DELAYED ADAPTATION FOR IMPROVED DOUBLETALK RESILIENCE IN ADAPTIVE ECHO CANCELLERS

*James D. Gordy**
School of Info. Tech. and Engg.
University of Ottawa, Ottawa, Canada
jdgordy@site.uottawa.ca

Franck Beaucoup†
Broadcom Corp.
Richmond, BC, Canada
franckb@broadcom.com

Rafik A. Goubran
Dept. of Systems and Comp. Engg.
Carleton University, Ottawa, Canada
goubran@sce.carleton.ca

ABSTRACT

This paper proposes a simple modification to adaptive echo cancellers to prevent rapid divergence due to doubletalk conditions. The proposed structure introduces a delay only into the adaptation algorithm, which compensates for doubletalk detector response time at the onset of near-end speech. The structure is described and analyzed for the NLMS and cross-correlation-based doubletalk detector algorithms. The method introduces no delay into the return path, and no additional constraints on adaptation step size arise. Simulations with ITU-T G.168 tests confirm that the proposed structure prevents divergence at the onset of doubletalk for near-end to echo signal ratios of -26 to -6 dB.

Index Terms— *adaptive echo cancellation, delayed adaptation, doubletalk detection, NLMS*

1. INTRODUCTION

Adaptive echo cancellers have been used for several decades to cancel network echoes generated by hybrid transformers, and for acoustic echoes in hands-free terminals [1]. Longer round-trip delays introduced by parametric speech coding and VoIP networks exacerbate user perception of echo and increase the performance requirements of echo cancellers [2]. Adaptation algorithms such as normalized least-mean-square (NLMS) assume only echo and low background noise in the reference signal [3]. As shown in Fig. 1, a practical problem is doubletalk, or the simultaneous presence of near-end speech. Therefore, typical echo canceller implementations employ a doubletalk detector to halt adaptation during such periods to avoid rapid divergence of the adaptation algorithm. Fast detection at the onset of near-end speech is a key requirement, but accuracy is also desirable to avoid false positives that may affect echo canceller convergence and tracking.

Doubletalk detectors have been proposed based on measures of energy, cross-correlation, frequency-domain coherence, and robust statistics [4] – [8]. In general, algorithms that employ time averaging to estimate detection statistics offer improved reliability, but at a cost of latency in response time. As a result, divergence can still occur at the onset of doubletalk. Previous solutions insert delays into the reference or error signals to allow time for the doubletalk detector to react [9], [10]. However, side effects are severe limits on adaptation step size and delay introduced to the return-path signal. "Tap-rotation" algorithms introduce delay by maintaining previous sets of adaptive filter coefficients. In particular, a dual-filter solution was proposed in [11] employing a fixed foreground filter for echo cancellation and a constantly adapting background filter. The latter is brought to the fore when its echo power reduction improves over the fixed filter, which naturally guards against doubletalk at a cost of increased storage and computational requirements.

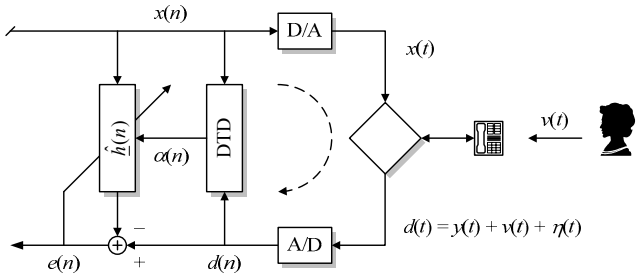

Fig. 1 – Block diagram of a typical echo canceller and doubletalk detector (DTD) at an analog hybrid transformer.

This paper proposes a simple modification to adaptation algorithms to increase their immunity to divergence at the onset of doubletalk. The method improves upon [9] – [11] in that it introduces no additional constraint on step size, no delay into the return-path signal, and a minimal increase in complexity. Section 2 reviews echo canceller structures and the problem of doubletalk detector response time. The proposed structure is described in Section 3 for NLMS and the cross-correlation-based doubletalk detector of [5], with simulation results presented in Section 4.

