

ROBUST AND ELEGANT, PURELY STATISTICAL ADAPTATION OF ACOUSTIC ECHO CANCELER AND POSTFILTER

Gerald Enzner and Peter Vary

Institute of Communication Systems and Data Processing (IND)
Aachen University (RWTH), D-52056 Aachen, Germany

Phone: +49-241-8026960 E-mail: {enzner, vary}@ind.rwth-aachen.de

ABSTRACT

We present an extremely simple and effective signal processing solution to the acoustic echo control problem. The approach is based on the concept of synchronous statistical adaptation of an acoustic echo canceler and a postfilter for residual echo suppression. The required convergence state of the echo canceler is estimated by a new statistical element of our system. Double talk detection is not required explicitly. The whole solution combines an excellent output signal quality with a high degree of elegance and simplicity.

1. INTRODUCTION AND SYSTEM OVERVIEW

Fast and robust adaptation of the acoustic echo control (AEC) unit in hands-free communication systems is often considered to be difficult in non-stationary noisy environments. The adaptation of many AEC units explicitly relies on the sophisticated detection and classification of different acoustic events (e.g. room impulse response changes or double talk situations). In our paper we will demonstrate that the AEC problem can be solved in a simpler, purely statistical framework, at least if the AEC unit relies on the filter arrangement with acoustic echo canceler and postfilter for residual echo suppression as shown in Figure 1.

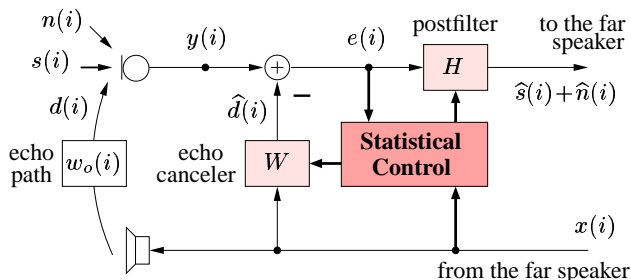


Fig. 1. Acoustic feedback and the arrangement of adaptive FIR filters W and H to perform feedback attenuation.

The local microphone signal at sampling time index i , $y(i) = s(i) + n(i) + d(i)$, is additively composed of clean near speech $s(i)$, local background noise $n(i)$, and acoustic echo $d(i)$. The echo signal is due to the acoustic echo path with impulse response $w_o(i)$ from the local loudspeaker to the local microphone. In this paper we assume that the length of the FIR filter W sufficiently models the length of the echo path. Nevertheless, the echo canceler leaves a residual echo $b(i) = d(i) - \hat{d}(i)$ due to the time-varying nature of the echo path and the presence of observation noise $s(i) + n(i)$. The residual echo in the error signal $e(i) = s(i) + n(i) + b(i)$ is further reduced by the FIR postfilter H .

Recently it has been shown that there is a simple and close relation between the optimum statistical adaptation processes of echo canceler and postfilter [1, 2]. This concept is referred to as joint control or synchronous adaptation of echo canceler and postfilter. This technique achieves a very consistent interaction between both filters: The adaptation control triggers a postfilter attenuation in the sending path of the hands-free system only when the echo canceler is temporarily not able to reduce the acoustic echo sufficiently. That happens for example after double talk situations or after strong echo path changes. With this intelligent interaction of echo canceler and postfilter, distortions of the useful signal in the sending path of the system can be almost avoided. In Section 2 of our paper, the details of this important concept will be recalled and applied in the Discrete Fourier Transform (DFT) domain.

The implementation of the synchronous adaptation concept requires knowledge about the convergence state of the echo canceler (i.e. the frequency-dependent system distance between the echo canceler and the echo path). It has been observed in other work, too, that the estimation of this convergence state is the major difficulty in acoustic echo control systems [1, 2, 3]. If, however, the concept of synchronous adaptation is applied, we claim that it is sufficient to use a simple statistical convergence state estimator in order to support the optimum adaptation of both filters. Our estimator will be derived in Section 3 of the paper. Results produced with our system are discussed in Section 4.

