Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol II,
IMECS 2010, March 17 - 19, 2010, Hong Kong

A Robust Double-Talk Detector for Acoustic Echo Cancellation

Junghsi Lee and Hsu-Chang Huang

***Abstract*—A reliable double-talk detector (DTD) plays an important role in acoustic echo cancellation (AEC). This paper proposes a robust DTD equipped with a near-end voice detector, a double-to-single detector, and two auxiliary filters. The new scheme gives the AEC a better chance of saving good estimate of echo path when entering double-talk mode and properly determining double to single-talk mode. The proposed DTD system responds well in the transition of double-talk and single talk and is capable of differentiating echo path changes from double-talk. Extensive simulations demonstrated that our DTD outperforms existing schemes during double-talk and echo path changes.**

***Index Terms*—Acoustic echo cancellation, double-talk detector, adaptive filter, NLMS.**

I. Introduction

Acoustic echoes are major sources of annoyance in hands-free communications, where the presence of coupling from the far-end signal (loudspeaker) to the near-end signal (microphone) would result in undesired acoustic echo. A reliable acoustic echo canceller (AEC) should include good solutions to the problems of estimating the echo path and double-talk detection. When the near-end speech $s(n)$ and far-end speech $x(n)$ occur simultaneously, the so-called double-talk (DT) mode, the adaptation of the adaptive filter will be severely disturbed by the near-end signal. Therefore, a dependable double-talk detector is required to decide whether it enters DT mode. If so, the AEC has to either slow down or freeze the adaptation of the adaptive filter to prevent it from divergence [3-4, 6-7].

In hands-free communications, the movement of objects or people produce changes in the acoustic echo paths and it has introduced a more difficult problem for double-talk detector (DTD). The DTD might treat the double-talk situation as echo path change. Consequently, the adaptive filter will keep updating and result in the divergence of the system. On the other hand, the DTD may declare an echo-path change as DT mode. It is important for the DTD to efficiently differentiate between echo path change and DT mode.

Numerous schemes have been presented to tackle the DT problem in the past two decades [1-7]. For example, Ye and Wu proposed to use the cross-correlation coefficient vector between the input vector $\mathbf{x}(n)$, and the error signal $e(n)$ for double-talk detection [7]. The error signal is defined as

$$e(n) = y(n) - \hat{d}(n), \tag{1}$$

where $\hat{d}(n)$ is the estimate of the echo,

$$\hat{d}(n) = \mathbf{w}^T(n)\mathbf{x}(n), \tag{2}$$

and $y(n)$ is the signal received by the microphone. This microphone signal can be written as

$$y(n) = s(n) + v(n) + d(n), \tag{3}$$

where $s(n)$ is the near-end speech, $v(n)$ is the noise in the near-end, and $d(n)$ is the acoustic echo

$$d(n) = \mathbf{h}^T(n)\mathbf{x}(n). \tag{4}$$

Benesty, Morgan and Cho [3] utilized the cross-correlation coefficient vector between $\mathbf{x}(n)$ and $y(n)$ for double-talk detection. The idea is to compare the normalized cross-correlation coefficient ξ_1 to a threshold level, and the DT mode is declared if ξ_1 is smaller than the threshold. Later, Park [6] argued that appropriate thresholds are hardly available in real environments, and proposed a DTD that uses two cross-correlations, $\rho_{\hat{d},y}(n)$ and $\rho_{e,y}(n)$. Park employed these two cross-correlations as indicators for DT and developed an acoustic echo and noise canceller.

This paper extends two-filter structure of our previous work [4] to introduce a new robust DTD that includes a near-end voice detector (NEVD), a double-to-single detector (DSD), two auxiliary filters, and new decision rules of the DTD. The proposed DTD performs well during double-talk stage in an echo path changing scenario. The superiority of our method was verified by extensive simulations.

II. The Proposed DTD

The block diagram of an acoustic echo canceller equipped with the proposed DTD is illustrated in Fig. 1. Our DTD includes a NEVD, a DSD and two auxiliary filters. The roles of NEVD and DSD are to control the adaptive filter either working or halting. The auxiliary filters are to save good estimates of echo path and to prevent the AEC from divergence.

