Evaluation of an Improved Deviation Measure for Two-Path Echo Cancellation

Christian Schüldt
Blekinge Institute of Technology
Department of Signal Processing
SE-37225, Ronneby, Sweden.
Email: christian.schuldt@bth.se

Fredric Lindstrom
Limes Audio AB
Döbelnsgatan 19
SE-90330, Umeå.

Ingvar Claesson
Blekinge Institute of Technology
Department of Signal Processing
SE-37225, Ronneby, Sweden.

***Abstract*—The two-path algorithm is a well-known approach for overcoming the dead-lock problem in echo cancellation systems. Typically, a fixed foreground filter is producing the echo cancelled output while a continuously updating background filter adapts to the echo-path. When the background filter is considered to perform better than the foreground filter, the coefficients of the background filter are copied into the foreground filter. To determine which filter is better adjusted to the true echo-path, a filter deviation measure can be used.**

Recently, a method which introduces a delay in the calculation of the filter deviation measure, yielding a more reliable estimate has been proposed. However, a thorough evaluation of the effect of different delay settings has not yet been performed.

Thus, in this paper a number of simulations with different delay parameter settings are carried out to show how this parameter affects the overall performance of the filter deviation measure.

I. Introduction

The use of parallel adaptive filters, the so-called *two-path approach* [1], is frequent in many adaptive filter based echo cancellation systems to overcome the dead-lock problem. The dead-lock problem occurs when a change of the echo-path is mistaken for local noise, causing the adaptive filter control to halt the updating of the adaptive filter. In the two-path approach, two parallel adaptive filters are used: a continuously updating *background filter* and a fixed *foreground filter*. The fundamental idea is to constantly compare the echo cancellation performance of the background- and the foreground filter and to copy the background filter coefficients into the foreground filter whenever the background filter is considered to be better adjusted to the echo-path than the foreground filter.

A straight-forward approach for determining which filter is better adjusted to the echo-path is to compare the output errors from the filters [1], [2]. However, during *double-talk*, i.e. situations where both the signal driving the filter as well as the near-end signal are active simultaneously, cancellation of the near-end signal can occur, making the output error an unreliable measure for determining how well adjusted the filter is to the echo-path [3]. In [4] an alternative filter deviation measure was introduced, although this measure also suffers from problems due to cancellation of near-end speech in double-talk situations. As a solution to this problem, [5] introduced a delay in the calculation of the filter deviation measure to obtain a more reliable estimate. However, setting of the delay parameter and its effect on the performance of the algorithm was not entirely evaluated - in essence the paper just concluded that a large delay was better than no delay (i.e. the method in [4]).

The main purpose of this paper is to show how different settings of the delay parameter affects the performance of the algorithm. It is concluded that having a delay is not always better than no delay. It is also concluded that a negative delay parameter on the other hand always seem to give better performance than no delay and that the step-size parameter in the adaptive filter update strongly affects the performance of the filter deviation measure.

II. Two-path echo cancellation

In this paper, the two-path approach is considered in an acoustic echo cancellation environment where the foreground filter is denoted as $\hat{\mathbf{h}}_\mathbf{f}(k) = [\hat{h}_{f_0}(k), \hat{h}_{f_1}(k), \cdots, \hat{h}_{f_{N-1}}(k)]^T$ and the updating background filter as $\hat{\mathbf{h}}_\mathbf{b}(k) = [\hat{h}_{b_0}(k), \hat{h}_{b_1}(k), \cdots, \hat{h}_{b_{N-1}}(k)]^T$, where N is the filter length and k is the sample index. The echo cancelled signals are calculated by subtracting the filtered echo estimates from the microphone signal $y(k)$ according to

$$e_f(k) = y(k) - \hat{y}_f(k), \tag{1}$$

where $\hat{y}_f(k) = \hat{\mathbf{h}}_\mathbf{f}(k)^T\mathbf{x}(k)$ and

$$e_b(k) = y(k) - \hat{y}_b(k), \tag{2}$$

where $\hat{y}_b(k) = \hat{\mathbf{h}}_\mathbf{b}(k)^T\mathbf{x}(k)$, $e_f(k)$ is the output error from the foreground filter, $e_b(k)$ is the output from the background filter, $\mathbf{x}(k) = [x(k), x(k-1), \cdots, x(k-N+1)]^T$ is the regressor vector and $x(k)$ is the loudspeaker signal.

