AN IMPROVED DEVIATION MEASURE FOR TWO-PATH ECHO CANCELLATION

Christian Schüldt^a, Fredric Lindstrom^b, Ingvar Claesson^a

^aBlekinge Institute of Technology, Department of Signal Processing, SE-37225, Ronneby, Sweden.

^bLimes Technology AB, Döbelnsgatan 19, SE-90330, Umeå, Sweden.

ABSTRACT

Parallel adaptive filters have been proposed for echo cancellation to solve the dead-lock problem, occurring when the echo is detected as near-end speech after a severe echo-path change; causing the updating of the adaptive filter to halt. To control the parallel filters and monitor their performance, estimates of the filter deviation (i.e. the squared norm of the filter mismatch vector) are typically used.

This paper presents a modification of a filter mismatch estimator. The proposed modification requires slightly more computational resources than the original measure, but provides a significant improvement in terms of robustness during double-talk. This is shown both analytically and through simulations.

Index Terms— Echo cancellation, adaptive filters, two-path, transfer logic

1. INTRODUCTION

In systems using adaptive filters for echo cancellation, it is of outmost importance to have a mechanism controlling the adaptation of the filter to avoid divergence in the case of local disturbances. Such mechanism is commonly referred to as a *double-talk detector* (DTD), with the purpose to differentiate between situations where only echo is present (single-talk) and situations where echo and local disturbances are present (double-talk). Several DTDs have been proposed, such as the Geigel detector [1] and detectors based on correlation [2] and coherence [3]. However, a problem related to all DTDs is the *dead-lock* problem occurring when the echo is detected as a local disturbance, preventing the adaptive filter from updating when it is in fact needed. This can happen after a severe change of the echo-path (i.e. an *echo-path change*). In an acoustic echo cancellation environment, an echo-path change constitutes a change in the acoustic environment such as dislocation of the loudspeaker and/or microphone or people moving in the room.

As a solution to the dead-lock problem, at the cost of an extra adaptive filter, the two-path algorithm has been proposed [4, 5, 6]. This paper presents an improved filter de-

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

Fig. 1. Scheme of the two-path algorithm in an acoustic environment.

viation measure intended for the two-path algorithm. The proposed filter deviation measure is compared to the one presented in [5], both analytically and through simulations. It is shown that the proposed measure has similar desirable properties as the measure in [5], while being much more robust in double-talk situations.

2. TWO-PATH ECHO CANCELLATION

A scheme illustrating the two-path echo cancellation approach in an acoustic echo cancellation environment is shown in figure 1. The updating of the background filter, $\hat{\mathbf{h}}_b(k) = [\hat{h}_{b0}(k), \hat{h}_{b1}(k), \dots, \hat{h}_{b_{N-1}}(k)]^T$, of length N could be performed with a variety of algorithms, and is in this paper performed with the normalized least mean square (NLMS) [7] owing to its simplicity, according to

$$\begin{aligned} e_b(k) &= y(k) - \hat{\mathbf{h}}_b(k)^T \mathbf{x}(k) \\ \hat{\mathbf{h}}_b(k+1) &= \hat{\mathbf{h}}_b(k) + \mu \frac{e_b(k) \mathbf{x}(k)}{\mathbf{x}(k)^T \mathbf{x}(k) + \epsilon}, \end{aligned} \quad (1)$$

where $x(k)$ is the loudspeaker signal, $y(k)$ is microphone signal, $\mathbf{x}(k) = [x(k), x(k-1), \dots, x(k-N+1)]^T$ is the regressor vector, μ is the step-size control variable, ϵ is a regu-

larization term to avoid division by zero and k is the sample index. $[\cdot]^T$ denotes transpose.

The foreground filter, denoted $\hat{\mathbf{h}}_f(k) = [\hat{h}_{f_0}(k), \hat{h}_{f_1}(k), \dots, \hat{h}_{f_{N-1}}(k)]^T$, gives the output error

$$e_f(k) = y(k) - \hat{\mathbf{h}}_f(k)^T \mathbf{x}(k). \quad (2)$$

Updating of the foreground filter is done by copying the filter coefficients of the background filter. At which time instances this copying is performed is controlled by the transfer logic. The transfer logic is a set of conditions which should be fulfilled in order to initiate copying of the filter. Typical transfer logic conditions, in addition to trivial conditions such as sufficient loudspeaker and microphone energy, are [4, 5, 6]

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

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

2.1. Filter deviation

During double-talk the background filter can occasionally produce lower output error than the foreground filter due to the cancellation of near-end speech [6]. This means that the first transfer logic condition presented in the previous section is not always reliable, imposing the need of additional conditions for updating the foreground filter.

