript $(\cdot)^T$ denotes transpose of a vector, $\mathbf{x}(n)=[x(n), x(n-1), \ldots, x(n-L+1)]^T$ is a vector containing the L most recent time samples of the received signal $x(n)$, or the far-end signal.

The microphone signal $d(n)$ includes the echo signal $y(n)$ and the background noise signal $b(n)$ as:

$$d(n) = y(n) + b(n) \tag{2}$$

The estimated echo signal $\hat{y}(n)$ generated by the adaptive filter, which is a linear combination of several inputs at time n. This copy version of echo is subtracted from the microphone signal and it is given by:

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

where, $\mathbf{w}(n) = [w_0(n)\, w_1(n), \ldots,\, w_{L-1}(n)]^T$ is the weight vector of the adaptive filter. The length of the adaptive filter is generally equal to the length of the acoustic echo path **h**.

The error signal $e(n)$ at time n is given by:

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

Several adaptive filtering algorithms are proposed to adapt the weights $\mathbf{w}(n)$ of the filter using the feedback of the estimation error. The ideal algorithms should have a speed convergence rate and good tracking capabilities but achieving low misalignment. One of the most used due to its stability and low complexity is the normalized least mean squares (NLMS) algorithm [20]. It is defined by the following update equation:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}^T(n)\,\mathbf{x}(n) + \varepsilon} e(n)\mathbf{x}(n) \tag{5}$$

where, $\mathbf{w}(n+1)$ is the next tap weight value and $\mathbf{w}(n)$ is the present tap weight value of the adaptive filter. μ is the step-size parameter used in the weight vector updating with $0<\mu<2$ for the stability considerations, and $\varepsilon>0$ is a regularization constant used to improve adaptation stability and to avoid division by zero.

2.2 Double-talk scenario

In this scenario, the near-end speaker is in the speech-state and the DT situation is occurred when the echo signal $y(n)$ and the near-end signal $s(n)$ appear simultaneously. This situation causes a quick divergence of the adaptive filter coefficients $\mathbf{w}(n)$ from their optimum values where the microphone signal $d(n)$ in this case combines the echo $y(n)$, the near-end $s(n)$ and background noise $b(n)$ signals. This combination affects the comparison with the estimated echo signal $\hat{y}(n)$ in the adaptive cancellation where a better process requires that the microphone observation contains only the echo signal $y(n)$.

For this raison, DTD is use for controlling the update of the filter coefficients as is shown in Figure 2. In essence, the main goal of DTD is to detect the presence of the near-end speech and freeze the adaptation process in DT situations [21].

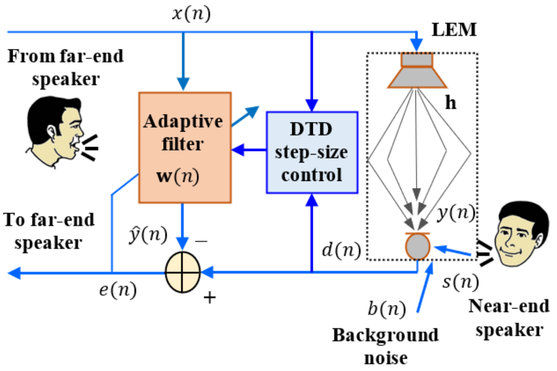

Figure 2. Representation of AEC with conventional DTD

Typically, the DTD calculates a statistic decision $\xi(n)$, and the DT is declared when $\xi(n)$ is lower than the threshold value T. The optimum decision variable $\xi(n)$ for DTD will behave as follows [19]:

$$\begin{cases} \text{If } s(n) = 0 \text{ (double} - \text{talk is not present)}, \xi(n) \geq T. \\ \text{If } s(n) \neq 0 \text{ (double} - \text{talk is present)}, \xi(n) < T. \end{cases}$$

The control of the adaptive filter by DTD is defined as:

$$\text{Control} = \begin{Bmatrix} \xi(n) \geq T, & \text{DTD} = 0 & \text{adaptation} & \mu \neq 0 \\ \xi(n) < T, & \text{DTD} = 1 & \text{no adaptation} & \mu = 0 \end{Bmatrix}$$

where, the update of the adaptive filter coefficients will be stopped during DT situations (μ=0), that means $\mathbf{w}(n+1)=\mathbf{w}(n)$ according to Eq. (5).