This work is supported by Nokia Research Center, Tampere, Finland, and Nokia Mobile Phones, Bochum, Germany.

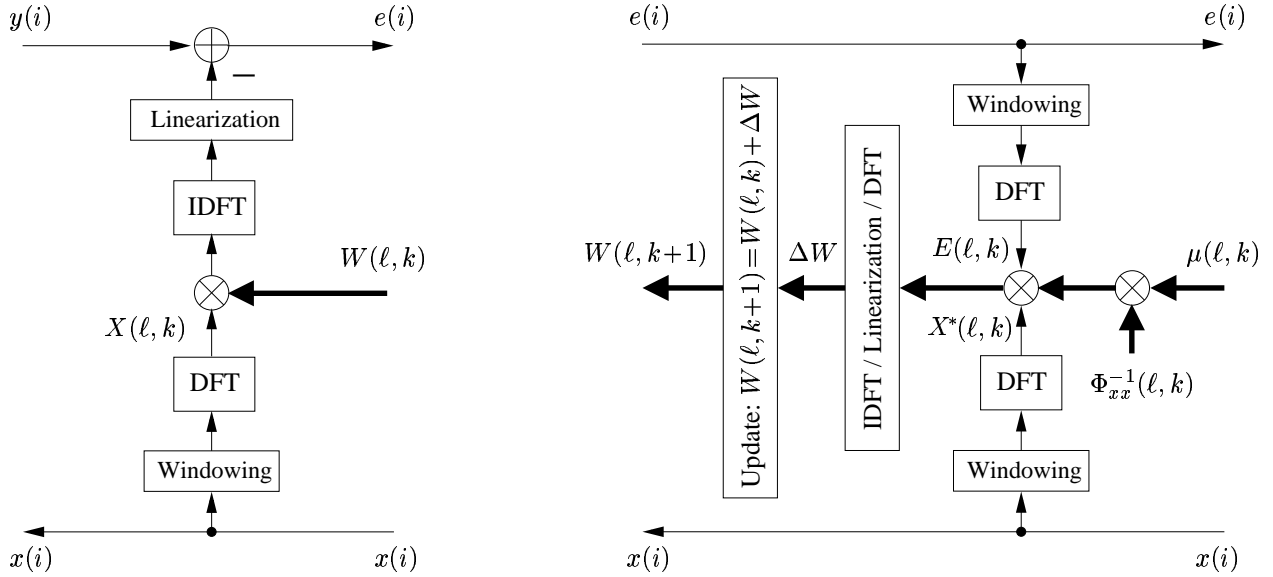


Fig. 2. Schematics of the filtering path (left picture) and the adaptation loop (right picture) of an echo canceler W realized with the FDAF algorithm.

2. BUILDING BLOCKS OF THE AEC UNIT

We discuss the statistical optimization of echo canceler and postfilter in our AEC unit. That clarifies the need of the adaptation process for a convergence analysis of the echo canceler.

2.1. Frequency-Domain Adaptive Filter (FDAF)

The frequency-domain adaptive filter [4, 5] has become a first choice in acoustic echo cancellation because of its ability to realize high-order adaptive filters W with a high convergence rate and moderate computational complexity. The two basic modules of the FDAF – filtering and adaptation – are shown in Figure 2.

The FDAF uses the fast convolution/correlation technique to implement an acoustic echo canceler in the DFT domain. The signal frames are obtained by the “Windowing” operation, $k \in \mathbb{Z}$ is the *frame time index*, and $\ell \in \{0, 1, \dots, M-1\}$ denotes the *discrete frequency index* for a DFT length of M . The “Linearization” (according to the overlap/save method) removes cyclic convolution/correlation components produced by DFT and IDFT.

The filtering stage in the left picture of Figure 2 applies a set of coefficients $W(\ell, k)$ to compute an estimate $\hat{d}(i)$ of the acoustic echo $d(i)$. Then the error signal $e(i)$ is given to the adaptation stage in the right picture. The cross-correlation between $e(i)$ and $x(i)$ is computed to improve the set of coefficients $W(\ell, k)$ gradually. The time- and frequency-dependent normalization of the gradient by the sampled power spectral density¹ (PSD) $\Phi_{xx}(\ell, k)$ of the

input signal $x(i)$ is in fact responsible for the fast convergence rate of the FDAF. The step-size $\mu(\ell, k)$ guarantees the robustness of the LMS type adaptive filter W in the presence of observation noise $s(i) + n(i)$.

