IMPROVED DECISION LOGIC FOR TWO-PATH ECHO CANCELERS

Eric J. Diethorn

Agere Systems[1], Research
600 Mountain Avenue, Murray Hill, NJ, 07974-0636
eric.diethorn@agere.com

ABSTRACT

Two-path echo cancelers are distinguished by their use of two digital filters, only one of which is adaptive. The other filter is nonadaptive, but is periodically updated with the coefficients of the adaptive filter. In this work, a new decision method is described for governing the filter coefficient transfer operation that is basic to all echo cancelers employing the two-path structure. This new decision logic differs from that of prior works in that it does not use decision thresholds (constants). Moreover, the new logic applies to both lossy and gain-incurring echo paths, and possesses favorable convergence properties for many scenarios encountered in practice.

1 INTRODUCTION

The design of effective doubletalk detectors is recognized as the most difficult aspect of any real-world application of speech echo cancelers. For conventional single-path cancelers used in network echo applications, the popular method of A. A. Geigel [1] is widely used for doubletalk detection. Geigel's method uses signal-power-level comparison tests to determine when doubletalk or any near-side disturbance of significant power is present. If doubletalk is detected, filter adaptation is inhibited. Signal-level-based doubletalk detectors are, however, notoriously poor at detecting near-side disturbances. Additionally, level-based doubletalk detectors cannot, in general, be used in applications where the echo path induces signal gain (loud echo).

A two-path echo canceler (Fig. 1) is distinguished by its use of two digital filters: a quasi-static foreground filter and an adaptive background filter. The primary benefit of the two-path canceler structure lies in its ability to perform very well in the presence of doubletalk. First introduced some twenty-five years ago by Ochiai, Araseki and Ogihara [2], the two-path echo canceler offers a simple, yet elegant solution to the problem of doubletalk detection. Because the output of the background adaptive filter is not in the audio path, temporary degradation of its coefficients does not directly affect the performance of the foreground canceler.

The most complicated aspect of the two-path structure is the design of decision logic used to determine when the background filter coefficients should be copied to the foreground filter. The original logic described in [2] relies on several user-selected constants, including thresholds and timers. Furthermore, because this logic incorporates signal-level comparison tests for doubletalk detection, it does not apply to applications where the echo path induces signal gain, a common condition in acoustic echo cancellation applications.

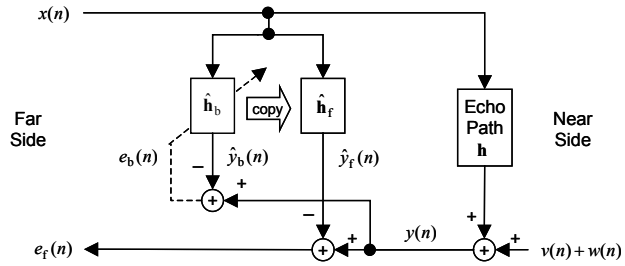

Figure 1. Two-path echo canceler.

This work describes new decision logic for the two-path structure. This new logic differs from that of prior works in that it is devoid of decision thresholds. Moreover, the new logic applies to both lossy and gain-incurring echo paths.

2 TWO-PATH ECHO CANCELER

2.1 Echo Canceler Structure

Figure 1 shows a signal-flow diagram of the classic two-path echo canceler. The physical echo path of interest may be electrical (network) or acoustic. At each sample time *n,* the estimation error between the foreground filter output and the *near-side* signal *y(n)* is

$$e_{\mathrm{f}}(n) = y(n) - \hat{y}_{\mathrm{f}}(n)\,, \tag{1}$$

where

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

is the foreground filter output,

$$\mathbf{x}(n) = [x(n), x(n-1),..., x(n-N+1)]^{T} \tag{3}$$

is the length-*N* history of the receive signal, or *far-side* input signal, and

$$\hat{\mathbf{h}}_{\mathrm{f}}(n) = [\hat{h}_{\mathrm{f},0}(n), \hat{h}_{\mathrm{f},1}(n),..., \hat{h}_{\mathrm{f},N-1}(n)]^{T} \tag{4}$$