One such condition based on the filter deviation has recently been proposed [5], where the adaptive (background) filter deviation is estimated as

$$\nu_b(k) = \left| \frac{r_{e_b y}(k) - \sigma_{e_b}^2(k)}{\sigma_y^2(k) - r_{e_b y}(k)} \right| = \left| \frac{r_{\hat{y} e_b}(k)}{r_{\hat{y} y}(k)} \right|, \quad (3)$$

where $r_{e_b y}(k) = \mathbb{E}[e_b(k)y(k)]$, $r_{\hat{y} e_b}(k) = \mathbb{E}[\hat{y}(k)e_b(k)]$, $r_{\hat{y} y}(k) = \mathbb{E}[\hat{y}(k)y(k)]$, $\hat{y}(k) = \hat{\mathbf{h}}_b(k)^T \mathbf{x}(k)$ and $\mathbb{E}[\cdot]$ denotes expectation (ensemble average).

The microphone signal is modeled as

$$y(k) = \mathbf{h}^T \mathbf{x}(k) + n(k), \quad (4)$$

where the unknown echo-path $\mathbf{h} = [h_1, h_2, \dots, h_{N-1}]^T$ is of length N , i.e. same length as the adaptive filters, and $n(k)$ is near-end noise and/or speech. Using equations (4) and (1) and assuming that $x(n)$ and $n(k)$ are zero mean and uncorrelated, yields that equation (3) can be written as

$$\nu_b(k) = \left| \frac{(\mathbf{h} - \hat{\mathbf{h}}_b(k))^T \mathbf{R}_{xx} \hat{\mathbf{h}}_b(k) + \rho_b(k)}{\mathbf{h}^T \mathbf{R}_{xx} \hat{\mathbf{h}}_b(k) + \rho_b(k)} \right|, \quad (5)$$

where $\mathbf{R}_{xx} = \mathbb{E}[\mathbf{x}(k)\mathbf{x}(k)^T]$ and $\rho_b(k) = \mathbb{E}[\hat{y}(k)n(k)]$. It should be noted that in [5] $\hat{y}(k)$ and $n(k)$ are assumed to be uncorrelated, leading to $\rho_b(k) = 0$. In this case equation (5) provides an estimate of the filter deviation, resulting in $\nu_b(k) \approx 0$ when $\mathbf{h} \approx \hat{\mathbf{h}}_b(k)$ (i.e. when the adaptive filter is well adjusted to the echo-path) and $\nu_b(k) \gg 0$ when $\mathbf{h} \not\approx \hat{\mathbf{h}}_b(k)$.

3. PROPOSED DEVIATION MEASURE

The problem with the described filter deviation estimator in equation (5) is that during double-talk, the disturbing near-end speech present in the microphone signal $y(k)$ and the adaptive filter error signal $e_b(k)$ will corrupt the filter update (see equation (1)). This means that the signal $\hat{y}(k+1)$ will indeed be correlated with $n(k)$. Thus, if $n(k)$ is non-white, which certainly is the case for speech signals, the term $\rho_b(k)$ will not be 0, causing the previously described deviation estimate to be inaccurate.

Because of this problem, an alternative filter deviation estimator

$$\nu_{b_D}(k) = \left| \frac{r_{\hat{y} e_{b_D}}(k)}{r_{\hat{y} y_D}(k)} \right|, \quad (6)$$

where $r_{\hat{y} e_{b_D}}(k) = \mathbb{E}[\hat{y}_D(k)e_b(k-D)]$, $r_{\hat{y} y_D}(k) = \mathbb{E}[\hat{y}_D(k)y(k-D)]$, $\hat{y}_D(k) = \hat{\mathbf{h}}_b(k)^T \mathbf{x}(k-D)$ and D is a delay constant, is proposed.