The optimum step-size $\mu(\ell, k)$ for the FDAF was derived in [3] as the ratio of the PSDs $\Phi_{bb}(\ell, k)$ and $\Phi_{ee}(\ell, k)$ corresponding to the residual echo signal $b(i)$ and the error signal $e(i)$:

$$\mu(\ell, k) = \frac{\Phi_{bb}(\ell, k)}{\Phi_{ee}(\ell, k)} = \frac{|G(\ell, k)|^2 \Phi_{xx}(\ell, k)}{\Phi_{ee}(\ell, k)}. \quad (1)$$

$|G(\ell, k)|^2$ is the statistical expectation of the sampled residual echo power transfer function and basically defines the convergence state of the echo canceler in the DFT domain. The step-size in (1) minimizes the convergence state $|G(\ell, k+1)|^2$ at the frame index $k+1$, given the convergence state $|G(\ell, k)|^2$ at the current frame index k .

This type of minimum mean-square error (MMSE) optimization highly motivates the usage of a Wiener postfilter for residual echo suppression. Our estimator for $|G(\ell, k)|^2$ will be presented in Section 3.

2.2. Wiener Postfilter in the DFT Domain

In the original papers [6, 7] the postfilter H was proposed for combined residual echo and background noise suppression. Here we emphasize the important special case of residual echo suppression: Consider Figure 1 in which a linear filter H shall be applied to the echo compensated signal $e(i)$ such that $\hat{s}(i) + \hat{n}(i)$ approximates $s(i) + n(i)$ in the MMSE sense.

¹The PSD of a non-stationary signal is a function of time and frequency.

The MMSE solution can be computed approximately by spectral weighting of the DFT coefficients $E(\ell, k)$ corresponding to the echo compensated signal $e(i)$, i.e.

$$\widehat{S}(\ell, k) + \widehat{N}(\ell, k) = H(\ell, k)E(\ell, k), \quad (2)$$

if the postfilter weights are determined according to the Wiener rule as

$$\begin{aligned} H(\ell, k) &= \frac{\Phi_{ee}(\ell, k) - \Phi_{bb}(\ell, k)}{\Phi_{ee}(\ell, k)} \\ &= \frac{\Phi_{ee}(\ell, k) - |G(\ell, k)|^2 \Phi_{xx}(\ell, k)}{\Phi_{ee}(\ell, k)}. \end{aligned} \quad (3)$$

Note that the DFT coefficients $E(\ell, k)$ in (2) are computed by the FDAF already. Finally, the synthesis of the postfilter output $\widehat{s}(i) + \widehat{n}(i)$ requires an IDFT of $\widehat{S}(\ell, k) + \widehat{N}(\ell, k)$ for each frame index k and the overlap/save method to recombine an output signal stream.

2.3. Synchronous Adaptation

The concept of synchronous adaptation directly follows from the MMSE optimization of both echo canceler and postfilter: From the solutions in Equations (1) and (3) it can be easily seen that the optimum step-size of the echo canceler depends on the same parameters than the optimum postfilter weights and both are exactly complementary [1, 2]:

$$\mu(\ell, k) + H(\ell, k) = 1. \quad (4)$$

This fundamental equation establishes the close relation between the optimum statistical adaptation of echo canceler and postfilter. This relation can be exploited to synchronize the echo canceler W and the postfilter H very effectively and therefore leads to a simplified structure of AEC units. The synchronization further achieves a very consistent interaction between echo canceler and postfilter and therefore results in an excellent output signal quality of the AEC unit.