The well-known method of DTD is the Geigel algorithm [7], where the variable statistic decision $\xi_G(n)$ of the latter is defined as follows:

$$\xi_G(n) = \frac{\max[|x(n)|\ \ |x(n-1)|, \ldots,\ |x(n-L+1)|]}{|d(n)|} \tag{6}$$

where, max [.] and |.| denote the maximum and the absolute value, respectively.

3. PROPOSED METHOD OF DTD

In this paper, we propose a new method of DTD based on the variance of the error signal. As is known to all, most of the existing methods of DTD based on similarity comparison between the microphone signal $d(n)$ and the received far-end signal $x(n)$.

In the proposed method, we focus only on the variation of the error signal $e(n)$ for detecting DT periods. For this reason, let us assume a frame $\boldsymbol{f}$ of M recent sample history of the error signal which is defined as:

$$\mathbf{e}_{\boldsymbol{M}}(n) = [e(n)\ \ e(n-1), \ldots, e(n-M+1)]^T \tag{7}$$

where, $M \leq L$.

This frame is used for sensing the DT periods where we exploit the change i.e., variance of the error signal to detect the activity of near-end signal from the error signal by calculating the decision variable as follow:

$$\xi_V(n) = 1 - |\max[|\mathbf{e}_{\boldsymbol{M}}(n)|] - \mathrm{var}[\mathbf{e}_{\boldsymbol{M}}(n)]| \tag{8}$$

where, max[.] and var[.] are the maximum and the variance of the frame $\boldsymbol{f}$, respectively.

Double-talk is declared (DTD=1) if $\xi_V(n) < T_V$, where T_V is a constant threshold.

To avoid a false alarm declaration, especially in the beginning of the adaptive filtering process when the energy of the error signal is high, we combine the proposed DTD with the structure of AEC proposed in our previous work [22]. Figure 3 shows the structure of the proposed DTD, the main objective behind this structure is to accelerate the convergence speed of the adaptive filtering algorithm and reduce the steady-state error. Its principle based on a superposition of a short ineffective stationary segment of additive white Gaussian noise (AWGN) in the beginning of the received far-end speech signal.

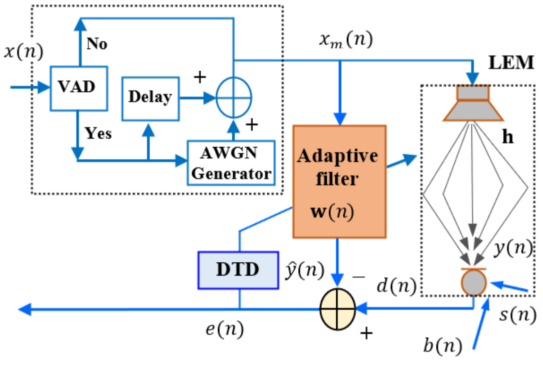

Figure 3. Structure of the proposed DTD

The added block in this new structure of acoustic echo cancellation system is considered as a pre-processing block which accelerates the convergence speed and facilitates the double-talk detection process. This block is composed of three principal sub-blocks as:

(a) The voice activity detection (VAD) sub-block which is used for detecting the presence of the far-end speech signal $x(n)$ and triggering the next sub-blocks, the activity of the far-end speech signal is declared as follow:

$$\mathrm{VAD} = \begin{cases} 0, & \text{if } x(n) \text{ is absent} \quad \text{(No)} \\ 1, & \text{if } x(n) \text{ is present} \quad \text{(Yes)} \end{cases} \tag{9}$$

(b) The delay sub-block is considered for creating a short delay period t_d in the received far-end signal $x(n)$ if the far-end speaker is active.