is the vector of foreground filter coefficients at time *n. y(n)* is the combination of the physical echo, the near-side speech *v(n)* and the near-side background noise *w(n)*:

$$y(n) = \mathbf{h}^{T}\mathbf{x}(n) + v(n) + w(n)\,, \tag{5}$$

where **h** is the true echo path (for discussion also length-*N).*

A second filter, background filter $\hat{\mathbf{h}}_{\mathrm{b}}(n)$, accepts the same two inputs as the foreground filter and produces a like error signal

$$e_b(n) = y(n) - \hat{y}_b(n) \tag{6}$$

but this error signal is used only for control. Filter $\hat{\mathbf{h}}_{\mathrm{b}}(n)$ is a continuously adapting filter, while $\hat{\mathbf{h}}_{\mathrm{f}}(n)$ is quasi-static, changing only when $\hat{\mathbf{h}}_{\mathrm{b}}(n)$ is determined to be better at

[1] Formerly, the Microelectronics and Communications Division of Lucent Technologies, Inc.

canceling echo, in which case $\hat{\mathbf{h}}_b(n)$ is copied to $\hat{\mathbf{h}}_f(n)$. Regarding adaptation, the normalized least-mean-squares (NLMS) algorithm is used. The background filter is updated at each time *n* using

$$\hat{\mathbf{h}}_b(n) = \hat{\mathbf{h}}_b(n-1) + \frac{\mu}{\mathbf{x}(n)^T \mathbf{x}(n) + \delta} e_b(n)\ \mathbf{x}(n), \quad (7)$$

where μ, $0 < \mu < 2$, is the adaptation step size, and $\delta > 0$ is a regularization constant used to improve adaptation stability.

2.2 Ochiai, Araseki and Ogihara Decision Logic

Ochiai, Araseki and Ogihara (OAO) [2] presented the following logic to determine when the background coefficients should be copied to the foreground. $\hat{\mathbf{h}}_b(n)$ is copied to $\hat{\mathbf{h}}_f(n)$ if and only if:

i) $L_j[e_b(n)] < \gamma L_j[y(n)]$,

ii) $L_j[e_b(n)] < \beta L_j[e_f(n)]$, and

iii) $L_j[y(n)] < L_j[x(n)]$,

for all $j = 0,1,...,D-1$, D a positive block count, where

$$L_j[a(n)] = \sum_{i=0}^{M-1} |a(n - jM - i)| \ , \quad n = 0, \pm M, \pm 2M, ... \quad (8)$$

and where $0 < \gamma < 1$, $0 < \beta < 1$ are decision thresholds. The coefficient copy is performed if i)-iii) are satisfied for *D* consecutive *M*-sample time intervals *j*. In addition to i)-iii), the coefficient transfer is inhibited for a total duration of *T* seconds if for any *j*

$$L_j[y(n)] > L_j[x(n)] \ . \quad (9)$$

In [2], these values are used for 8 kHz sampling: $\gamma = 0.125$ (–18 dB), $\beta = 0.875$ (~ –1 dB), *M* = 128 (16 ms), *D* = 3 and *T* = 128 ms (1024 samples).

Condition i) ensures the background adaptive filter is canceling echo, while condition ii) ensures the background filter is outperforming the foreground filter. Both iii) and (9) define a Geigel-like doubletalk detector.

