

ROBUST LOW-COMPLEXITY TRANSFER LOGIC FOR TWO-PATH ECHO CANCELLATION

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

^aBlekinge Institute of Technology, School of Engineering, SE-37179, Karlskrona, Sweden.

^bLimes Audio AB, Box 7961, SE-90719, Umeå, Sweden.

ABSTRACT

A well used approach for echo cancellation is the two-path method, where two adaptive filters in parallel are utilized. Typically, one filter is continuously updated, and when this filter is considered better adjusted to the echo-path than the other filter, the coefficients of the better adjusted filter is transferred to the other filter. When this transfer should occur is controlled by the transfer logic.

This paper proposes transfer logic that is both more robust and more simple to tune, owing to fewer parameters, than the conventional approach. Extensive simulations show the advantages of the proposed method.

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

1. INTRODUCTION

A common approach in many speech communication applications where echo arise is to use the two-path adaptive filter structure [1], consisting of two echo cancellation filters, here denoted the *background filter* (BG filter) and the *foreground filter* (FG filter), respectively. The BG filter is continuously updated and the FG filter, which is producing the echo cancelled output, is fixed until the BG filter is considered to be better adjusted to the echo-path than the FG filter. When this occurs, the BG filter coefficients are transferred to the FG filter. Fundamental to the two-path filter structure is the *transfer logic*, which determines when the BG filter coefficients should be transferred to the FG filter. In the original two-path filter approach [1] several conditions are used for controlling the FG filter update. The main condition is that the output error magnitude of the BG filter must be less than that of the FG filter for a FG filter update to take place. However, due to cancellation of local speech during double-talk, this conditions is sometimes fulfilled despite the BG filter being severely mis-adjusted [2]. Additional and alternative transfer logic control conditions have also been proposed [2, 3].

This paper presents a simple, yet robust transfer logic approach for two-path echo cancellation using the delay based

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

filter deviation measure presented in [4] and the delay based double-talk detector presented in [5]. A more efficient way of calculating the delay based measures than in the original papers is also presented. It is shown through extensive simulations that the proposed transfer logic is more robust to erroneous FG filter updating during double-talk than previous approaches.

2. TWO-PATH FILTERING

The two-path filtering scheme considered in this paper constitutes a constantly adapting BG filter and a fixed FG filter producing the echo cancelled output, as explained in the previous section. In this paper, as in [1, 2, 3, 4], the normalized least mean square (NLMS) is used to update the BG filter 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 $e_b(k)$ is the BG filter error signal, $\hat{\mathbf{h}}_b(k) = [\hat{h}_{b_0}(k), \hat{h}_{b_1}(k), \dots, \hat{h}_{b_{N-1}}(k)]^T$ is the adaptive BG filter of length N , $x(k)$ is the driving signal fed to the echo-path (e.g. loudspeaker signal in case of acoustic echo cancellation (AEC), or signal fed to the telephone network in case of line echo cancellation (LEC)), $y(k)$ is the echo contaminated input signal (microphone signal in case of AEC, or signal from the telephone network in case of LEC), $\mathbf{x}(k) = [x(k), x(k-1), \dots, x(k-N+1)]^T$ is the regressor vector, μ is the step-size control variable, ϵ is a regularization term to avoid division by zero and k is the sample index. $[\cdot]^T$ denotes transpose.

The FG 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

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

and is updated by copying the BG filter coefficients, i.e. $\hat{\mathbf{h}}_b(k)$ into the FG filter $\hat{\mathbf{h}}_f(k)$. This is performed when the BG filter is considered to be better adjusted to the echo-path than the FG filter.

The process of determining when the FG filter should be updated is controlled by the transfer logic, which constitutes a set of conditions which has to be fulfilled for an update to take place. Typical transfer logic conditions are [1, 2, 3]

1. $\sigma_x^2(k) > T_1$ (sufficient excitation energy must exist)
2. $\sigma_y^2(k) > T_2$ (sufficient echo/near-end signal energy must exist)
3. $\frac{\sigma_{e_f}^2(k)}{\sigma_{e_b}^2(k)} > T_3$ (the BG filter must produce lower output error than the FG filter)
4. $\frac{\sigma_x^2(k)}{\sigma_b^2(k)} > T_4$ (sufficient echo cancellation and acoustic isolation must be present)

where T_1, T_2, T_3 and T_4 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 corresponding signals.