(c) The role of the AWGN generator is to produce a short stationary segment of additive white Gaussian noise $n_\sigma(n)$ in the generated delay period with a zero mean and ineffective values of variance σ compared to the near-end speech signal energy. This step can accelerate the convergence process of the adaptive filtering and help the learning of their coefficients.

The resulting near-end speech signal $x_m(n)$ from the new structure is defined as follow:

$$x_m(n) = \begin{cases} x(n), & \text{if } \ \mathrm{VAD} = 0 \\ n_\sigma(n) + x(n), & \text{if } \ \mathrm{VAD} = 1 \end{cases} \tag{10}$$

4. EXPIREMENT AND RESULTS

This section presents the experimental tests with the evaluation to prove the effectiveness of our proposed DTD method. We used AEC based on NLMS algorithm controlled by the step-size μ=0.9 and ε=2.2204×10^{-16} where the length of the adaptive filter is L=1024 equals to the length of the acoustic echo path (car cockpit impulse response sampled at 16 kHz) [22]. We use also speech signals for simulating the far-end and the near-end speakers which are sampled at 16 kHz. Furthermore, we adopted the proposed structure in [22] which investigates insignificant samples of the received far-end signal by creating a delay and adding a short period time of the AWGN. We take a period of delay t_d=125 ms with σ^2=0.005 is the variance value of AWGN.

In order to further demonstrate the performance of the proposed DTD we compare it with Geigel [6], normalized cross-correlation (NCC) [9] methods and zero-crossing rate (ZCR) method based DTD proposed in the study [4].

The optimal threshold values are given by: T_G=1.5, T_{NCC}=0.92, T_{ZCR}=0.25 and T_V=0.96 of Geigel, NCC, ZCR and the proposed DTD, respectively. Also, we choice M=512 the length of the frame $\boldsymbol{f}$ contents 512 recent sample history of the signals in ZCR and the proposed DTD.

The real environment is modeled by a white Gaussian background noise signal $b(n)$ that is added to the echo signal $y(n)$ at various signal-to-noise ratio (SNR) values, where the SNR value is defined by:

$$\mathrm{SNR\ (dB)} = 10\log_{10}\left(\frac{E\{|y(n)|^2\}}{E\{|b(n)|^2\}}\right) \tag{11}$$

We evaluated our method by calculating the different probabilities P_d, P_m and P_f of detection, miss and false alarm, respectively, where the output DTD signal is compared with the VAD of near-end signal $s(n)$. Also, we have used three performances measures: the normalized mean square deviation (NMSD) (mismatch system), mean square error (MSE) and echo return loss enhancement (ERLE) which are calculated by:

$$\text{NMSD (dB)} = 10\log_{10}\left(\frac{\|\mathbf{w}(n)-\mathbf{h}\|^2}{\|\mathbf{h}\|^2}\right) \qquad (12)$$

where, $\|\mathbf{w}(n)-\mathbf{h}\|$ is the Euclidian distance between the adaptive coefficients vector and the true echo path vector.

$$\text{MSE (dB)} = 10\log_{10}(E\{|e(n)|^2\}) \qquad (13)$$

where, $E\{\cdot\}$ denotes the mathematical expectation.

$$\text{ERLE (dB)} = 10\log_{10}\left\{\frac{E[|y(n)|^2]}{E[|e(n)|^2]}\right\} \qquad (14)$$

Good performance of AEC system is indicated by the capability to minimize the misalignment, the MSE and maximize the ERLE values.

Figures 4 and 5 illustrate the decision variable and the DTD signal of the proposed method using echo with 2.5s duration and near-end speech signals pronounced in English and background noise modeled by SNR value equals to 60 dB.