2. REVIEW OF ECHO CANCELLATION AND DOUBLETALK DETECTION

2.1. Echo Canceller Structure and Conventions

Fig. 1 shows a typical echo canceller and doubletalk detector at a hybrid transformer. The input signal $x(n)$ is sent to the hybrid at the near end, and the resulting reference signal $d(n)$ consists of echo $y(n)$, near-end speech $v(n)$, and background noise $\eta(n)$. The echo canceller models and tracks the echo path as an N-sample finite impulse response. The output $e(n)$ is obtained by subtracting the estimated echo from the reference signal:

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

$$e(n) = d(n) - \underline{x}^T(n)\hat{\underline{h}}(n) \tag{2}$$

where $\underline{x}(n) = [x(n) \quad x(n-1) \quad \cdots \quad x(n-N+1)]^T$ is the $N \times 1$ tap-input vector, and $\hat{\underline{h}}(n) = [\hat{h}_0(n) \quad \hat{h}_1(n) \quad \cdots \quad \hat{h}_{N-1}(n)]^T$ is the $N \times 1$ adaptive filter coefficient vector at time n. It is assumed that coefficients are updated using NLMS [3]:

---

* This work was performed while at Carleton University.

† The author was previously with Mitel Networks, and this paper is the subject of a patent application filed by Mitel.

$$\hat{\underline{h}}(n+1) = \hat{\underline{h}}(n) + \mu \frac{\underline{x}(n)e(n)}{\|\underline{x}(n)\|^2 + \delta} \quad (3)$$

where $0 \le \mu \le 2$ is the step size parameter, δ is a small regularization parameter, and $\|\cdot\|$ denotes the l^2 norm.

2.2. Cross-Correlation-Based Doubletalk Detection

For a stationary input signal, the expected echo signal variance can be written in terms of $\underline{R}_{xx}$, the $N \times N$ input signal autocorrelation matrix, and $\underline{r}_{xd}$, the $N \times 1$ cross-correlation vector between the input and reference signals. From this representation, a normalized detection statistic ξ was proposed in [5] as the ratio of expected to measured reference signal variances:

$$\sigma_y^2 = E\{y^2(n)\} = \underline{h}^T \underline{R}_{xx} \underline{h} = \underline{r}_{xd}^T \underline{R}_{xx}^{-1} \underline{r}_{xd} \quad (4)$$

$$\xi = \sqrt{\sigma_y^2 / \sigma_d^2} = \sqrt{\underline{r}_{xd}^T \underline{R}_{xx}^{-1} \underline{r}_{xd} / \sigma_d^2} \quad (5)$$

When doubletalk is absent, the numerator and denominator terms are approximately equal and $\xi = 1$. When doubletalk is present, the denominator will increase and $\xi < 1$. Practical simplifications are obtained by assuming the adaptive filter has converged, and by estimating parameters over a window of K samples [5]:

$$\underline{R}_{xx}^{-1} \underline{r}_{xd} = \underline{h} \approx \hat{\underline{h}}(n) \quad (6)$$

$$\hat{\underline{r}}_{xd}(n) = \frac{1}{K} \sum_{k=0}^{K-1} \underline{x}(n-k) d(n-k) \quad (7)$$

$$\hat{\sigma}_d^2(n) = \frac{1}{K-1} \sum_{k=0}^{K-1} [d(n-k) - \frac{1}{K} \sum_{j=0}^{K-1} d(n-j)]^2 \quad (8)$$

Substituting (6) – (8) into (5) results in an estimated doubletalk detection statistic at time n:

$$\xi(n) = \sqrt{\hat{\underline{r}}_{xd}^T(n) \hat{\underline{h}}(n) / \hat{\sigma}_d^2(n)} \quad (9)$$

A decision is made by comparing $\xi(n)$ to a threshold T. If $\xi(n) < T$, adaptation is slowed or halted by applying a scaling factor $0 \le \alpha(n) \le 1$ to the step size parameter. The threshold may be chosen using empirical or statistical calibration techniques [7], [12].