Additional transfer-logic conditions can be double-talk detectors of either Geigel-type [1] or based on the normalized cross-correlation [3] according to

$$1 - \frac{r_{ye_b}(k)}{\sigma_y^2(k)} > T_5, \quad (3)$$

where $r_{ye_b}(k) = \mathbb{E}[y(k)e_b(k)]$ and T_5 is a threshold. $\mathbb{E}[\cdot]$ denotes expectation (ensemble average; which in practice is approximated using time-averaging). Conditions related to the filter misalignment has also been presented [2, 3]. In [2], an artificial delay of L was introduced in the signal path of $y(k)$, causing the L first coefficients of the impulse response to be zero. This means that the first L coefficients of the impulse response are known (to be zero) and since the misalignment spreads over the whole filter [2], an estimate of the total filter misalignment can be made. While the solution in [2] works well for a fullband echo canceller, problems arise when trying to implement it for subband echo cancellation. The reason is that although the fullband echo-path is causal, the subband filters are not, due to temporal spreading, see [6] and references therein. Another filter misalignment condition has been proposed in [3] as

$$\left| \frac{r_{\hat{y}_f e_f}(k)}{r_{\hat{y}_f y}(k)} \right| > \left| \frac{r_{\hat{y}_b e_b}(k)}{r_{\hat{y}_b y}(k)} \right| \quad (4)$$

where $r_{\hat{y}_f e_f}(k) = \mathbb{E}[\hat{\mathbf{h}}_f(k)^T \mathbf{x}(k)e_f(k)]$, $r_{\hat{y}_f y}(k) = \mathbb{E}[\hat{\mathbf{h}}_f(k)^T \mathbf{x}(k)y(k)]$, $r_{\hat{y}_b e_b}(k) = \mathbb{E}[\hat{\mathbf{h}}_b(k)^T \mathbf{x}(k)e_b(k)]$ and $r_{\hat{y}_b y}(k) = \mathbb{E}[\hat{\mathbf{h}}_b(k)^T \mathbf{x}(k)y(k)]$.

3. PROPOSED APPROACH

First of all, it should be noted that the transfer logic conditions in [1, 2, 3] all contain numerous thresholds and require careful tuning. This can be problematic if the echo canceller is expected to function well in different environments. Further, there are some problems with the previous approaches. Perhaps most notably is the problem of the filter misalignment measure in [3] (Equation (4) in this paper). This problem was illustrated and discussed in [4]. Moreover, in [4] an improved filter misalignment measure was also presented, but only as a stand-alone measure, i.e. not as a part of a complete two-path transfer logic solution.

Here, the improved deviation measure in [4] is incorporated in a complete two-path transfer logic solution together with the double-talk detector presented in [5]. It is also shown that the deviation measure, as well as the double-talk measure, can be calculated much more efficiently than originally presented.

It has been shown that the deviation measure in Equation (4) is not reliable for the BG filter during double-talk and an improved (BG filter) deviation measure was presented as [4]

$$\nu_{b_D}(k) = \left| \frac{r_{\hat{y}_{b_D} e_{b_D}}(k)}{r_{\hat{y}_{b_D} y}(k)} \right|, \quad (5)$$

where $r_{\hat{y}_{b_D} e_{b_D}}(k) = \mathbb{E}[\hat{y}_{b_D}(k)(y(k) - \hat{y}_{b_D}(k))]$, $r_{\hat{y}_{b_D} y}(k) = \mathbb{E}[\hat{y}_{b_D}(k)y(k)]$, $\hat{y}_{b_D}(k) = \hat{\mathbf{h}}_b(k+D)^T \mathbf{x}(k)$ and D is a delay constant. The purpose of the delay is to avoid cancellation of near-end speech by the constantly updating BG filter. For more details regarding this matter, the reader is referred to [4] and [7].

The first proposed transfer logic condition is

$$\sigma_x^2(k) > T_1 \quad (6)$$

according to the traditional scheme (see previous section). This condition is also coupled with the NLMS update of the BG filter so that Equation (6) is true if the BG filter is adapted.