Updating of the adaptive filter can be achieved with a number of algorithms. In this paper, as well as in [3], [4], [5], normalized least mean square (NLMS) was used in order to simplify the analysis. Updating of the background filter is thus performed as

$$\hat{\mathbf{h}}_\mathbf{b}(k+1) = \hat{\mathbf{h}}_\mathbf{b}(k) + \mu\frac{e_b(k)\mathbf{x}(k)}{\mathbf{x}(k)^T\mathbf{x}(k) + \epsilon}, \tag{3}$$

where μ is the step-size control variable and ϵ is a regularization term to avoid division by zero [6].

The foreground filter is updated by copying the coefficients of the background filter into the foreground filter. When this copying is performed is controlled by the *transfer logic*, which essentially is a set of conditions that should be fulfilled in order to initiate a filter copying operation. Typical transfer logic conditions, in addition to trivial conditions such as sufficient loudspeaker and microphone energy, are [1], [3], [4]

- $\frac{\sigma^2_{e_f}(k)}{\sigma^2_{e_b}(k)} > T_1$ (i.e. the background filter must produce a lower output error signal than the foreground filter)
- $\frac{\sigma^2_{x}(k)}{\sigma^2_{e_b}(k)} > T_2$ (i.e. the acoustic coupling and echo return loss enhancement must be lower than T_2)

where T_1 and T_2 are thresholds and $\sigma^2_x(k)$, $\sigma^2_y(k)$, $\sigma^2_{e_b}(k)$, $\sigma^2_{e_f}(k)$ denote the short-time energy of the loudspeaker signal, microphone signal, background filter error signal and foreground filter error signal, respectively.

III. Filter deviation measure

The problem related to comparing output errors of the filters is that during double-talk, cancellation of near-end speech by the updating background filter can occur [3], making the output error from the background filter less than then output error from the foreground filter, even though the background filter is misadjusted due to updating during near-end speech. This means that the first transfer logic condition in the previous section is not always reliable, imposing the need of additional conditions for updating the foreground filter.

In [5] the (background) filter deviation measure

$$\nu_{b_D}(k) = \left| \frac{r_{\hat{y}e_{b_D}}(k)}{r_{\hat{y}y_D}(k)} \right|, \tag{4}$$

where $r_{\hat{y}e_{b_D}}(k) = \mathbb{E}[\hat{y}_D(k)e_b(k-D)]$, $r_{\hat{y}y_D}(k) = \mathbb{E}[\hat{y}_D(k)y(k-D)]$, $\hat{y}_D(k) = \hat{\mathbf{h}}_\mathbf{b}(k)^T\mathbf{x}(k-D)$, D is a delay constant and $\mathbb{E}[\cdot]$ denotes expectation (ensemble average) was introduced.

Assuming that the microphone signal can be modeled as

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

where the (unknown) echo-path $\mathbf{h} = [h_1, h_2, \cdots, h_{N-1}]^T$ is of length N, i.e. same length as the adaptive filters, and $n(k)$ is near-end noise and/or speech, allows combination of equations (4), (5) and (3) as [5]

$$\nu_{b_D}(k) = \left| \frac{\left(\mathbf{h} - \hat{\mathbf{h}}_\mathbf{b}(k-D)\right)^T \mathbf{R}_{\mathbf{xx_D}} \hat{\mathbf{h}}_\mathbf{b}(k) + \rho_{b_D}(k)}{\mathbf{h}^T\mathbf{R}_{\mathbf{xx_D}}\hat{\mathbf{h}}_\mathbf{b}(k) + \rho_{b_D}(k)} \right|, \tag{6}$$

where $\mathbf{R}_{\mathbf{xx_D}} = \mathbb{E}[\mathbf{x}(k-D)\mathbf{x}(k-D)^T]$ and $\rho_{b_D}(k) = \mathbb{E}[\hat{y}_D(k)n(k-D)]$. By setting $D = 0$ one obtains the filter deviation estimate proposed in [4]. Further, in [4] it was assumed that $\rho_{b_D}(k) = 0$. In that case it can clearly be seen that $\nu_{b_D}(k) \approx 0$ if $\mathbf{h} \approx \hat{\mathbf{h}}_\mathbf{b}(k)$ (i.e. if the adaptive background filter is well adjusted to the echo-path) and $\nu_b(k) \gg 0$ if $\mathbf{h} \not\approx \hat{\mathbf{h}}_\mathbf{b}(k)$.

Thus, the resulting additional transfer logic condition is then $\nu_b(k) < \nu_f(k)$, hence the deviation measure must indicate that the background filter is better adjusted to the echo path than the foreground filter for an update of the foreground filter to occur.