2.3. Doubletalk Detector Response Time

As the window size K increases for the estimators of (6) – (8), for stationary signals the accuracy of (9) increases in terms of probabilities of false alarm (P_F) and miss (P_M). However, in practice $\xi(n)$ will not fall below T until a number of samples after the onset of near-end speech. Fig. 2 shows the average response time in samples as a function of near-end to echo signal power ratio (NER) for $K = 50$, 100, and 200 samples, along with the 95% confidence intervals. It was obtained by averaging over 50 pairs of input and doubletalk signals from the TIMIT speech database downsampled to f_s = 8 kHz, and for a threshold T chosen to provide of $P_F \le 0.1$ [7]. The response time increases with K and with decreasing NER, so there is a tradeoff incurred by improving accuracy by increasing the estimation window size. The response time is in the range of 5 – 40 samples (at f_s = 8 kHz), indicating the need for a mechanism to avoid divergence of the echo canceller until the detection threshold is reached.

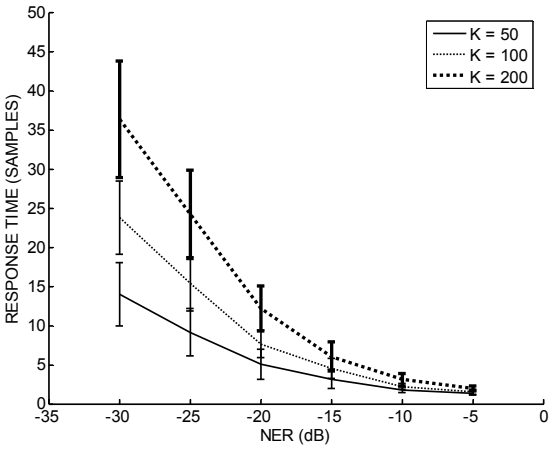

Fig. 2 – Doubletalk detector response time as a function of NER with $P_F \le 0.1$ (95% C.I.).

3. DELAYED ADAPTATION OF NLMS-TYPE ALGORITHMS

3.1. Algorithm Description

From the results of Section 2.3, it is natural to introduce some look-ahead processing for the doubletalk detector, ideally without increasing the return-path signal delay and therefore the round-trip time [2]. In particular, it is desirable to delay the echo canceller's adaptation by, say, D samples with respect to the doubletalk detector. If D is at least as long as the average response time, then this delay will increase the probability that adaptation will be halted before the adaptive filter diverges. One approach to delaying adaptation is to delay the error signal $e(n)$ fed back to the echo canceller's adaptation algorithm in (3), which corresponds to the well-known delayed LMS algorithm [9]. In [10] an explicit stability bound is given as a function of D, and is shown to be smaller than that of LMS [9]. For delays on the order of 5 – 40 samples, the stability bound is far too small for practical echo canceller implementations.

An alternative approach to delaying adaptation is proposed as follows, where $\hat{\underline{h}}_D(n)$ now represents the adaptive filter coefficient vector at time $n \ge 0$. The return-path error signal $e(n)$ is still constructed in accordance with (2). However, adaptation of the filter coefficients is modified to employ versions of the input and reference signals that are both delayed by D samples. In addition, the error signal fed back to the adaptation algorithm is constructed using these delayed input and reference signals:

$$e(n) = d(n) - \underline{x}^T(n) \hat{\underline{h}}_D(n) \quad (10)$$

$$e_D(n) = d(n-D) - \underline{x}^T(n-D) \hat{\underline{h}}_D(n) \quad (11)$$

$$\hat{\underline{h}}_D(n+1) = \hat{\underline{h}}_D(n) + \mu \frac{\underline{x}(n-D) e_D(n)}{\|\underline{x}(n-D)\|^2 + \delta} \quad (12)$$

Fig. 3 shows a block diagram of a hybrid echo canceller employing

the proposed delayed adaptation structure described above. Note that two filtering operations per sample are required, but only a single adaptive filter coefficient vector is maintained, and no additional signal delay is introduced into the return path.