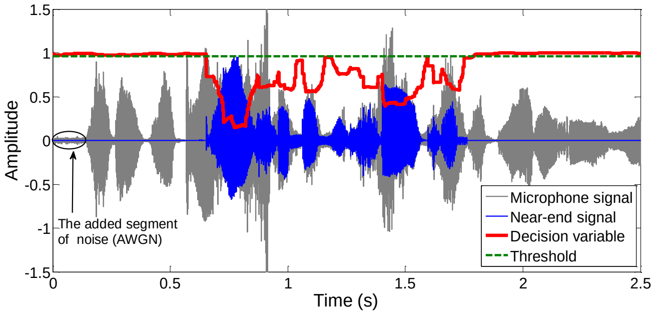

Figure 4. Decision variable of the proposed DTD

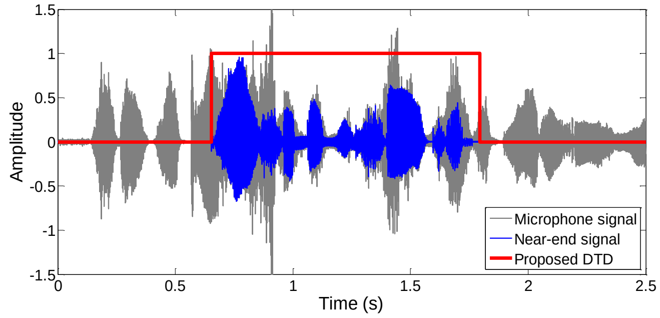

Figure 5. Representation of the proposed DTD decision

In addition, temporal evolutions and spectrograms of the near-end and the output signals are depicted in Figures 6 and 7. The obtained results indicate a great similarity between the near-end and the output signals temporal evolution, also as well as in their spectrograms.

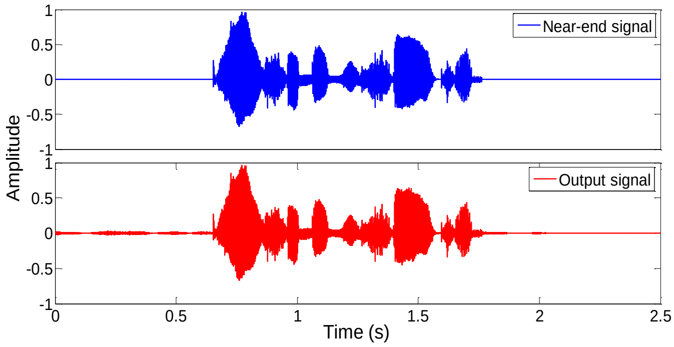

Figure 6. Temporal evolution of speech signals, (blue) near-end, (red) output

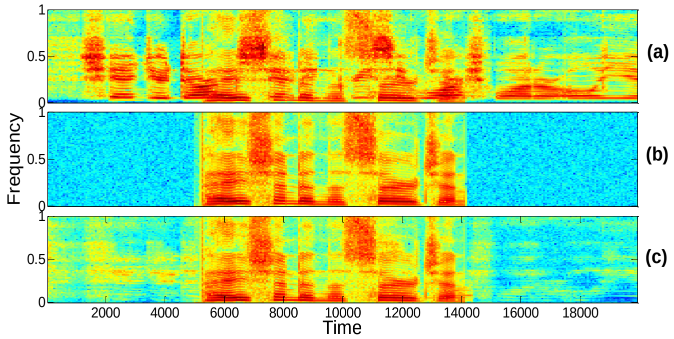

Figure 7. Spectrograms of speech signals, (a) microphone, (b) near-end, (c) output

To confirm effectiveness of the proposed DTD, we include long speech signals pronounced in French in DT scenario with duration of 30 seconds. The near-end speech (double-talk) appears between times 10 and 16.5 seconds.

