Hands–Free Communication: A Unified Concept of Acoustic Echo Cancellation and Residual Echo Suppression

Gerald Enzner
Institute of Communication Systems and Data Processing (IND)
Aachen University (RWTH), D–52056 Aachen, Germany
Phone: +49-241-80-26960, E-mail: enzner@ind.rwth-aachen.de

Introduction

An essential feature of a hands-free communication system is the acoustic echo control (AEC) unit. The need of an AEC unit basically arises from the acoustic echo path with impulse response $h(i)$ from the local loudspeaker to the local microphone. This is illustrated in Figure 1. The local microphone signal at the sampling time index i,

$$y(i) = s(i) + n(i) + d(i) \ , \qquad (1)$$

is additively composed of clean near speech $s(i)$, local background noise $n(i)$, and acoustic echo $d(i)$.

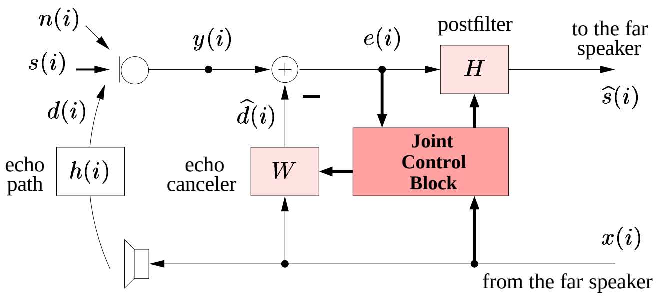

Fig. 1. Acoustic feedback and the arrangement of adaptive FIR filters W and H to perform feedback attenuation.

In most of the cases AEC units contain an acoustic echo canceler W and a postfilter H for residual echo suppression. In a duplex connection the AEC unit is supposed to maximize the attenuation of the feedback $d(i)$ and to minimize the distortion of the useful signal $s(i) + n(i)$. The key to that is the synchronization (joint control) of acoustic echo cancellation and postfiltering. An efficient realization of the AEC unit is possible when both echo cancellation and postfiltering are implemented in the Discrete Fourier Transform (DFT) domain. These facts will be explained in the remainder of the paper.

Building Blocks of an AEC Unit

This section presents an overview of important technical milestones in the field of acoustic echo cancellation and postfiltering:

- Frequency–domain adaptive filter (FDAF) for acoustic echo cancellation [1]. **Ferrara 1985.**
- Wiener postfilter in the DFT domain for combined residual echo and noise reduction [2]. **Gustafsson,Martin,Vary 1998.**
- Optimum time– and frequency–dependent stepsize for the FDAF in the MMSE sense [3]. **Nitsch 2000.**
- Joint control of acoustic echo cancellation and postfiltering [4, 5]. **Hänsler, Schmidt 2000. Enzner, Martin, Vary 2002.**

Acoustic Echo Canceler

The frequency–domain adaptive filter (FDAF) [1] has become a first choice in acoustic echo cancellation because of its ability to realize higher 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 stream of signal samples is processed on a frame by frame basis. Signal frames are obtained by the "Windowing" operation, $k \in \mathbb{Z}$ is the frame time index, and $\ell \in \{0, 1, ..., M-1\}$ denotes the discrete frequency index for a DFT length M. The "Linearization" is useful to remove cyclic convolution/correlation components produced by the DFT/IDFT.

The filtering stage on the left hand side applies a set of coefficients $W(\ell, k)$ to compute an estimate $\widehat{d}(i)$ of the acoustic echo $d(i)$. Then the error signal $e(i)$ is given to the adaptation stage on the right hand side of Figure 2. 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 power spectral density (PSD) $\Phi_{xx}(\ell, k)$ of the input signal $x(i)$ is in fact responsible for the fast convergence rate of the FDAF. The stepsize $\mu(\ell, k)$ guarantees the robustness of the LMS type adaptive filter W in the presence of observation noise $s(i) + n(i)$.

We assume that the length of the filter W matches the length of the echo path $h(i)$. Nevertheless, a quickly converging echo canceler W often yields an insufficient estimate of the acoustic echo. The reason is that the impulse response $h(i)$ is continuously changing and the local signal $s(i) + n(i)$ clearly means a disturbance for the identification process of the filter W. The residual echo $b(i) = d(i) - \widehat{d}(i)$ can be statistically suppressed by a postfilter H with input signal $e(i) = s(i) + n(i) + b(i)$.

Wiener Postfilter for Residual Echo Suppression

According to the original paper [2], the postfilter H could be used for combined residual echo and background noise suppression. Here we emphasize a special case of that. Consider Figure 1 in which a linear filter H shall be applied to the echo compensated signal $e(i)$ such that $\widehat{s}(i)$ approximates $s(i)+n(i)$ in the minimum mean–square error (MMSE) sense.