A. NEVD indicator $\xi(n)$

We use an NEVD indicator $\xi(n)$ defined as

$$\xi(n) = \sqrt{\sigma_{\hat{d}}^2(n) / \sigma_y^2(n)}, \tag{5}$$

where $\sigma_{\hat{d}}^2(n)$ and $\sigma_y^2(n)$ are estimated variance of $\hat{d}(n)$ and $y(n)$, respectively. They are calculated recursively as

Manuscript received Jan. 12, 2010. This work was supported by the TAIWAN National Science Council under grant NSC 98-2221-E-155-037.

J. Lee and H. C. Huang are with the Department of Electrical Engineering, Yuan-Ze University, Taoyuan, 32026, TAIWAN (Phone: +886-3-4638800, Ext.7116; Fax:+886-3-4639355; E-Mail:eejlee@saturn.yzu.edu.tw).

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$\sigma_{\hat{d}}^2(n) = \beta\sigma_{\hat{d}}^2(n) + (1-\beta)\hat{d}^2(n), \tag{6}$$

and

$$\sigma_y^2(n) = \beta\sigma_y^2(n) + (1-\beta)y^2(n), \tag{7}$$

where β is a smoothing factor, $0 < \beta < 1$. The idea is to compare $\xi(n)$ with a preset threshold T_1 to determine whether near-end speech $s(n)$ occurs.

B. DSD indicators $\zeta(k)$

One major problem encountered in AEC is to determine whether it really leaves DT mode and enters single-talk (ST) mode. If AEC declares ST incorrectly, the adaptive filter would be forced to update coefficients in DT stage. This situation might occur in weak speech or short silence periods of near-end speech, and would result in diverged adaptive filter. This paper proposes a new scheme with DSD indicators to alleviate this problem. The DSD indicators $\zeta(k)$, $k = 1, 2, \cdots, J$, are reset to zero when AEC declares DT mode from ST mode. After that, the DSD indicators are updated for every L signal samples as follows. Firstly, the value of $\zeta(k)$ is copied to $\zeta(k-1)$ for $k = 2, 3, \cdots, J$. Then, $\zeta(J)$ is calculated as the mean of the L latest NEVD indicators.

C. Decision rules of the proposed DTD s

The decision rules of the proposed DTD are as follows. Set a DT status parameter $S_{DT}(n) = 1$ if it is in the DT mode, and $S_{DT}(n) = 0$ for the ST mode. At sample n, assuming $S_{DT}(n-1) = 0$, $\xi(n)$ is calculated and compared with a threshold T_1. If $\xi(n)$ is bigger than T_1, the system is still in the ST mode. Otherwise, the system declares entering the DT mode and set $S_{DT}(n) = 1$. Now consider the situation that $S_{DT}(n-1) = 1$. If all DSD indicators $\zeta(k)$ are bigger than a preset threshold T_2, the system declares ST mode and set $S_{DT}(n) = 0$. Otherwise, the system is still in the DT mode.

D. Adaptive Filter

We use a modified NLMS algorithm in this AEC system. The coefficients $\mathbf{w}(n)$ are updated as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu_c}{\mathbf{x}^T(n)\mathbf{x}(n) + M \cdot \hat{\sigma}_x^2}\mathbf{x}(n)e(n), \tag{8}$$

where μ_c is a fixed step size, M is a positive constant, and $\hat{\sigma}_x^2$ is an estimate of the variance of $x(n)$.

E. Auxiliary Filters

AEC uses two auxiliary filters: AF_1 and AF_2, both filters simply store echo path estimates for later usage. The idea is as below. If $S_{DT}(n) = 0$, it is in the ST mode, the adaptive filter keeps updating the coefficients and the counter C_{AF} increases by one. Once C_{AF} reaches a preset threshold T_{AF}, copy the coefficients of AF_1 to AF_2 first, then copy the coefficients of adaptive filter to AF_1, and reset counter C_{AF} to zero. If $S_{DT}(n) = 1$, it is in the DT mode, AEC freezes adaptation of adaptive filter, copy the coefficients of AF_2 to AF_1 and to the adaptive filter as well. The role of AF_2 is to provide the frozen-adaptive filter "right coefficients" in DT stage so that it can produce good estimated echo $\hat{d}(n)$, and AF_1 is used as a buffer between adaptive filter and AF_2 so that the AEC would have a higher probability of saving good acoustic echo path during the transition from single talk to double talk.