The above decision logic is effective for certain applications, but is not without shortcomings. First, conditions i) and ii) are not always sufficient to prevent coefficient transfer, even in the presence of doubletalk and/or high background noise. For speech or any other non-spectrally diverse excitation, the inequalities in i) and ii) can be satisfied in the short term (over duration *D* in (8), for example) even though the actual misalignment error of the background coefficients is worse than that of the foreground coefficients. Second, i) and ii) employ thresholds that limit the responsiveness of the logic to changes in the performance of the background canceler and to changes in the physical echo path. Condition i) requires the background canceler to achieve a certain degree (18 dB in [2]) of cancellation before the foreground can be updated. In the presence of an echo path change, for example, i) can prolong the presence of annoying echo. The threshold in ii) ensures that the foreground is updated only in steps, in effect quantizing the convergence trajectory of the echo canceler. Last, condition iii) [with (9)] ensures no update is performed unless $|y(n)| < |x(n)|$. But, this property can be used to inhibit adaptation only in cases for which the physical echo path introduces signal loss ($\mathbf{h}^T\mathbf{h} < 1$). If the echo path introduces gain ($\mathbf{h}^T\mathbf{h} > 1$), condition iii) prevents adaptation even in the absence of near-side speech and noise. For this reason, these rules cannot in general be used in echo-canceling speakerphones, where $\mathbf{h}^T\mathbf{h} > 1$.

Some of these characteristics are evident in the examples presented in section 5.

2.3 Threshold-Free Decision Logic

In addition to the beneficial aspects of the OAO logic, a two-path canceler decision logic should possess these characteristics:

- Faster initial convergence and reconvergence following echo path changes.
- Applicability to echo paths having signal gain ($\mathbf{h}^T\mathbf{h} > 1$).
- Reduced dependence upon user-selected constants, such as thresholds and timers.

To this end, consider the decision methodology described below. For each time *n,* let $\hat{\mathbf{h}}_b(n)$ be copied to $\hat{\mathbf{h}}_f(n)$ if

$$\frac{\bar{e}_b(n)}{\bar{y}(n)} < ERLE_{best}(n) = \frac{e_{best}(n)}{y_{best}(n)} \quad (10)$$

or, equivalently, if

$$\bar{e}_b(n)\, y_{best}(n) < \bar{y}(n)\, e_{best}(n) \ , \quad (11)$$

where $\bar{e}_b(n)$ and $\bar{y}(n)$ are the smoothed envelopes

$$\bar{e}_b(n) = \alpha\, \bar{e}_b(n-1) + (1-\alpha)\,|e_b(n)| \ , \quad (12)$$

$$\bar{y}(n) = \alpha\, \bar{y}(n-1) + (1-\alpha)\,|y(n)| \ , \quad (13)$$

and $0 < \alpha < 1$ is the smoothing parameter. Here, $e_{best}(n)$ and $y_{best}(n)$ are previous values of (12) and (13), and at any point in time their ratio in (10) represents the best-attained background echo return loss enhancement (ERLE). When (11) is satisfied, the coefficient copy is performed, and the best-attained ERLE is updated with the envelopes of the better-performing background filter at the time of the copy:

$$e_{best}(n) = \bar{e}_b(n),\ y_{best}(n) = \bar{y}(n) \ . \quad (14)$$

By comparing the ERLE of the background filter,

$$ERLE_b(n) = \frac{\bar{e}_b(n)}{\bar{y}(n)} \ , \quad (15)$$

to $ERLE_{best}(n)$, the foreground filter is updated only with coefficients that provide an ERLE that is smaller than the smallest yet achieved since the best-attained ERLE was initialized. Because (11) uses no threshold, the foreground canceler is continually (smoothly) updated during convergence of the background canceler.

Of course, $ERLE_{best}(n)$ is meaningful only over a period of time for which the echo path is fixed. Should the path change, the value of the measure no longer applies; it must be reset, or "leaked." The following mechanism is used to achieve this. At each time *n*, whether or not (11) is satisfied, if the following two conditions are satisfied

$$1)\ \bar{e}_b(n) < \bar{y}(n) \ , \quad (16a)$$

$$2)\ \bar{e}_b(n) < \bar{e}_f(n) \ , \quad (16b)$$

then

$$y_{best}(n) = \alpha\, y_{best}(n-1) + (1-\alpha)\bar{y}(n) \text{ , and} \quad (17a)$$

$$e_{best}(n) = e_{best}(n-1) + (1-\alpha)[\bar{e}_f(n) - \bar{e}_b(n)] \ . \quad (17b)$$

Equation (17b) can be rewritten

$$e_{best}(n) = \alpha\, e_{best}(n-1) + (1-\alpha)[e_{best}(n-1) + \bar{e}_f(n) - \bar{e}_b(n)] \quad (18)$$

to convey more clearly that $e_{best}(n)$ is also smoothed using a single-pole filter, the input to which is the current output plus the differential $\bar{e}_f(n) - \bar{e}_b(n)$. Note that, because of (16b), the differential in (17b) is always positive.