However, as pointed out in [5], if $n(k)$ is non-white (i.e. $\mathbb{E}[n(k)n(k+l)] \neq 0 \quad \forall l \neq 0$) and $D = 0$ it can be seen that $\rho_{b_D}(k) \neq 0$ by observing the adaptive filtering equations (2) and (3). Due to the characteristics of speech, it was argued in [5] that $|\rho_{b_D}(k)|$ is more likely to be lower for $D \neq 0$ than for $D = 0$, resulting in a more accurate filter deviation estimator.

However, in the simulations in [5] the variable D was simply set to $D = 32$ without further evaluation. Because of this, the following sections will show how different settings of the delay parameter D affects the performance of the described filter deviation measure.

It should be noted that although $D < 0$ indicate non-causality, delaying the respective signals by D before performing the filtering and adaptive filter updating will have the same effect and also allow real-time implementation.

IV. Simulations

To evaluate the performance of the algorithm for different settings of D, two sets of speech signals sampled at 8 kHz were used. The first set consisted of a speech signal from a male speaker used as the driving loudspeaker signal $x(k)$ and a speech signal from a female speaker used as near-end speech signal $s(k)$. The second set consisted of a speech signal from a female speaker used as the driving loudspeaker signal $x(k)$ and a speech signal from a male speaker used as near-end speech signal $s(k)$. A total of 8 different signal constellations (simulation scenarios), with 36 seconds duration each, was created by using both signal sets and varying the starting time of the near-end speech to occur after either 16, 19, 21 or 24 seconds. The results presented in this paper are the average results form the 8 different simulation scenarios. In all scenarios the same impulse response, a filter $\mathbf{h}$ of length $N = 500$ measured in a normal office, was used. The driving loudspeaker signal $x(k)$ was convoluted with $\mathbf{h}$ to obtain the echo signal $d(k)$ and the simulated microphone signal was then formed by summing the signals $d(k)$, $s(k)$ as well as a local stationary noise signal $w(k) \sim \mathcal{N}(0, 10^{-6})$, i.e. $y(k) = d(k) + s(k) + w(k)$.

To isolate the performance of the deviation measure, the two-path transfer logic was in each simulation scenario set to transfer the background filter into the foreground filter at all instances from 0 seconds up to the sample index where near-end speech starts and thereafter halt the update of the foreground filter. This means that the background filter and the foreground filter will be identical up to the occurrence of near-end speech (double-talk) and then the background filter will diverge while the foreground filter stays converged. This is the same procedure as in [5].

Exponential recursive weighting was used to obtain approximations of the ensemble averages used in the deviation

estimates [4] as

$$\begin{aligned}\hat{r}_{\hat{y}e_{b_D}}(k) &= \lambda\hat{r}_{\hat{y}e_{b_D}}(k-1) + (1-\lambda)\hat{y}_D(k)e_b(k-D)\\ \hat{r}_{\hat{y}e_{f_D}}(k) &= \lambda\hat{r}_{\hat{y}e_{f_D}}(k-1) + (1-\lambda)\hat{y}_D(k)e_f(k-D)\\ \hat{r}_{\hat{y}y_D}(k) &= \lambda\hat{r}_{\hat{y}y_D}(k-1) + (1-\lambda)\hat{y}_D(k)y(k-D) \quad (7)\end{aligned}$$

where the forgetting factor was set to $\lambda = 0.9995$, which is the same as in [5].

Two measures were used to evaluate the performance: *the number of errors* defined as the number of sample indices where $\nu_{b_D}(k) < \nu_{f_D}(k)$ during double-talk, i.e. the number of sample indices where the algorithm *falsely* indicates that the background filter is better adjusted to the echo-path than the foreground filter (despite having diverged due to double-talk), and *the averaged difference between the foreground- and background filter deviation measure* defined as

$$\frac{1}{N_d - p_i}\sum_{i=p_i}^{N_d}\left(\nu_{b_D}(i) - \nu_{f_D}(i)\right) \quad (8)$$

where p_i is the sample index where near-end speech (and thus double-talk) starts and N_d is the length (in samples) of the double-talk sequence. The averaged results over the 8 different simulation scenarios and for three different step-size parameter settings are shown in figures 1 and 2, respectively. Interesting to note is that the results for the individual scenarios where fairly equal, i.e. no results from a single scenario was very different from the average results.