III. Simulation Results

In this section, we present the results of several experiments that demonstrate the efficiency of our DTD. The adaptive filter is used to identify a 512-tap acoustic echo impulse response. The acoustic echo path was measured at 8000 Hz sampling rate in a small office. The excitation signal is a 16-second-long Chinese speech. White Gaussian noise with SNR 39dB is added to the acoustic echo. The adaptive filter is run with as many taps as the echo path. We compare the performance of our DTD to Benesty's DTD using the normalized cross-correlation (NCC), ξ and Park's DTD using two cross-correlations, $\rho_{\hat{d},y}(n)$ and $\rho_{e,y}(n)$. The thresholds of ξ_1, $\rho_{\hat{d},y}(n)$,and $\rho_{e,y}(n)$ are chosen as 0.9, 0.8, and 0.35, respectively. We set $L = 512$, $J = 4$, $T_{AF} = 1000$, $T_1 = 0.7$, $T_2 = 0.95$, $\mu_c = 0.4$, and $M = 512$. The echo return loss enhancement (ERLE) and the normalized squared coefficient error (NSCE) are defined as

$$ERLE(n) = 10\log_{10}\frac{\left(y(n)-s(n)\right)^2}{\left(e(n)-s(n)\right)^2}, \tag{9}$$

and

$$NSCE(n) = 10\log_{10}\frac{\|\mathbf{h}(n)-\mathbf{w}(n)\|^2}{\|\mathbf{h}(n)\|^2}, \tag{10}$$

respectively. Note that we have included $s(n)$ in (9) so as to better evaluate the performance of AEC during DT stages.

The first experiment has DT periods from 11 to 14 seconds. Fig. 2 shows the far-end speech $x(n)$. The acoustic echo impulse response is fixed for all times except from time 5.3 to 6.3 which is described by

$$\mathbf{h}(n) = \mathbf{h}_o + g(n), \tag{11}$$

where $g(n)$ is a white Gaussian noise with variance 10^{-5}. Simulation results are illustrated in Figs. 3-5. Fig. 3 shows that the Benesty's method makes a wrong decision to declare DT mode when echo-path changes. Fortunately, this does not cause any extra ERLE loss comparing to other methods. During DT stage, Park's method declares ST status during 11.9 to 12.05. Note that there is a slight decision delay occurring at the beginning of DT stage for all three methods. As a result, the error increases more at the beginning of DT mode and all methods have an NSCE drop by 25dB as demonstrated in Fig. 4. Roughly, the same amount of ERLE loss can be seen in Fig. 5. However, our AEC efficiently recovers to good NSCE and ERLE in less than 1 second while the other two methods perform badly during DT phase. The good performance of our DTD is mainly due to the employment of auxiliary filters. It can be seen from Fig. 5 that our method is about 30dB ERLE better than the other

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

two methods during DT phase.

We conduct another experiment involving abrupt echo path changes: $\mathbf{h}(n)$ is circularly shifted by 200 samples at time 5.3. The DT phase is set from 11 to 14 seconds. Simulation results are illustrated in Figs. 6-8. Fig. 6 shows that Benesty's scheme declares DT at time 5.3, which is actually the echo path change. The damage is obvious: the AEC freezes the adaptation thereafter. Park's method does not perform well during DT stage. The NSCE and ERLE curves demonstrate that the proposed method outperforms other two algorithms at DT stage.

IV. Conclusions

In this paper, we presented a robust DTD equipped with a near end voice detector, a double-to-single detector, two auxiliary filters, and new decision rules of the DTD. The proposed scheme gives the AEC a better chance of saving good estimate of echo path. As a result, our DTD system performs well in the transition of double-talk and single-talk and is capable of differentiating echo path changes from double-talk situation. Extensive simulations demonstrated that our DTD outperforms other existing schemes during double-talk phase and echo path changes.