To summarize, the new logic consists of test (11), with update (14) and tests (16a) and (16b) with updates (17a) and (17b). Note this logic uses no thresholds or timers. Smoothing parameter α is the only constant used.

The update mechanism for $ERLE_{\text{best}}(n)$ is a crucial component of the proposed logic. Consider, first, its behavior for the case of a fixed echo path in the absence of near-side speech or noise. With far-side speech excitation, (16a) is continually satisfied because the background filter is converging. When $\bar{e}_{\text{b}}(n)$ is even infinitesimally smaller than $\bar{e}_{\text{f}}(n)$, (16b) is satisfied and the update of $y_{\text{best}}(n)$ and $e_{\text{best}}(n)$ is performed. Because $\bar{e}_{\text{f}}(n) - \bar{e}_{\text{b}}(n)$ is very small, $e_{\text{best}}(n)$ grows slowly; in fact, because time constant α is relatively large (> 100 ms), the growth rate of $e_{\text{best}}(n)$ is lower than the rate of decrease of $\bar{e}_{\text{b}}(n)$. As a result, coefficient update condition (11) is satisfied almost continually, and the foreground filter is smoothly updated with better performing coefficients.

Now consider a doubletalk condition, following a period of convergence. At the onset of near-side speech, $\hat{\mathbf{h}}_{\text{b}}(n)$ diverges and, therefore, $\bar{e}_{\text{b}}(n) > \bar{e}_{\text{f}}(n)$. Equation (16b) is not satisfied, thus preventing update of $ERLE_{\text{best}}(n)$, as desired.

Finally, in the case of an echo path change, both $\bar{e}_{\text{f}}(n)$ and $\bar{e}_{\text{b}}(n)$ rise instantaneously because neither filter matches the new echo path. Eventually, $\bar{e}_{\text{b}}(n)$ decreases, both (16a) and (16b) are satisfied, and $ERLE_{\text{best}}(n)$ is updated. As $e_{\text{best}}(n)$ increases, $\bar{e}_{\text{b}}(n)$ decreases, and eventually (11) is satisfied and the foreground filter is updated.

At initialization $y_{\text{best}}(0) = \bar{y}(0) = \bar{e}_{\text{f}}(0) = \bar{e}_{\text{b}}(0) = 0$ dB, and $e_{\text{best}}(0) = -1$ dB, where maximum scale is referenced to 0 dB. $ERLE_{\text{best}}(n)$ is initialized to −1 dB so that (11) is not transiently satisfied at start-up.

3 EXAMPLES

3.1 Input Signals and Echo Paths

Figure 2 shows the input speech time series used in the examples. Time series are confined to a full-scale range of [–0.5,0.5], are sampled at 8 kHz, and are filtered to a frequency range of approximately [200, 3400] Hz. The far-side signal *x(n)* consists of 20 seconds of phonetically balanced sentences uttered by a mixture of males and females. The near-side speech consists of 5 seconds of male speech formed by repeating a single utterance. The level of near-side speech is intentionally lower, by about 6 dB, than the far-side speech. This is done to showcase one benefit of the two-path structure, namely, its robustness to lower-level doubletalk. Single-path cancelers employing signal-level-based doubletalk detectors are known to perform worse as the near-speech-to-echo ratio drops [3]. Noise sequence $w(n)$ in (5) consists of Gaussian deviates having zero-mean and standard deviation $\sigma_w = 0.00025$ [$w(n)$ is not shown in Fig. 2(b)]. This results in a near-speech-to-noise ratio of about 40 dB.