The deviation measure in Equation (5) is used in the second proposed transfer logic condition as

$$\left| \frac{r_{\hat{y}_f e_f}(k)}{r_{\hat{y}_f y}(k)} \right| > \left| \frac{r_{\hat{y}_{b_D} e_{b_D}}(k)}{r_{\hat{y}_{b_D} y}(k)} \right|. \quad (7)$$

Furthermore, the third transfer logic condition involves double-talk detection and is [5]

$$1 - \frac{r_{ye_b D}}{\sigma_y^2} > T_5, \quad (8)$$

where $r_{ye_b D} = \mathbb{E}[y(k)(y(k) - \hat{y}_{b_D}(k))]$. (The reader is referred to [5] for more details.)

The fourth and final condition of the proposed transfer logic is a straight-forward output error comparison according to

$$\sigma_{e_f}^2(k) > \sigma_{e_{b_D}}^2(k), \quad (9)$$

where $\sigma_{e_{b_D}}^2(k) = \mathbb{E}[(y(k) - \hat{y}_{b_D}(k))^2]$ i.e. the squared output error magnitude of the (delayed) BG filter must be lower than that of the FG filter.

Hence, the proposed transfer logic involves only three tuning parameters: T_1, T_5 and D .

3.1. Practical considerations and complexity

The essence of the proposed transfer logic is the calculation of the echo estimate $\hat{y}_{b_D}(k)$ which should be decoupled as much as possible from the current adaptive filter update, using the delay D . It has been shown that a negative D achieves this decoupling better than a positive D [4] for the filter deviation measure. A straight-forward approach to calculate $\hat{y}_{b_D}(k)$ is

to store all old filters and perform a filtering operation. However, a much more efficient solution is presented below.

Inserting the NLMS update Equation (1) into the expression for $\hat{y}(k)$ gives

$$\begin{aligned}\hat{y}_b(k) &= \mathbf{x}(k)^T \hat{\mathbf{h}}_b(k) \\ &= \mathbf{x}(k)^T \left(\hat{\mathbf{h}}_b(k-1) + \beta(k-1)\mathbf{x}(k-1) \right) \\ &= \hat{y}_{-1}(k) + \beta(k-1)\mathbf{x}(k)^T \mathbf{x}(k-1),\end{aligned}\quad (10)$$

where $\hat{y}_{-1}(k) = \mathbf{x}(k)^T \hat{\mathbf{h}}_b(k-1)$ and $\beta(k) = \mu \frac{e_b(k)}{\mathbf{x}(k)^T \mathbf{x}(k) + \epsilon}$. From Equation (10) it can be seen that continuing to expand the expression using the NLMS update Equation (1) yields

$$\hat{y}_b(k) = \hat{y}_{b_D}(k) + \sum_{i=1}^{|D|} \beta(k-i)\alpha_i(k),\quad (11)$$

where $\alpha_i(k) = \mathbf{x}(k)^T \mathbf{x}(k-i)$. Thus it is clear that $\hat{y}_{b_D}(k)$ can be calculated as

$$\hat{y}_{b_D}(k) = \hat{y}_b(k) - \beta(k)^T \boldsymbol{\alpha}(k),\quad (12)$$

where the vectors $\beta(k) = [\beta(k-1), \beta(k-2), \dots, \beta(k-|D|)]^T$ and $\boldsymbol{\alpha}(k) = [\alpha_1(k), \alpha_2(k), \dots, \alpha_{|D|}(k)]^T$ are both of length $|D|$.

Since the echo estimate $\hat{y}_b(k)$ is calculated in the adaptive filtering update procedure, what remains for obtaining $\hat{y}_{b_D}(k)$ is the scalar product $\beta(k)^T \boldsymbol{\alpha}(k)$. This scalar product requires $|D|$ multiplications and additions. The vector $\beta(k)$ can be obtained at practically no additional computational cost, since $\beta(k)$ is calculated in the NLMS update Equation (1) just has to be delayed/stored in the vector. The elements of the vector $\boldsymbol{\alpha}(k)$ can be obtained recursively at low computational cost as $\alpha_i(k) = \alpha_i(k-1) - x(k-N)x(k-N-i) + x(k)x(k-i)$, i.e. using just two multiplications, one addition and one subtraction. In total, this means that calculating $\hat{y}_{b_D}(k)$ using the proposed method requires $3|D|$ multiplications and the same number of additions/subtractions. This should be compared to the straight-forward approach in [4], requiring N multiplications and the same number of additions. Since typically $|D| \ll N$, the proposed approach saves significant computational cost.