We further observe from Equations (1) and (3) that the unknown convergence state $|G(\ell, k)|^2$ essentially controls the optimum statistical adaptation of echo canceler and postfilter. Despite the relation (4), the convergence state is still absolutely necessary to compute either the echo canceler step-size or the postfilter weights. This convergence state is, however, not available (measurable) explicitly and therefore has to be estimated from the known signals.

3. A SIMPLE, PURELY STATISTICAL CONVERGENCE STATE ESTIMATOR

In this section we propose a very simple statistical estimation technique for the convergence state required in the algorithm. The proposed technique fills an important gap in the control theory of AEC units and overcomes the more heuristic nature of other approaches.

3.1. Statistical Convergence Analysis

In the following derivation we consider the FDAF approximately as a normalized LMS type adaptive filter with one tap-weight in each frequency bin. A statistical convergence analysis for the LMS algorithm was performed for example in [8, Equation 34]. Thus, given the step-size $\mu(\ell, k)$, the dynamic convergence behavior of the FDAF is described by a first order difference equation for the residual echo PSD $\Phi_{bb}(\ell, k)$:

$$\begin{aligned} \Phi_{bb}(\ell, k+1) &= \mu^2(\ell, k) \cdot \Phi_{(s+n)(s+n)}(\ell, k) \\ &+ \Phi_{bb}(\ell, k) \cdot (1 - \mu(\ell, k))^2 + |\Delta W_o(\ell, k)|^2 \Phi_{xx}(\ell, k). \end{aligned} \quad (5)$$

In contrast to [8, Equation 34], we added the PSD $|\Delta W_o(\ell, k)|^2 \Phi_{xx}(\ell, k)$ to account for time-varying characteristics of the acoustic echo path $w_o(i)$. This approach is motivated by a Markov model for the echo path variations as used in [5]. The statistical parameter $|\Delta W_o(\ell, k)|^2$ is the expectation of the magnitude-squared frequency response corresponding to acoustic echo path changes from frame index k to $k+1$. Our statistical model for the *expected room variations* $|\Delta W_o(\ell, k)|^2$ will be provided in Section 3.2.

Using $\Phi_{ee} = \Phi_{bb} + \Phi_{(s+n)(s+n)}$ Equation (5) can be rewritten as

$$\begin{aligned} \Phi_{bb}(\ell, k+1) &= \mu^2(\ell, k) \cdot \Phi_{ee}(\ell, k) \\ &+ \Phi_{bb}(\ell, k) \cdot (1 - 2\mu(\ell, k)) + |\Delta W_o(\ell, k)|^2 \Phi_{xx}(\ell, k). \end{aligned} \quad (6)$$

Substitution of the optimum step-size from Equation (1) into Equation (6) and normalization of the result by the input signal PSD $\Phi_{xx}(\ell, k)$ yields the following recursion for the convergence state $|G(\ell, k)|^2$ of the echo canceler:

$$\begin{aligned} |G(\ell, k+1)|^2 &= |G(\ell, k)|^2 \cdot (1 - \mu(\ell, k)) \\ &+ |\Delta W_o(\ell, k)|^2. \end{aligned} \quad (7)$$

Equation (7) represents a central result of our paper. It can be seen that the predicted convergence state $|G(\ell, k+1)|^2$ solely depends on the result of the previous iteration, the time- and frequency-dependent forgetting factor $1 - \mu(\ell, k)$, and the statistical model parameter $|\Delta W_o(\ell, k)|^2$ for the expected room variations.

The initialization of the convergence state can be chosen for example as $|G(\ell, 0)|^2 = 0$, and the recursion in Equation (7) converges to the true value of $|G(\ell, k+1)|^2$. Eventually, $|G(\ell, k+1)|^2$ can be easily substituted into Equations (1) and (3) to control the adaptation of echo canceler and postfilter at the frame index $k+1$.