The examples use a synthetic echo path, $\mathbf{h}$, given by the following random-sequence-modulated, exponentially decaying window model:

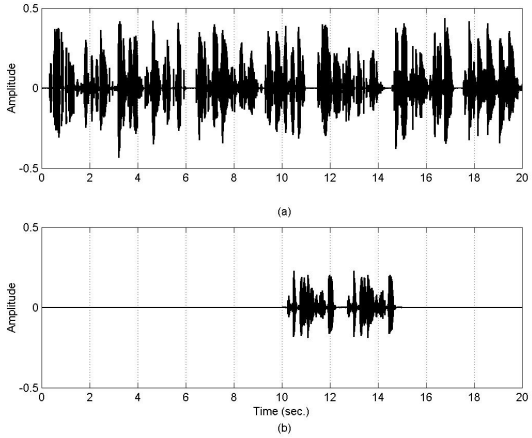

Figure 2. Input time series for examples. (a) far-side signal $x(n)$ used in all examples. (b) near-side signal $v(n)$ used only in doubletalk examples.

$$g(n) = [u(n-16) - u(n-N)]e^{-n/(N/5)}r(n), \quad n = 0,1,\ldots,N-1 \qquad (19a)$$

$$\mathbf{h} = \mathbf{g}/\|\mathbf{g}\| \qquad (19b)$$

where *u(n)* is the unit step function, *r(n)* is a sequence of Gaussian deviates of zero-mean and unit variance, and $\|\mathbf{g}\|^2 = \mathbf{g}^T\mathbf{g}$. The impulse response in (19b) has unit energy, and its frequency response reasonably approximates the multimodal nature of acoustic echo paths. Using scaling, echo paths of arbitrary energy can be generated.

Two measures are used to evaluate echo canceler performance. These are the foreground misalignment error (MAE),

$$MAE[\hat{\mathbf{h}}_{\text{f}}(n)] = \left\|\mathbf{h} - \hat{\mathbf{h}}_{\text{f}}(n)\right\|^2 \Big/ \|\mathbf{h}\|^2, \qquad (20)$$

which conveys the normalized mean-squared error of the foreground filter coefficients relative to the known echo path, and the foreground ERLE, $ERLE_{\text{f}}(n)$, given by (15) but with $\bar{e}_{\text{f}}(n)$ in place of $\bar{e}_{\text{b}}(n)$. While both the MAE and the ERLE decrease as the adaptive filter converges, the instantaneous ERLE is a function of excitation signal *x(n)* while the MAE is not.

3.2 Example 1. Doubletalk, ERL = 12dB

This first example demonstrates the similarity of the OAO logic and proposed logic for the case of a lossy echo path. Figure 3 shows the MAE and ERLE for the two-path echo canceler using the proposed logic (solid line) and OAO logic (dashed line). The OAO logic uses the original constants listed in section 2.2. So that the OAO logic can be used, the echo path generated by (19a) and (19b) is scaled to provide an echo return loss (ERL) of 12 dB, that is, $\|\mathbf{h}\| = 0.25$. For both the OAO logic and proposed logic, N in (4) is 512 (64 ms), μ in (7) is 0.5, and $\delta = 0.001$ (a regularization constant is used in [2] for the form (7), but its value is not stated). Smoothing constant α for the proposed logic is chosen to provide a time constant of 150 ms. Note this same α is used in both (12) and (13) as well as for the updates in (17a) and (17b).

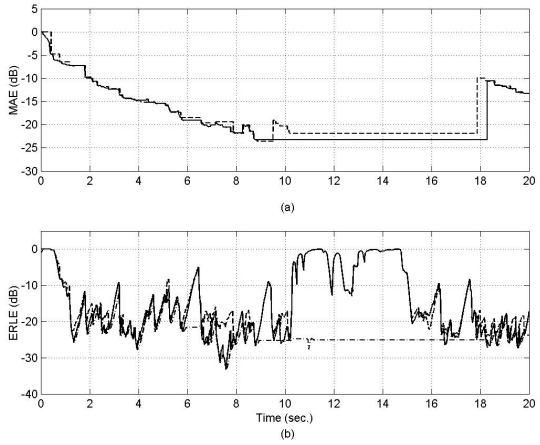

Figure 3. Proposed-logic vs. OAO-logic performance during doubletalk. Proposed-logic (solid), OAO-logic (dash), and best-attained ERLE (dot-dash). ERL = 12 dB.