V. Results

By observing the upper and middle plot of figure 1, representing simulations with step-size parameter $\mu = 0.95$ and $\mu = 0.5$ respectively, it can clearly be seen that for $D < 0$ the number of occasions where the deviations measure falsely indicates that the background filter is better adjusted than the foreground filter during double-talk decreases as the time-lag D decreases. Interesting to note is also that $D = 1$ gives the worst result for $\mu = 0.95$ (upper plot) and $D = 2$ gives the worst result for $\mu = 0.5$ (middle plot). Also interesting to note is that by observing the lower plot of figure 1, representing simulations with step-size parameter $\mu = 0.1$, it can be seen that the deviation measure performs significantly better for $D < 0$ than for $D > 0$ in this case. It can be speculated that this is due to the relatively slow convergence of the adaptive background filter and the properties of speech: it seems that for this step-size setting the adaptive filter captures the characteristics of the near-end speech, making $|\rho_{b_D}(k)| \gg 0$. It is clear that increasing D up to at least 64 does not improve the performance. (Of course, increasing D enough will indeed improve the performance since speech is only considered stationary up to about 20 ms [6].) For relatively large step-sizes on the other hand, it might be that the adaptive filter does not get a chance to adjust to the near-end speech owing to fluctuation of the NLMS update vector.

Figure 2 shows the the averaged difference between the foreground- and background filter deviation measure calculated as equation (8) averaged over all 8 simulation scenarios.

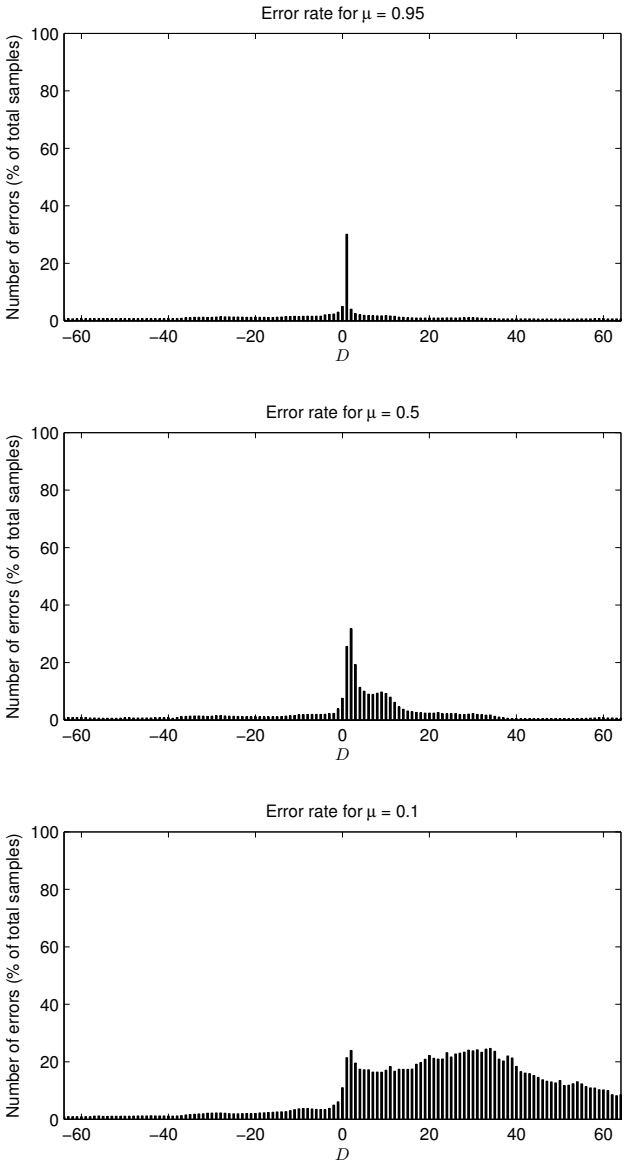

Fig. 1. Number of errors, defined as the number of sample indices where $\nu_{b_D}(k) < \nu_{f_D}(k)$ during double-talk, in percent for three different step-sizes ($\mu = \{0.95, 0.5, 0.1\}$) and a number of different time-lags, i.e. settings of D. The plots show the average results of 8 different simulations.

It can be seen that for step-size parameter $\mu = 0.95$ and $\mu = 0.5$ (upper and middle plot), increasing $|D|$ sufficiently gives an improved margin between $\nu_{b_D}(i)$ and $\nu_{b_D}(i)$ during double-talk, which is highly desirable. However, the margin seems to decrease with the step-size parameter - which is expected since an adaptive filter with a large step-size diverges more rapidly than an adaptive filter with a small step-size. Similar conclusions as for figure 1 can be drawn from figure 2, i.e. that using $D < 0$ for $\mu = 0.1$ gives significant performance improvement over using $D > 0$, while for large step-sizes it