Using equations (4) and (1), equation (6) can be rewritten as

$$\nu_{b_D}(k) = \left| \frac{(\mathbf{h} - \hat{\mathbf{h}}_b(k-D))^T \mathbf{R}_{xxD} \hat{\mathbf{h}}_b(k) + \rho_{b_D}(k)}{\mathbf{h}^T \mathbf{R}_{xxD} \hat{\mathbf{h}}_b(k) + \rho_{b_D}(k)} \right|, \quad (7)$$

where $\mathbf{R}_{xxD} = \mathbb{E}[\mathbf{x}(k-D)\mathbf{x}(k-D)^T]$ and $\rho_{b_D}(k) = \mathbb{E}[\hat{y}_D(k)n(k-D)]$. The significant difference between equations (7) and (5) lies in the terms $\rho_{b_D}(k)$ and $\rho_b(k)$. As discussed earlier, in the event of disturbing near-end speech ($n(k)$), the near-end speech will disturb the filter update, imposing correlation between $\hat{y}(k+1)$ and $n(k)$. Since the autocorrelation of a speech signal usually decrease rapidly as the lag increases [8], it is obvious that $|\rho_{b_D}(k)|$ is more likely to be lower than $|\rho_b(k)|$, resulting in a more accurate filter deviation estimator.

The extra computational cost for this is one additional filtering operation, calculating $\hat{y}_D(k)$. The constant D should be chosen as large as possible to ensure a low disturbance term $|\rho_{b_D}(k)|$. However, increasing D also increases the memory requirement for storing old samples of the signals.

4. SIMULATIONS

Speech signals sampled at 8 kHz, shown in figure 2, were used in the simulations. The echo signal $d(k)$ was generated by convolution with a $N = 2000$ coefficient impulse

Fig. 2. Signals used in the simulations. Plot (a) shows the echo signal $d(k)$ and plot (b) shows the near-end speech signal $s(k)$.

response measured in a normal office. A stationary noise signal $w(k) \sim \mathcal{N}(0, 10^{-6})$ was added to $s(k)$, forming the near-end disturbance signal $n(k) = s(k) + w(k)$. The final microphone signal was then calculated as $y(k) = d(k) + n(k)$. Exponential recursive weighting was used to obtain approximations of the ensemble averages used in the transfer logic conditions [5] as

$$\begin{aligned}\hat{r}_{\hat{y}e_b}(k) &= \lambda \hat{r}_{\hat{y}e_b}(k-1) + (1-\lambda)\hat{y}(k)e_b(k) \\ \hat{r}_{\hat{y}e_f}(k) &= \lambda \hat{r}_{\hat{y}e_f}(k-1) + (1-\lambda)\hat{y}(k)e_f(k) \\ \hat{r}_{\hat{y}y}(k) &= \lambda \hat{r}_{\hat{y}y}(k-1) + (1-\lambda)\hat{y}(k)y(k) \\ \hat{r}_{\hat{y}e_{bD}}(k) &= \lambda \hat{r}_{\hat{y}e_{bD}}(k-1) + (1-\lambda)\hat{y}_D(k)e_b(k-D) \\ \hat{r}_{\hat{y}e_{fD}}(k) &= \lambda \hat{r}_{\hat{y}e_{fD}}(k-1) + (1-\lambda)\hat{y}_D(k)e_f(k-D) \\ \hat{r}_{\hat{y}y_D}(k) &= \lambda \hat{r}_{\hat{y}y_D}(k-1) + (1-\lambda)\hat{y}_D(k)y(k-D)\end{aligned}\quad (8)$$

where the forgetting factor was set to $\lambda = 0.9995$.

Equations (3) and (6) can then be expressed in terms of the estimates in (8), with a regularization parameter $\epsilon_a = 0.001$ introduced to avoid division by zero, yielding the filter deviation estimators used in the simulations as

$$\begin{aligned}\hat{\nu}_b(k) &= \frac{\hat{r}_{\hat{y}e_b}(k)}{|\hat{r}_{\hat{y}y}(k) + \epsilon_a|}, \\ \hat{\nu}_f(k) &= \frac{\hat{r}_{\hat{y}e_f}(k)}{|\hat{r}_{\hat{y}y}(k) + \epsilon_a|}, \\ \hat{\nu}_{bD}(k) &= \frac{\hat{r}_{\hat{y}e_{bD}}(k)}{|\hat{r}_{\hat{y}y_D}(k) + \epsilon_a|}, \\ \hat{\nu}_{fD}(k) &= \frac{\hat{r}_{\hat{y}e_{fD}}(k)}{|\hat{r}_{\hat{y}y_D}(k) + \epsilon_a|}.\end{aligned}\quad (9)$$

The delay constant was set to $D = 32$.