The results in Fig. 3 demonstrate performance typical of the two-path structure using the OAO logic. After a slight initial delay, caused by threshold γ, the OAO MAE (Fig. 3a, dash) decreases in small, abrupt steps, a result of the background-to-foreground performance comparison threshold β. The ERLE (Fig. 3b, dash) does not clearly convey these small steps because the excitation *x(n)* is non-white. As seen from the MAE over the region 10-15 s, robustness to doubletalk is excellent. The MAE degrades slightly near the onset of near speech at 10 s, but is stable over the doubletalk period. Following the doubletalk period, near the 18 s mark, the foreground filter is updated with coefficients having worse MAE. This occurs because the background filter need only achieve a certain level of ERLE performance relative to that of the foreground filter for the coefficient transfer to occur.

The proposed logic (Fig. 3, solid) shows performance comparable to the OAO logic. The MAE for the proposed logic (Fig. 3a, solid) follows a more smooth and continual convergence curve and its performance during the doubletalk is slightly better. The dot-dash curve in Fig. 3b shows the proposed logic's best-attained ERLE in (10).

3.3 Example 2. Path Change, ERL = 12 dB to ERL = –12dB

This example demonstrates the behavior of the two decision methods under a severe echo-path change, showing the superiority of the proposed logic under such conditions. No doubletalk is present in this example. At the 10 s mark, the ERL of the echo path changes from 12 dB ($\|\mathbf{h}\| = 0.25$) to -12 dB ($\|\mathbf{h}\| = 4.0$). First, note in Fig. 4 the unusually long delay in the OAO-logic's initial update of the foreground canceler. Again, this is a function of the thresholds used by the OAO logic, but the effect is exacerbated in this example by the echo path generated in (19a) [each experiment uses a different random sequence in (19a)]. At the 10 s mark, the ERL of the echo path changes from 12 dB to –12 dB. Following the 10 s mark, the OAO-logic is unable to update the foreground because condition (9) is satisfied continually. The proposed logic, in comparison, converges to the new path. Note the best-attained ERLE (Fig. 4b, dot-dash) of the proposed logic is updated following the path change, as desired.

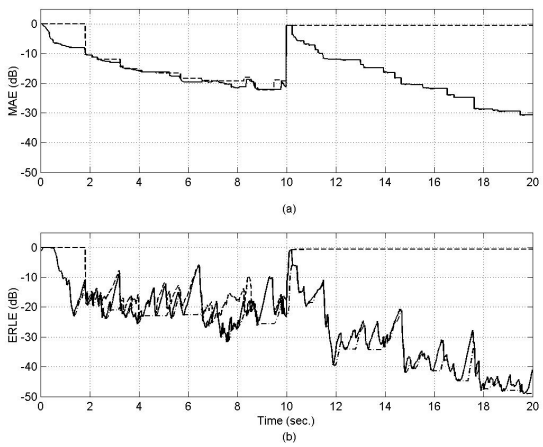

Figure 4. Proposed-logic vs. OAO-logic performance for an echo path change. Proposed-logic (solid), OAO-logic (dash), and best-attained ERLE (dot-dash). ERL changes from 12 dB to –12 dB at the 10 s mark.

4 SUMMARY

A new decision methodology has been presented for use with two-path echo canceler structures. This logic is relatively compact and uses fewer user-selected constants. Additionally, the resulting echo canceler converges more rapidly and smoothly than that described in [2] and is applicable to applications in which the echo path introduces a signal gain.

5 REFERENCES

[1] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.,* vol. 26, No. 5, pp. 647-653, May 1978.

[2] K. Ochiai, T. Araseki and T. Ogihara, "Echo Canceler with Two Echo Path Models," *IEEE Trans. Commun.,*Vol. COM-25, No. 6, pp. 589-595, June 1977.

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