References

[1] P. Ahgren and A. Jakobsson, "A study of doubletalk detection performance in the presence of acoustic echo path changes," in *Proc. IEEE ICASSP, 2005*, vol. III, pp. 141-144.

[2] J. Benesty and T. Gansler, "A multidelay double-talk detector combined with the MDF adaptive filter," *EURASIP J. Applied Signal Process.*, pp. 1056-1063, 2003.

[3] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 2, pp. 168-172, Mar. 2000.

[4] H. C. Huang and J. Lee, "A new double-talk detection using two auxiliary filters," in *Proc. 4th IASTED Int. Conf. Circuits, Signals, and Systems, 2006*, 269-272.

[5] J. C. Jeng and S. F. Hsieh, "Decision of double-talk and time-variant echo path for acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 317-319, Nov. 2003.

[6] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, "Integrated echo and noise canceller for hands-free applications," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 49, no. 3, pp. 188-195, Mar. 2002.

[7] H. Ye and B. X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Commun.*, vol. 39, pp. 1542-1545, Nov. 1991.

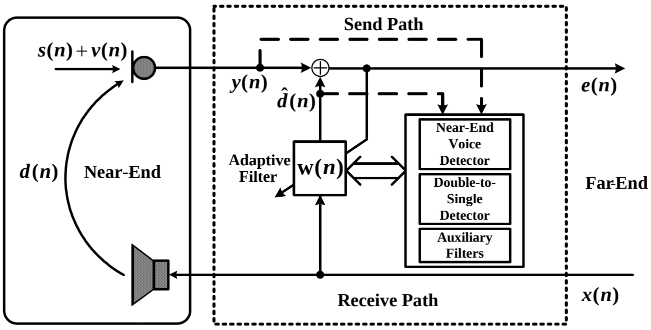

Fig. 1, Basic structure of acoustic echo cancellation with double-talk detection.

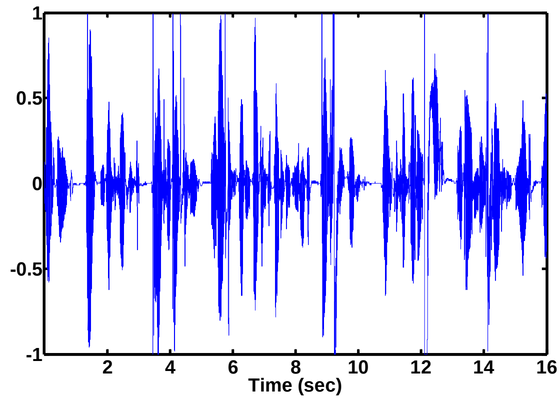

Fig. 2, Far-end speech used in the simulations.

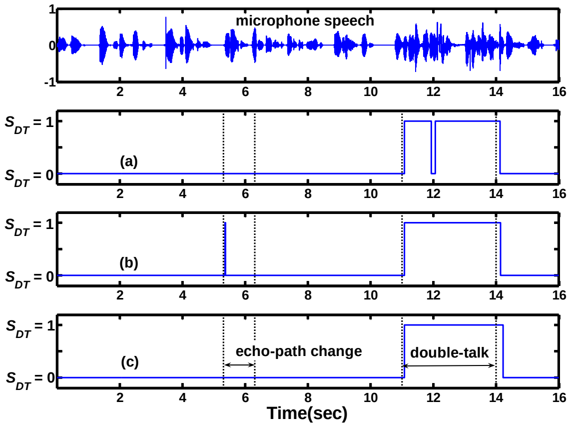

Fig. 3, DTD decision results of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