Similarly to the previous tests, consider Figures 8, 9 and 10, which depict DTD signal and comparison of the near-end and the output (estimated near-end) signals, it is clear that the proposed method also has good performance in terms of echo cancellation for a long duration of speech conversation.

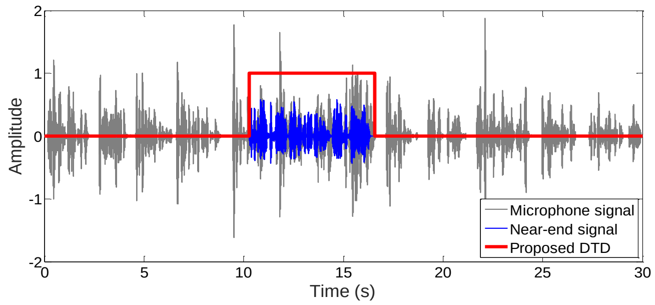

Figure 8. Representation of the proposed DTD decision for long speech signals, SNR=60 dB

A comparison between the proposed DTD, Geigel, NCC and ZCR methods is shown in Figures 11 and 12 using NMSD curves for short and long speech signals. According to the obtained results, the proposed method outperforms the other methods in terms of minimizing the steady-state NMSD where the proposed DTD maintains the NMSD level during DT period and avoids NMSD divergence. In other words, the proposed DTD allows the NLMS algorithm to ignore the effect of the near-end signal on NMSD convergence, such as its

NMSD curves bring closer to the single-talk (ST) curves i.e., absence of the near-end signal.

On the contrary, the presence of DT affects the Geigel, NCC and ZCR behaviours as well as an increase in the NMSD level.

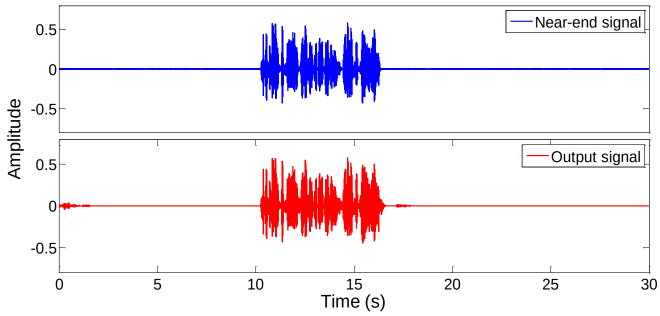

Figure 9. Temporal evolution for long speech signals, (blue) near-end, (red) output

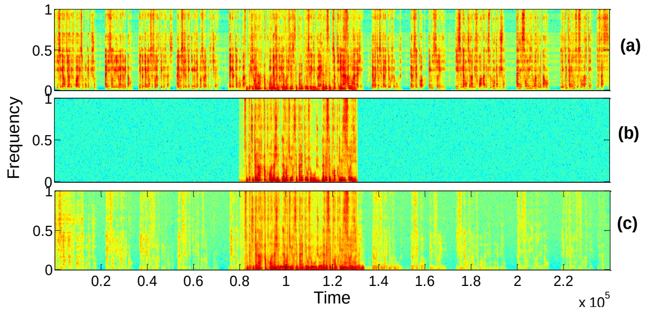

Figure 10. Spectrograms of long speech signals, (a) microphone, (b) near-end, (c) output

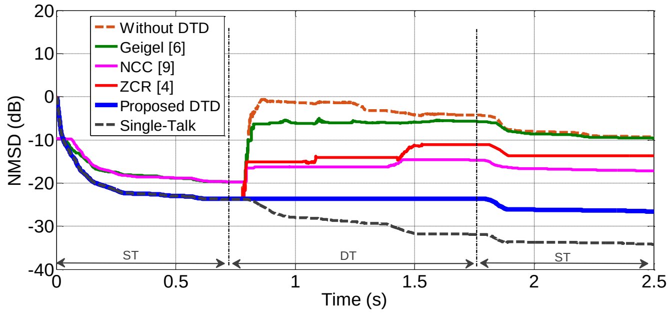

Figure 11. NMSD curves for short speech signals, SNR=60 dB, the near-end speech (double-talk) appears between times 0.625 and 1.76 seconds