3.2. Proportional Room Variation Model

In our statistical framework we further assume that the expected level of room variations $|\Delta W_o(\ell, k)|^2$ is proportional

to the acoustic coupling $|W_o(\ell, k)|^2$ between loudspeaker and microphone:

$$|\Delta W_o(\ell, k)|^2 = C \cdot |W_o(\ell, k)|^2. \quad (8)$$

The acoustic coupling $|W_o(\ell, k)|^2$ is the expectation of the magnitude-squared frequency response corresponding to the echo path $w_o(i)$. It can be determined, for example, from the last estimated echo path $W(\ell, k) \approx W_o(\ell, k)$. A similar technique has been reported in the context of the PNLMS algorithm [9], where magnitude information about the estimated echo path is used as feedback to control the adaptation step-size. Another option to find magnitude a priori information about the echo path is the utilization of a background filter.

It should be noted that it might be helpful to adjust the time- and frequency-resolution of the expected room variations $|\Delta W_o(\ell, k)|^2$ to the specific application using AEC functionality. We further recommend a fixed constant of proportionality, $0 < C \ll 1$, which is chosen appropriately according to the frame-shift of the AEC algorithm.

4. RESULTS AND BENEFITS

4.1. General Benefits from the Proposed Technique

The main feature of our AEC algorithm is the structural elegance and simplicity: The synchronous adaptation concept can reduce the control effort for AEC units considerably. The reason is that a simple statistical approach is sufficient to estimate the required convergence state. A double talk detection is not required for the stability of the adaptation.

4.2. Simulation Results and Realtime Verification

Test Environment: For the offline simulation, real speech signals $x(i)$ were reproduced by a hands-free loudspeaker in the passenger footwell inside of a car. The acoustic echo $d(i)$ was mixed with various levels of local speech $s(i)$ and car background noise $n(i)$ and recorded by a hands-free microphone located next to the driver's mirror. The acoustic coupling between loudspeaker and microphone was -5 dB. Inside the car, the local talker performed typical movements to simulate acoustic echo path changes.

For the realtime verification, we have so far used a small office room which produces similar reverberation as the car environment. Otherwise the setup was comparable to the offline scenario.

Processing Results: The offline simulation and the realtime scenario basically produced the same results: The adaptive filter W adapts very quickly to the room variations ($\ll 1s$), but a weak residual echo remains audible after the echo canceler. By the postfilter H , the attenuation of acoustic echo is considerably improved (depending on the amount of near speech and background noise). A distortion of the useful signal is hardly ever present.

In the informal subjective evaluation of the whole system according to [10], the speech transmission quality was rated as *excellent* and acoustic echo was *not noticeable*.

5. CONCLUSION

We proposed a very elegant solution to the acoustic echo control problem. It is based on the statistical adaptation of echo canceler and postfilter and it uses a simple statistical estimator for the convergence state of the echo canceler. The adaptation process has been verified to be fast and robust in offline simulations and in the realtime environment.

6. REFERENCES

- [1] E. Hänsler, G. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering," *Signal Processing*, vol. 80, no. 11, pp. 2295–2305, Nov. 2000.
- [2] G. Enzner, R. Martin, P. Vary, "Partitioned Residual Echo Power Estimation for Frequency-Domain Acoustic Echo Cancellation and Postfiltering," *European Trans. Telecommunications*, vol. 13, pp. 103–114, March/April 2002.
- [3] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Processing*, vol. 80, no. 9, pp. 1733–1745, Sept. 2000.
- [4] E. R. Ferrara, "Frequency-domain adaptive filtering," *In: C.F.N. Cowan, P.M. Grant (Eds.), Adaptive Filters, Prentice Hall*, pp. 145–179, 1985.
- [5] S. Haykin, "Adaptive Filter Theory," Prentice Hall, 1996.
- [6] R. Martin, J. Altmann, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction," *Proc. IEEE Intl. Conference on Acoustics, Speech, and Signal Processing*, pp. 3043–3046, May 1995.
- [7] S. Gustafsson, R. Martin, P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, no. 1, pp. 21–32, Jan. 1998.
- [8] T. Claasen, W. Mecklenbräuker, "Comparison of the Convergence of Two Algorithms for Adaptive FIR Digital Filters," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 29, no. 3, pp. 670–678, June 1981.
- [9] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 508–518, Sept. 2000.
- [10] ITU-T, "Recommendation P.832, Subjective performance evaluation of hands-free terminals," May 2000.