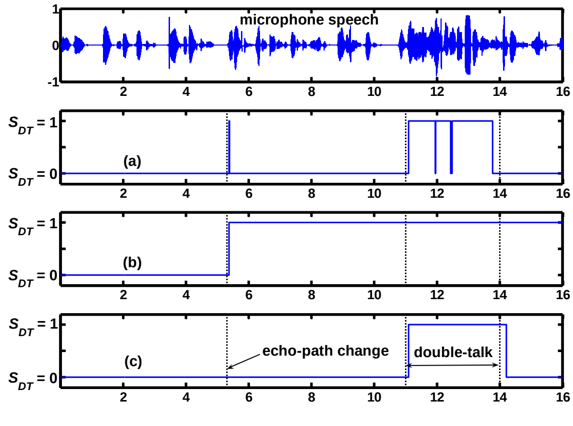

Fig. 6, DTD decision results of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

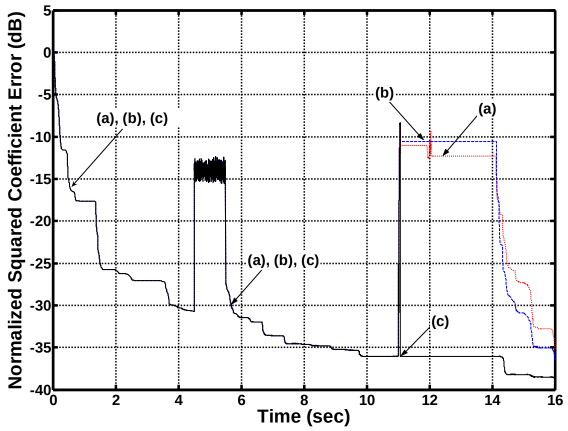

Fig. 4, NSCE curves of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

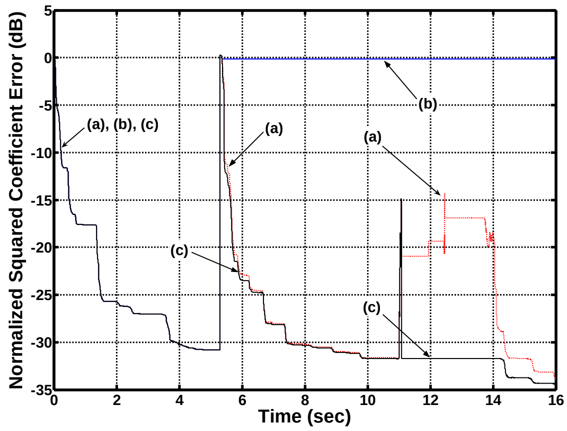

Fig. 7, NSCE curves of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

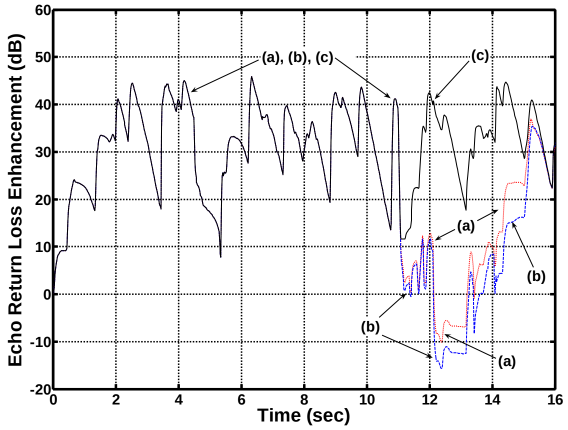

Fig. 5, ERLE curves of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

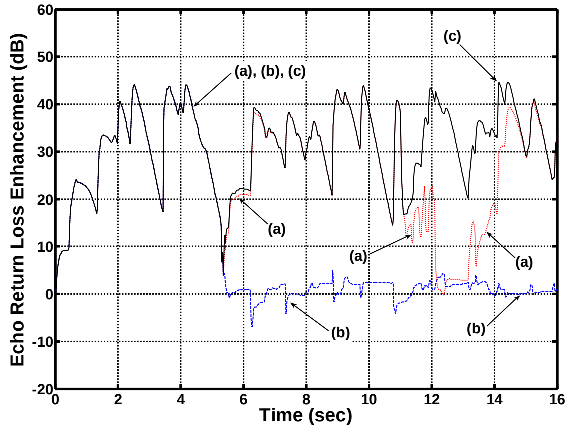

Fig. 8, ERLE curves of (a) Park's method, (b) Benesty's method, and (c) our proposed method.

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)