3.2. Adaptation and Complexity Analysis

A simple inductive proof can be used to show that the proposed algorithm of (10) – (12) introduces an adaptation delay of exactly D samples compared to $\hat{\underline{h}}(n)$, the coefficient vector produced by NLMS in (3). Assuming $\hat{\underline{h}}_D(n) = \hat{\underline{h}}(0)$ and $x(n) = 0$ for $n < 0$, note that (12) allows no adaptation for the first D samples:

$$\hat{\underline{h}}_D(n) = \hat{\underline{h}}(n-D), \quad n = D \tag{13}$$

Now assume (13) holds at some time $n > D$. Substituting (13) into (11) and (12) reveals that the adaptive filter coefficient vector at time n+1 is equal to the coefficient vector of (3) at time n–D+1:

$$e_D(n) = d(n-D) - \underline{x}^T(n-D)\hat{\underline{h}}(n-D) = e(n-D) \tag{14}$$

$$\hat{\underline{h}}_D(n+1) = \hat{\underline{h}}(n-D) + \mu \frac{\underline{x}(n-D)e(n-D)}{\left\|\underline{x}(n-D)\right\|^2 + \delta} = \hat{\underline{h}}(n-D+1) \tag{15}$$

By induction it follows that (11) and (12) produce an adaptive filter coefficient vector delayed by exactly D samples, or $\hat{\underline{h}}_D(n) = \hat{\underline{h}}(n-D)$, which in turn implies that the error signal in (10) is constructed using $\hat{\underline{h}}(n)$ delayed by D samples. This delay is achieved *without* requiring storage for $D - 1$ previous sets of adaptive filter coefficient vectors. In contrast, tap-rotation-like algorithms may require storage of multiple $N \times 1$ vectors to achieve a similar effect. Finally, an important corollary of the above result is that no additional constraint on the adaptation step size parameter is introduced compared to regular NLMS.

NLMS requires one filtering operation to obtain $e(n)$ and an update for each of the filter coefficients, for a total of approximately $2N$ operations per sample period. Two buffers of N samples are required to hold the tap-input and adaptive filter coefficient vectors. From (10) – (12), the proposed algorithm requires a second filtering operation (N operations) to obtain the delayed error signal $e_D(n)$. The tap-input buffer must be increased from N to $N + D$ samples, and a buffer of D samples added to hold the delayed reference signal $d(n - D)$. For many echo cancellers, particularly in acoustic environments, these increases are marginal given the results of Fig. 2 ($N >> D$). In comparison, the dual-filter approach of [11] involves a second filtering operation as well, but requires N additional coefficients of storage. The proposed algorithm may require the presence of a more complex doubletalk detector than the dual-filter approach, leading to a trade-off between storage and algorithmic complexity for the two structures.

4. SIMULATION RESULTS

4.1. Simulation Setup

The proposed delayed adaptation algorithm of (10) – (12), denoted DA-NLMS, was compared to regular NLMS using the same

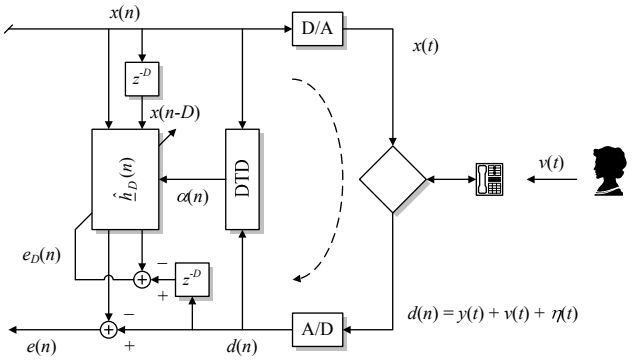

Fig. 3 – Proposed delayed adaptation structure employing a single adaptive filter and two filter operations per sample.