Two different cases were simulated, a double-talk scenario where the microphone signal was calculated as de-

scribed previously with double-talk occurring from 19 seconds to 26 seconds and an echo-path change scenario where the impulse response was changed at 19 seconds by moving all filter coefficients one step to the left (i.e. $h_{i-1} = h_i$, $i = \{1, \dots, N-1\}$, $h_{N-1} = 0$). In the echo-path change scenario, no near-end speech was present (i.e. $s(k) = 0$).

To isolate the performance of the deviation measures, the two-path transfer logic was set to transfer the background filter into the foreground filter at all instances from 0 seconds up to 19 seconds and thereafter halt the update of the foreground filter. This means that the background filter and the foreground filter will be identical from 0 seconds to 19 seconds and after that the background filter will either diverge (in the double-talk scenario) or converge to the new echo-path (in the echo-path change scenario). Since the foreground filter is fixed after 19 seconds, it will either stay converged (in the double-talk scenario) or be misadjusted to the new echo-path (in the echo-path change scenario).

Figure 3 shows the results from the double-talk scenario. Plot (a) shows the output from the deviation estimator described in [5], plot (b) shows the output from the proposed deviation estimator and plot (c) shows the actual normalized deviation calculated as $\mathcal{D}_f = \|\mathbf{h} - \hat{\mathbf{h}}_f(k)\|_2 / \|\mathbf{h}\|_2$ (foreground filter) and $\mathcal{D}_b = \|\mathbf{h} - \hat{\mathbf{h}}_b(k)\|_2 / \|\mathbf{h}\|_2$ (background filter), respectively. Plot (d) shows a magnified version of plot (a). All plotted signals are truncated so that their maximum value cannot exceed 1 for the sake of clarity. It can be seen from plots (a) and (d) that the deviation estimator in [5] does not accurately describe the true deviation shown in plot (c) during double-talk. Further, the background filter even seems to perform better than the foreground filter occasionally, even though it is severely misadjusted. The proposed deviation estimator in plot (b), on the other hand, does not suffer from this problem and models the true deviation in plot (c) much better.

Figure 4 shows the results from the echo-path change scenario. As can be seen, the deviation estimator in [5] shown in plot (a) and the proposed deviation estimator in plot (b), both show similar behavior, i.e. both the foreground- and background filter deviation rise after the echo path change and as the background filter converges, the estimation of the background filter deviation falls.

Thus, the simulations demonstrate that the proposed deviation estimator is more robust than the deviation estimator in [5] during double-talk, while having the same desirable properties after an echo-path change.

5. CONCLUSIONS

This paper has proposed a deviation estimator targeted for two-path echo cancellation. The proposed estimator is an extension of the deviation estimator presented in [5] and is showing similar desirable properties in an echo-path change situation in addition to greatly improved robustness during double-talk. The cost of the improved robustness is increased

Fig. 3. Adaptive filter deviation estimates in a double-talk situation. Plot (a) shows the output from the estimator used in [5], plot (b) shows the output from the proposed estimator, and plot (c) shows the true deviation, respectively. Plot (d) shows a magnified version of plot (a). By comparing plots (a), (d) and (c), it can be seen that output from the estimator in [5] is not reliable during double-talk.

computational complexity (an additional filtering operation). Comparisons between the original deviation estimator and the proposed estimator have been made analytically and through simulations.

6. REFERENCES

- [1] D.L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Transactions on Communications*, vol. COM-26, pp. 647–653, May 1978.
- [2] J. Benesty, D.R. Morgan, and J.H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 168–172, March 2000.

Fig. 4. Adaptive filter deviation estimates in an echo-path change situation. Plot (a) shows the output from the estimator used in [5], plot (b) shows the output from the proposed estimator, plot (c) shows the true deviation, respectively. It can be seen that both estimators show similar desirable properties after an echo-path change.

- [3] T. Gansler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Transactions on Communication*, vol. 44, pp. 1421–1427, November 1996.
- [4] 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.
- [5] M.A. Iqbal and S.L. 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*, 2007, pp. 167–170.
- [6] 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.
- [7] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 4th edition, 2002.
- [8] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.