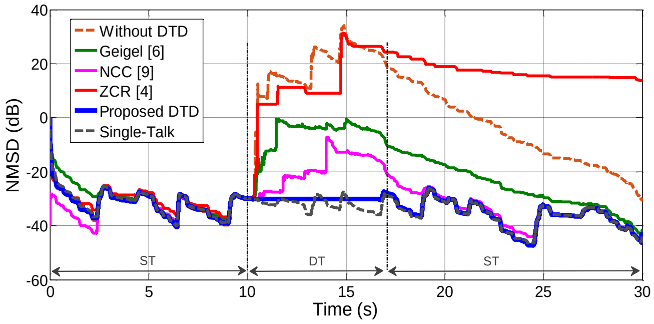

Figure 12. NMSD curves for long speech signals, SNR=60 dB, the near-end speech (double-talk) appears between times 10 and 16.5 seconds

MSE curves are shown in Figure 13, which are evaluated for each 2048 iterations. It is observed that the proposed method of DTD yields better performance in terms of MSE minimizing compared to the others methods. This result indicates a reduction in the error signal (returned echo) energy during the ST periods.

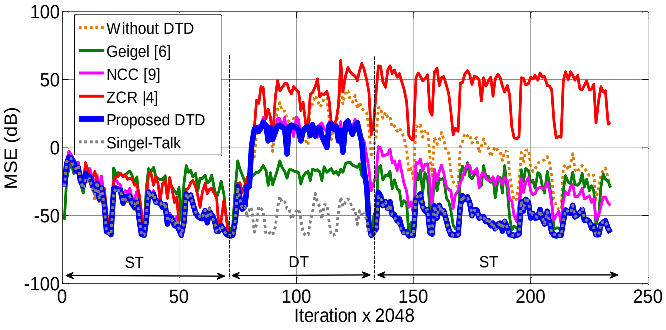

Figure 13. MSE evaluation for long speech signals, SNR=60 dB

According to the ERLE comparison of DTD methods in different SNR values (15 dB, 35 dB and 55 dB) given in Table 1, the obtained results confirm the superiority of the proposed method of DTD compared to the others methods which has large values of ERLE average that can reduce the effect of the acoustic echo.

Table 1. Evaluation of ERLE in function of SNR for different methods of DTD

SNR (dB)	Method	ERLE (dB)		
		Min	Max	Mean
55	Geigel [6]	–10.63	31.37	10.79
	ZCR [4]	–17.62	34.59	8.45
	NCC [9]	–14.03	35.65	15.25
	Proposed	–2.38	40.87	**16.83**
35	Geigel [6]	–13.37	19.59	5.13
	ZCR [4]	–23.09	23.51	1.45
	NCC [9]	–13.87	18.07	3.75
	Proposed	–23.74	23.06	**8.16**
15	Geigel [6]	–9.21	10.43	0.27
	ZCR [4]	–14.20	10.93	0.47
	NCC [9]	–13.35	6.43	0.21
	Proposed	–14.07	10.31	**1.48**

Table 2. Evaluation of PESQ in function of SNR for different methods of DTD

SNR (dB)	Method	PESQ
55	Geigel [6]	0.75
	ZCR [4]	2.23
	NCC [9]	2.23
	Proposed	**3.31**
35	Geigel [6]	1.09
	ZCR [4]	1.58
	NCC [9]	1.51
	Proposed	**2.01**
15	Geigel [6]	0.57
	ZCR [4]	0.88
	NCC [9]	0.61
	Proposed	**1.48**

A perceptual evaluation of speech quality (PESQ) [23] is also considered to evaluate the quality of the output speech signal of each method of DTD as is depicted in Table 2. The values of PESQ are ranging from 4.5 (the highest possible quality) to 0 (the worst quality) where the output signal i, e., the estimated near-end signal is compared with the original near-end signal.