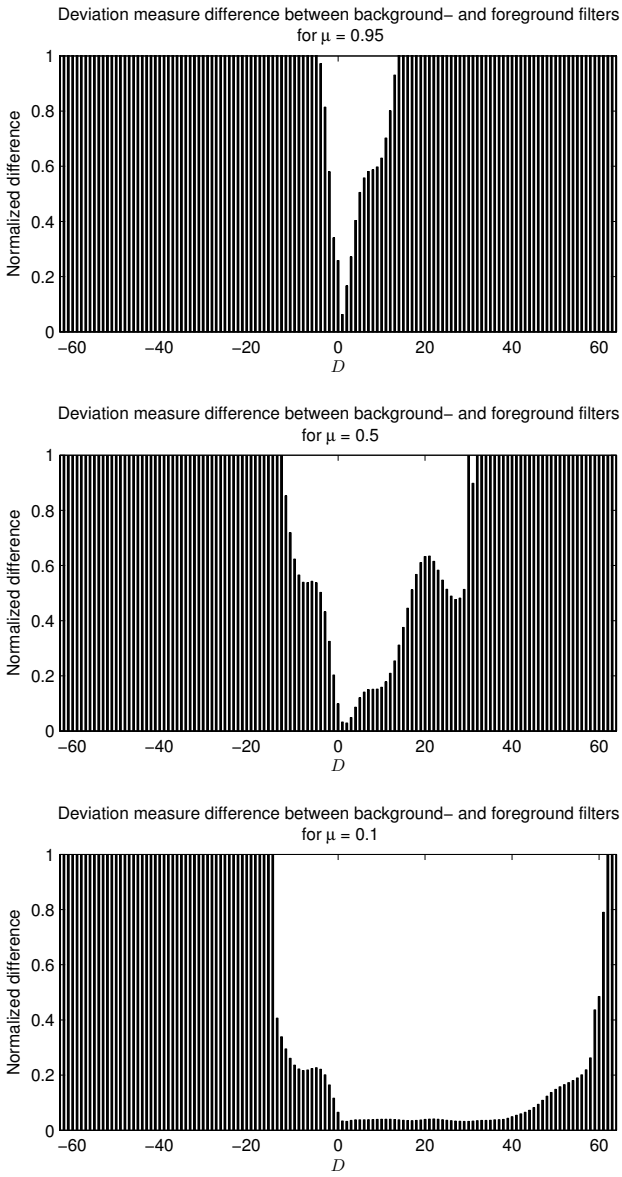

Fig. 2. Deviation measure difference between background and foreground filters for three different step-sizes ($\mu = \{0.95, 0.5, 0.1\}$) and a number of different time-lags, i.e. settings of D. The plots show the average results of 8 different simulations.

does not seem to matter as long as $|D|$ is fairly large.

Thus, the conclusion to be drawn from the simulations is that $D < 0$ always seem to give better performance than $D = 0$ and a smaller D always seem to give better performance than a larger D if D is negative, regardless of the step-size. For positive values of D however, starting with $D = 0$, the performance seem to worsen up to a point (depending on the step-size) if D increases, and then improve again if D increases further.

VI. Conclusions

In [4] an adaptive filter deviation measure for use in two-path echo cancellation was proposed. This adaptive filter deviation measure was improved in [5] by introducing a delay D in the calculation. This paper has evaluated how different settings of D affects the performance in practice, using simulations with speech signals. It has been concluded that that setting $D < 0$ consistently seem to give better performance than $D = 0$ and for smaller step-sizes ($\mu < 0.5$) the sign of D is more important for the performance than for larger step-sizes. A setting of $D > 0$ could also give better performance than $D = 0$, but not necessarily, as this depends on the actual value of D and the adaptive filter step-size parameter.

Acknowledgment

The funding from the Swedish Knowledge Foundation (KKS) is gratefully acknowledged.

References

[1] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceler with two echo path models," *IEEE Transactions on Communications*, vol. COM-25, no. 6, pp. 8–11, June 1977.

[2] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and evaluation of an acoustic echo canceller using the duo-filter control system," in *Proceedings of International Workshop on Acoustic Echo and Noise Control*, June 1995, pp. 79–82.

[3] F. Lindstrom, C. Schüldt, and I. Claesson, "An improvement of the two-path algorithm transfer logic for acoustic echo cancellation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 1320–1326, May 2007.

[4] M. Iqbal and S. Grant, "Novel and efficient download test for two path echo canceller," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 2007, pp. 167–170.

[5] C. Schüldt, F. Lindstrom, and I. Claesson, "An improved deviation measure for two-path echo cancellation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2010, pp. 305–308.

[6] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley, 2004.