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

$$\widehat{S}(\ell, k) = H(\ell, k)E(\ell, k) \ , \qquad (2)$$

if the postfilter weights are determined according to Wiener as

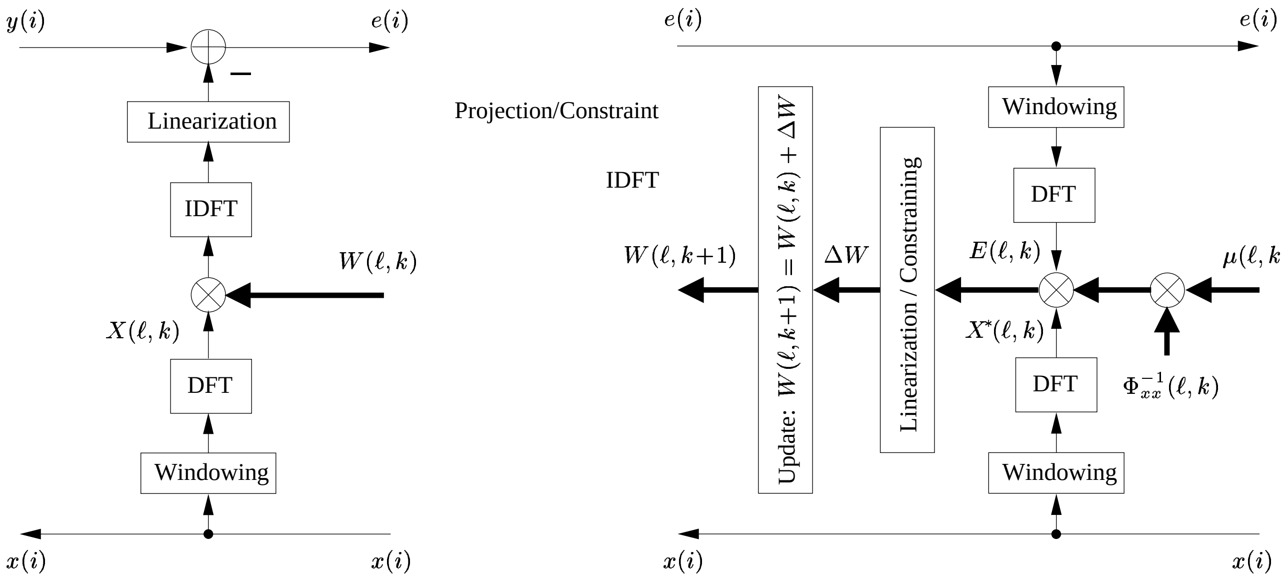

Fig. 2. Schematics of the filtering path (left picture) and the adaptation loop (right picture) of an echo canceler W realized with the FDAF algorithm.

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

For non–stationary signals like in speech communication, the PSDs as required in (3) are varying in time. $\Phi_{xx}(\ell,k)$ and $\Phi_{ee}(\ell,k)$ correspond to the measurable signals $x(i)$ and $e(i)$. $\Phi_{bb}(\ell,k)$ is the residual echo PSD and $|G(\ell,k)|^2$ the residual echo power transfer function. Either one must be estimated from the available signals. In the following section we will show that this problem is closely related to the stepsize control for the echo canceler W.

For completion, the synthesis of the postfilter output $\widehat{s}(i)$ requires an inverse DFT for each frame index k and the overlap/add or overlap/save method to recombine an output signal stream.

Joint Control of Echo Canceler and Postfilter

A relatively unknown but tight relation exists between the echo canceler W and the postfilter H [4, 5]. The FDAF echo canceler especially fits our Wiener postfilter in the DFT domain if we consider the optimum stepsize $\mu(l,k)$ for the FDAF in the MMSE sense [3]:

$$\mu(\ell,k) = \frac{\Phi_{bb}(\ell,k)}{\Phi_{ee}(\ell,k)} = \frac{|G(\ell,k)|^2\Phi_{xx}(\ell,k)}{\Phi_{ee}(\ell,k)} \quad (4)$$

We note that the stepsize $\mu(\ell,k)$ depends on the same parameters than the postfilter weights $H(\ell,k)$ in (3), especially on the (unknown) residual echo power transfer function $|G(\ell,k)|^2$. Interestingly, we even find that $\mu(\ell,k) + H(\ell,k) = 1$ in the optimum case.

This relation can be used to synchronize the echo canceler W and the postfilter H very effectively and therefore yields a simplified structure of AEC units. The synchronization (joint control) further achieves a very consistent interaction between echo canceler and postfilter and therefore results in an excellent output signal quality of the AEC unit.

Audio Demonstration

Scenario: The received signal $x(i)$ was reproduced by a hands–free loudspeaker in the passenger footwell inside of a car. The acoustic echo $d(i)$ mixed with various levels of local speech $s(i)$ and background noise $n(i)$ was recorded by a hands–free microphone located next to the driver's mirror. The acoustic coupling between loudspeaker and microphone was -10 dB. Inside the car, the driver performed typical movements to simulate acoustic echo path changes.

Results: Our system converges very quickly, but the residual echo after the echo canceler W is clearly audible. The residual echo is basically inaudible after the postfilter H. Nevertheless, a distortion of the useful signal is not noticeable. [Samples presented at DAGA.]

Summary and Conclusion

Echo canceler and postfilter are the essential building blocks of an AEC unit. The synchronization of both (joint control) simplifies the structure of AEC units and results in an excellent end user quality.

Acknowledgment

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

References

[1] E. R. Ferrara, "Frequency–domain adaptive filtering," *In: C.F.N. Cowan, P.M. Grant (Eds.), Adaptive Filters, Prentice–Hall, Englewood Cliffs, N.J.*, pp. 145–179, 1985.

[2] S. Gustafsson, R. Martin, and P. Vary, "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, pp. 21–32, 1998.

[3] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Processing*, vol. 80, pp. 1733–1745, 2000.

[4] E. Hänsler, G. Schmidt, "Hands-free telephones - joint control of echo cancellation and postfiltering," *Signal Processing*, vol. 80, pp. 2295–2305, 2000.

[5] G. Enzner, R. Martin, P. Vary, "Partitioned residual echo power estimation for frequency–domain acoustic echo cancellation and postfiltering," *European Transaction on Telecommunications*, vol. 13, March 2002.