4. SIMULATIONS AND RESULTS

The performance of the proposed transfer logic was verified through simulation with speech signals sampled at 8 kHz. Evaluation was performed for a DTF-modulated polyphase filterbank [6] with 32 subbands and a decimation ratio of 16. The number of prototype filter coefficients was 128. Each subband contained an individual setup of FG and BG filters with $N = 64$, and the BG filter was constantly updated when there was enough driving signal energy using the NLMS with variable step-size according to [8]. The two-path transfer logic of each subband was independent from the other subbands. This setup was used for both compared methods.

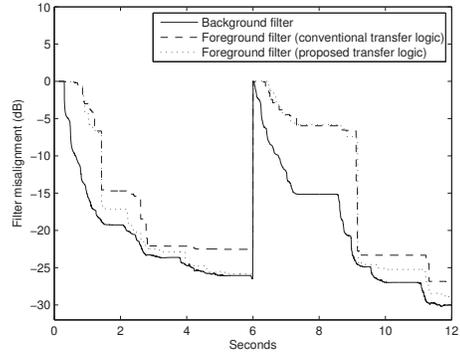


Fig. 1. Misalignment of the background filter and the foreground filters of both compared methods. An echo-path change occurs after 6 s.

Evaluation of the proposed approach was made through comparison with the transfer logic presented in [3] which basically is the same as in [1] (see Section 2) with the addition of the filter misalignment condition in Equation (4) and the double-talk detector in Equation (3) instead of a Geigel detector. This approach was denoted the *conventional transfer logic* and the thresholds were $T_1 = 10^{-8}$ (corresponding to the threshold for BG filter NLMS updating), $T_2 = 10^{-10}$, $T_3 = 2$, $T_4 = 2$ and $T_5 = 0.95$. The proposed approach uses only a total of four conditions, Equations (6), (7), (8) and (9), with three parameters: T_1 and T_5 which were set, as for the conventional transfer logic, to 10^{-8} and 0.95, respectively and $D = -4$, corresponding to a “full-band delay setting” of -64 with the selected decimation ratio [7].

For a BG-to-FG filter transfer to take place, the transfer logic conditions must be true for 50 consecutive subband samples, corresponding to 100 ms. This was used for both the conventional and the proposed method.

The ensemble averages used in the transfer logic conditions were estimated through time-averaging exponential recursive weighting [3, 4], e.g.

$$\hat{r}_{ye_b}(k+1) = \lambda \hat{r}_{ye_b}(k) + (1-\lambda)y(k)e_b(k)\quad (13)$$

and similarly for all other averages. The forgetting factor λ was set to 0.95.

To obtain the signal $y(k)$, the driving signal $x(k)$ was filtered with a known impulse response $\mathbf{h} = [h_0, h_1, \dots, h_{N_f-1}]^T$ of length $N_f = 1024$ obtained through measurement in a small office. A flat spectrum noise signal was also added to $y(k)$, giving an echo-to-noise ratio of approximately 30 dB.

Two sets of scenarios were used: an echo-path change situation and double-talk. In both scenarios, the same driving speech signal of length 12 s was used. The performance of the different transfer logics was evaluated using the corresponding fullband FG filter misalignment evaluated as $10 \log_{10} \|\mathbf{h} - \hat{\mathbf{f}}_f(k)\|^2 / \|\mathbf{h}\|^2$ where $\hat{\mathbf{f}}_f(k)$ is a fullband filter constructed from all subband FG filters.

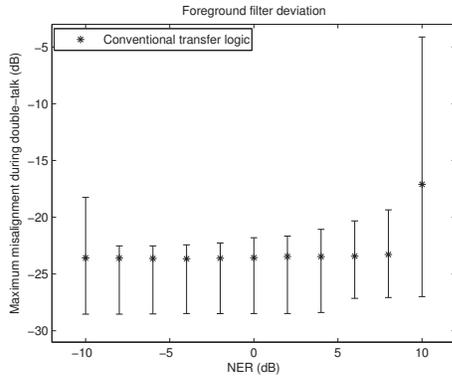


Fig. 2. Ensemble averages, together with the maximum and minimum, of the maximum foreground filter misalignment during double-talk for different near-end-to-echo ratios (NER) for the conventional transfer logic.