From the obtained results, we can observe that the proposed method has better performance in terms of the speech intelligibility.

Table 3 presents the obtained probabilities (P_d, P_m and P_f) of the different DTD methods utilizing several tests of speech signals for near-end and far-end speakers, under three SNR levels (15, 35 and 55 dB). It is worth noting that, the goal is to maximize the detection probability P_d and minimize the miss probability P_m for better detection of DT. On the other side, for avoiding the update stopping during the convergence process the false alarm probability P_f should be minimized.

The obtained results evince that the proposed DTD yields lower values of P_m and P_f, also higher value of P_d for the three levels of SNR compared to the others methods. We indicate that the increasing on P_f value causes convergence halting of the NLMS algorithm. We can therefore conclude that the proposed DTD performs better in terms of probabilities evaluation.

Table 3. Comparison of probabilities in function of SNR for different methods of DTD

SNR (dB)	Method	P_d	Pm	P_f
55	Geigel [6]	0.52	0.48	0.25
	ZCR [4]	0.81	0.19	0.27
	NCC [9]	0.92	0.08	0.22
	Proposed	**0.99**	**0.01**	**0.21**
35	Geigel [6]	0.52	0.48	0.31
	ZCR [4]	0.74	0.26	0.28
	NCC [9]	0.81	0.19	0.37
	Proposed	**0.90**	**0.10**	**0.25**
15	Geigel [6]	0.52	0.48	0.43
	ZCR [4]	0.70	0.30	0.38
	NCC [9]	0.80	0.20	0.59
	Proposed	**0.88**	**0.12**	**0.18**

5. CONCLUSION

In this paper, we have proposed a new fast double-talk detector that uses the error signal's variance to detect the presence of near-end signal. As opposed to the existing principle based on comparison between far-end and microphone signals, the key idea behind this work is to focus only on the error signal to declare DT situations.

We have carried out a comparison between Geigel, NCC, ZCR methods and the proposed DTD to show clearly the performance of this latter for acoustic echo cancellation in double-talk scenario. From the obtained results, we can be concluded that the proposed DTD performs better in terms of near-end signal detection and AEC controlling. In addition, this method is simple and it has low computational complexity. In the future we intend to study the effect of echo path change on the performance of the double-talk detector with the use of robust adaptive filtering algorithms to enhance DTD in low level of SNR. Also, we will investigate a dynamique thresolding in the DT decision for providing better performance.

REFERENCES

[1] Benesty, J., Gänsler, T., Morgan, D.R., Sondhi, M.M., Gay, S.L. (2001). Advances in network and acoustic echo cancellation. Digital SignalProcessing. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-04437-7_5

[2] Sondhi, M.M. (2006). The history of echo cancellation. IEEE Signal Processing Magazine, 23(5): 95-102. https://doi.org/10.1109/MSP.2006.1708416

[3] Enzner, G., Buchner, H., Favrot, A., Kuech, F. (2013). Acoustic echo control, in: S. Theodoridis, R. Chellappa (Eds.), Academic Press Library in Signal Processing, 4: 807-877.

[4] Ikram, M.Z. (2015). Double-talk detection in acoustic echo cancellers using zero-crossings rate. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1121-1125. https://doi.org/10.1109/ICASSP.2015.7178144

[5] Muzahid, A.A.M., Ingrid, K.M.R., Mondol, S.R., Zhou, Y. (2016). Advanced double-talk detection algorithm based on joint signal energy and cross-correlation estimation. In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN), pp. 303-306. https://doi.org/10.1109/ICCSN.2016.7586669

[6] Duttweiler, D. (1978). A twelve-channel digital echo canceler. IEEE Transactions on Communications, 26(5): 647-653. https://doi.org/10.1109/TCOM.1978.1094133