doubletalk detector, and assessed using conformance tests from ITU-T G.168 [13]. Performance was measured as combined signal power reduction in decibels (A_{COM}) due to echo return loss (A_{ECHO}) and cancellation achieved by the adaptive filter (A_{CANC}). ITU-T G.168 specifies hybrid impulse responses and input and doubletalk composite source signals (CSS) possessing speech-like power spectra, which the reader will find in [13]. In particular, these simulations employed echo paths 5 and 7 calibrated for $A_{ECHO} = 6$ dB, each consisting of $N = 96$ samples. The input CSS signal was set to a level of 0 dBm0 at a sampling rate of $f_s = 8$ kHz, with relative adjustment of doubletalk CSS level. Both algorithms employed parameters of $\mu = ⅓$, $\delta = 10^{-5}$ and, for delayed adaptation, $D = 40$. For both algorithms, the doubletalk detector of [5] employed an estimation window of $K = 200$ samples and a detection threshold calibrated for $P_F \le 0.1$ [7].

4.2. Results and Discussion

ITU-T G.168 Test 2B evaluates convergence and re-convergence using only input CSS (no doubletalk). For input CSS at a level of 0 dBm0, the test requires $A_{COM} \ge 20$ dB after one second of adaptation, and $A_{COM} \ge 30$ dB after ten seconds. In this experiment, convergence was assessed by applying input CSS to echo path 5, switching to echo path 7 after ten seconds of adaptation. Fig. 4(a) shows A_{COM} for both configurations, along with the minimum level of cancellation required by ITU-T G.168. Fig. 4(b) shows the system distance (error norm) between the true and estimated echo paths during one second of adaptation. It is clear from these figures that employing DA-NLMS has no effect on A_{COM}, which exceeds the test requirements for both configurations. A closer look at Fig. 4(b) reveals that adaptation is delayed by exactly $D = 40$ samples, as expected from Section 3.2. This amount (5 ms) is negligible compared to the overall convergence time requirements of the test.

ITU-T G.168 Test 3B evaluates echo canceller performance in doubletalk under both high and low near-end speech conditions. After convergence, doubletalk is applied for Δ seconds, after which adaptation is halted and A_{COM} is measured during singletalk. For input CSS at a level of 0 dBm0, the test requires $A_{COM} \ge 20$ dB after doubletalk CSS greater than or equal to 0 dBm0, and $A_{COM} \ge 27$ dB after doubletalk CSS of -30 to -6 dBm0. In this experiment the echo canceller was allowed to converge for three seconds using echo path 7, after which doubletalk CSS was applied at levels of 0 and -20 dBm0 (-6 and -26 dB NER, respectively) for $\Delta = 2$ seconds. Fig. 5 and Fig. 6 show A_{COM} and system distance for the two doubletalk signal levels, respectively, along with the minimum

level of cancellation required during the experiments. For loud near-end speech (-6 dB NER), the doubletalk detector for both NLMS and DA-NLMS was able to detect doubletalk conditions and halt adaptation, meeting the test requirements with $A_{COM} \geq 20$ dB (by a small margin for NLMS). For lower near-end speech (-26 dB NER), NLMS allowed significant divergence to occur at near-end speech onsets prior to doubletalk being declared, resulting in $A_{COM} \approx 20$ dB after doubletalk, far below the 27 dB required by ITU-T G.168. In contrast, DA-NLMS proved resilient to divergence at the onset of near-end speech, producing $A_{COM} \approx 40$ dB for both doubletalk signal levels, well above the requirements.

5. CONCLUSIONS

A modification was proposed to improve echo canceller resilience to divergence at the onset of doubletalk. Simulations showed that the structure mitigates the problem of response time, while introducing neither degradation of convergence rate nor additional signal delay. Though presented for network echo, the technique is also applicable to acoustic echo cancellers. Further work must be done to compare its performance with tap-rotation-like algorithms.

REFERENCES

[1] S. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunications*. Norwell, MA: Kluwer, 2000.

[2] International Telecommunication Union, *ITU-T G.131: Talker echo and its control*, ITU 2003.

[3] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1996.

[4] D. L. Duttweiler, "A twelve-channel digital echo canceller," *IEEE Trans. Commun.*, vol. 26, pp. 647–653, May 1978.

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

[6] T. Gänsler *et al.*, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, vol. 44, no. 11, pp. 1421–1427, Nov. 1996.