4.1. Echo-path change

In this scenario, the known impulse response \mathbf{h} was changed after 6 s by shifting all filter coefficients one step to the left (i.e. $h_{i-1} = h_i$, $i = \{1, \dots, N_f - 1\}$, $h_{N_f-1} = 0$). The resulting filter misalignment for the BG filter, the FG filter with the conventional transfer logic and the FG filter with the proposed transfer logic are shown in Figure 1. It can be seen that the proposed transfer logic allows the FG filter to track the adaptation of the BG filter slightly better than the conventional approach.

4.2. Double-talk

For the double-talk evaluation, a similar setup as in [5] where four different speech signals (two male and two female) of approximately 2 s each were used as near-end speech. The near-end speech was set to occur at four different positions in time within the 12 s far-end speech, yielding a total of 16 simulations. The gain of the near-end speech was also varied to give a range of different near-end speech to echo ratios (NER).

The transfer logic performance was evaluated by comparing the FG filter misalignment during double-talk. For each of the 16 simulations, the maximum FG filter deviation during double-talk was noted. The ensemble mean of the maximum FG filter deviation during double-talk was then calculated, together with the overall maximum and minimum misalignment. This procedure was carried out for a range of NERs and the results are shown in Figures 2 and 3. By comparing the figures it can clearly be seen that the proposed method is more robust than the conventional transfer logic as the FG filter is constantly kept at a lower level during double-talk.

5. CONCLUSIONS

An improved transfer logic scheme for two-path echo cancellation, based on the delay-based deviation measure in [4]

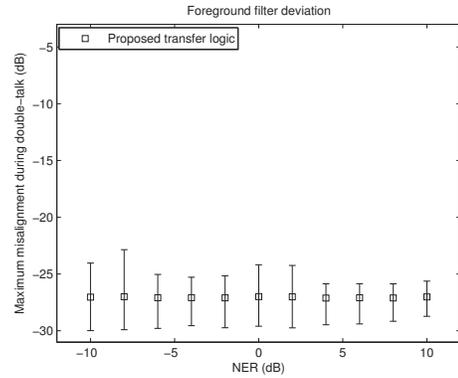


Fig. 3. Ensemble averages, together with the maximum and minimum, of the maximum foreground filter misalignment during double-talk for different near-end-to-echo ratios (NER) for the proposed transfer logic.

and the delay based double-talk detector in [5], has been proposed. Extensive simulations have shown that the proposed transfer logic is more robust to double-talk than the conventional method, while also exhibiting slightly improved performance during a change of the echo-path.

6. 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] F. Lindstrom, C. Schüldt, and I. Claesson, “An improvement of the two-path algorithm transfer logic for acoustic echo cancellation,” *IEEE Trans. on Audio, Speech and Language Proc.*, vol. 15, pp. 1320–1326, May 2007.
- [3] M.A. Iqbal, J.W. Stokes, and S.L. Grant, “Normalized double-talk detection based on microphone and AEC error cross-correlation,” in *Proc. of IEEE International Conf. on Multimedia and Expo*, July 2007, pp. 360–363.
- [4] C. Schüldt, F. Lindstrom, and I. Claesson, “An improved deviation measure for two-path echo cancellation,” in *Proc. of IEEE ICASSP*, March 2010, pp. 305–308.
- [5] C. Schüldt, F. Lindstrom, and I. Claesson, “A delay-based double-talk detector,” *Submitted to IEEE Trans. on Audio, Speech and Language Proc.*
- [6] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.
- [7] C. Schüldt, F. Lindstrom, and I. Claesson, “Evaluation of an improved deviation measure for two-path echo cancellation,” in *Proc. of IWAENC*, September 2010.
- [8] J. Benesty, H. Rey, L.R. Vega, and S. Tressens, “A non-parametric VSS NLMS algorithm,” *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 581–584, October 2006.