[7] Paleologu, C., Benesty, J., Gaensler, T., Ciochină, S. (2011). Class of double-talk detectors based on the holder inequality. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 425-428. https://doi.org/10.1109/ICASSP.2011.5946431

[8] Benesty, J., Morgan, D.R., Cho, J.H. (2000). A new class of doubletalk detectors based on cross-correlation. IEEE Transactions on Speech and Audio Processing, 8(2): 168-172. https://doi.org/10.1109/89.824701

[9] Gänsler, T., Benesty, J. (2006). The fast normalized cross-correlation double-talk detector. Signal Processing, 86(6): 1124-1139. https://doi.org/10.1016/j.sigpro.2005.07.035

[10] Szwoch, G., Czyżewski, A., Kulesza, M. (2008). A low complexity double-talk detector based on the signal envelope. Signal Processing, 88(11): 2856-2862. https://doi.org/10.1016/j.sigpro.2008.05.013

[11] Gansler, T., Hansson, M., Ivarsson, C.J., Salomonsson, G. (1996). A double-talk detector based on coherence. IEEE Transactions on Communications, 44(11): 1421-1427. https://doi.org/10.1109/26.544458

[12] Bao, H., Yang, Y., Liu, J., Bao, X., Yuan, Q. (2010). A robust algorithm of double talk detection based on voice activity detection. In 2010 International Conference on Audio, Language and Image Processing (ICALIP), pp. 12-15. https://doi.org/10.1109/ICALIP.2010.5685026

[13] Lee, K.H., Chang, J.H., Kim, N.S., Kang, S., Kim, Y. (2010). Frequency-domain double-talk detection based on the Gaussian mixture model. IEEE Signal Processing Letters, 17(5): 453-456. https://doi.org/10.1109/LSP.2010.2043891

[14] Rahbar, K. (2012). Double talk detection method based on spectral acoustic properties. U.S. Patent No 8,335,319. Washington, DC: U.S. Patent and Trademark Office.

[15] Low, S.Y., Venkatesh, S., Nordholm, S. (2011). A

spectral slit approach to doubletalk detection. IEEE Transactions on Audio, Speech, and Language Processing, 20(3): 1074-1080. https://doi.org/10.1109/TASL.2011.2168210

[16] Hamidia, M., Amrouche, A. (2017). A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation. Digital Signal Processing, 60: 99-112. https:// doi.org/10.1016/j.dsp.2016.09.001

[17] Szwoch, G., Czyzewski, A., Ciarkowski, A. (2009). A double-talk detector using audio watermarking. Journal of the Audio Engineering Society, 57(11): 916-926.

[18] Urakami, H., Kajikawa, Y. (2010). A double-talk-detector using sound and image information. In 2010 10th International Symposium on Communications and Information Technologies, pp. 447-452. https://doi.org/10.1109/ISCIT.2010.5664883

[19] Hamidia, M., Amrouche, A. (2014). A new structure for acoustic echo cancellation in double-talk scenario using auxiliary filter. In 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 253-257. https://doi.org/10.1109/IWAENC.2014.6954297

[20] Haykin, S.S. (2002). Adaptive filter theory. 4 ed.: Englewood Cliffs, NJ: Prentice-Hall.

[21] Benziane, M., Bouamar, M., Makdir, M. (2020). Simple and Efficient Double-Talk-Detector for Acoustic Echo Cancellation. Traitement du Signal, 37(4): 585-592. https://doi.org/10.18280/ts.370406

[22] Hamidia, M., Amrouche, A. (2019). Improving acoustic echo cancellation in hands-free communication systems. In 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), pp. 1-5. https:// doi.org/10.1109/ISPA48434.2019.8966823

[23] ITU-T P. 862 (2002). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Rec. ITU-T P. 862. https://www.itu.int/rec/T-REC-P.862-200102-I/en.