[7] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancellers," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 718–724, Nov. 1999.

[8] T. Gänsler *et al.*, "Double-talk robust fast converging algorithms for network echo cancellation", *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp.656–663, Nov. 2000.

[9] G. Long, F. Ling, and J. G. Proakis, "The LMS algorithm with delayed coefficient adaptation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 9, pp. 1397–1405, Sep. 1989.

[10] P. Kabal, "The stability of adaptive minimum mean square error equalizers using delayed adaptation," *IEEE Trans. Commun.*, vol. 31, no. 3, pp. 430–432, Mar. 1983.

[11] K. Ochiai *et al.*, "Echo canceler with two echo path models," *IEEE Trans. Commun.*, vol. 25, no. 6, pp. 589–595, Jun. 1977.

[12] J. D. Gordy and R. A. Goubran, "Statistical analysis of doubletalk detection for calibration and performance evaluation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1035–1043, Mar. 2007.

[13] International Telecommunication Union, *ITU-T G.168: Digital network echo cancellers*, ITU 2004.

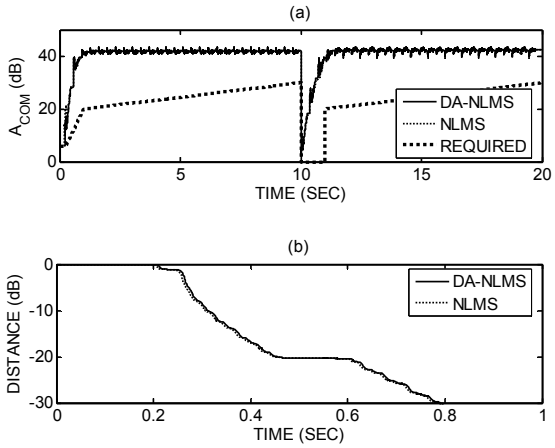

Fig. 4 – ITU-T G.168 Test 2B; (a) measured and required A_{COM}; (b) system distance during one second of adaptation.

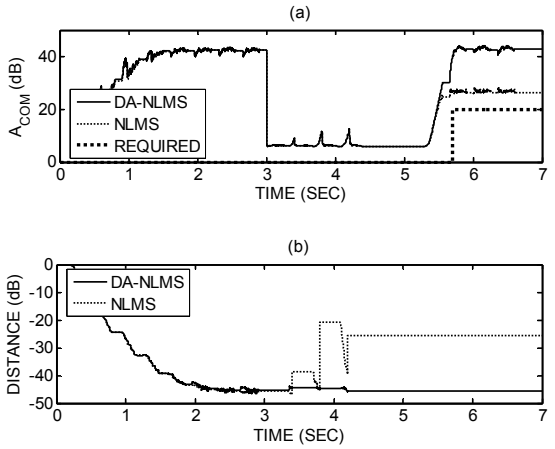

Fig. 5 – ITU-T G.168 Test 3B (NER = -6 dB); (a) measured and required A_{COM}; (b) system distance during and after doubletalk.

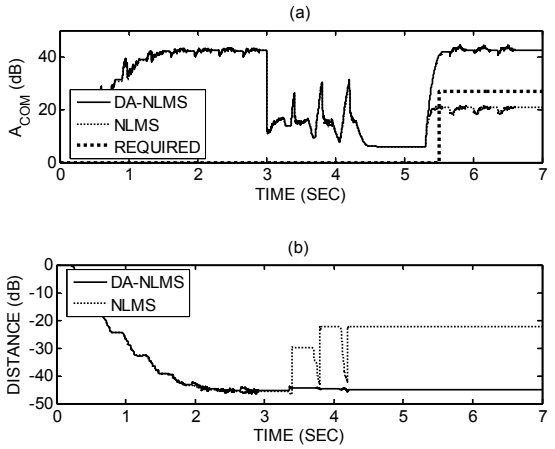

Fig. 6 – ITU-T G.168 Test 3B (NER = -26 dB); (a) measured and required A_{COM}; (b) system distance during and after doubletalk.