

A SOFT-PARTITIONED FREQUENCY-DOMAIN ADAPTIVE FILTER FOR ACOUSTIC ECHO CANCELLATION

Gerald Enzner and Peter Vary

Institute of Communication Systems and Data Processing (IND)

Aachen University (RWTH), D-52056 Aachen, Germany

Phone: +49-241-80-26960, E-mail: {enzner, vary}@ind.rwth-aachen.de

ABSTRACT

We present a *soft-partitioned* frequency-domain adaptive filter (SPFDAF) and compare it in the context of other LMS type acoustic echo cancelers. The SPFDAF uses a non-rectangular window to partition the model taps (impulse response) of an echo canceler. Our *soft window* can be approximated very efficiently in the DFT domain and still attenuates aliasing components of the DFT by 40 dB. This technique results in a very low computational complexity and maintains an excellent signal quality.

1. INTRODUCTION

We consider the family of LMS type algorithms in the application of acoustic echo cancellation. This class of adaptive filters is often represented by the normalized least mean-square algorithm (NLMS) and the frequency-domain adaptive filter (FDAF) [1]. The partitioned block frequency-domain adaptive filter (PBFDAF) [2] can be seen as a solution in-between NLMS and FDAF.

A rectangular sectioning, like in the PBFDAF, is however not the only possible way to partition the model taps of an echo canceler. The question arises which is the optimum window function for a partitioned block frequency-domain adaptive filter. We will demonstrate that an *overlapped soft-partitioning* of the unknown system at least introduces some benefits compared to non-overlapped rectangular partitioning. The approach will be referred to as the *soft-partitioned frequency-domain adaptive filter* (SPFDAF). In contrast to [3], we choose a cosine window function which better eliminates aliasing components produced by the DFT and hence delivers a better signal quality.

In order to compare the performances of NLMS, FDAF, PBFDAF, and SPFDAF to each other we define the following quality attributes for acoustic echo cancelers:

- **Computational Complexity.** This will be compared for a given signal delay. We assume that an algorithmic delay of a few milliseconds is acceptable in most applications.
- **Convergence Speed.** The rate of convergence will be compared for a given accuracy in the steady state.
- **Signal Quality.** This attribute addresses the degree of naturalness of the output signal, e.g. the presence of block artifacts.

Figure 1 introduces basic denotation to be used. The local microphone signal at the sampling time index i

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

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

is additively composed of clean near speech $s(i)$, local background noise $n(i)$, and acoustic echo $d(i)$, respectively. The linear echo canceler W produces an estimate $\hat{d}(i)$ of the acoustic echo, given the received speech signal $x(i)$. The error signal $e(i)$ is transmitted to the far speaker.

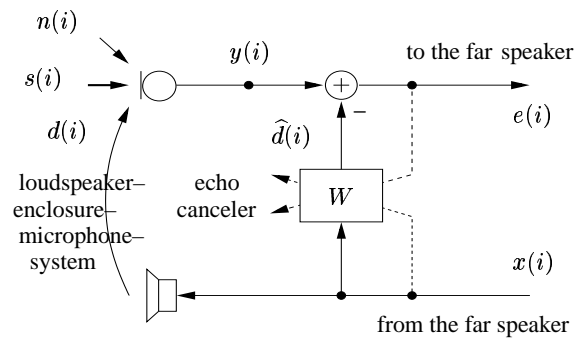


Fig. 1. Acoustic echo cancellation in a hands-free scenario.

Our paper is organized as follows: Section 2 recalls the frequency-domain adaptive filter (FDAF) and the partitioned block concept based on a rectangular window (PBFDAF). In Section 3, the soft-partitioned frequency-domain adaptive filter (SPFDAF) will be developed in two steps from the PBFDAF: modification and approximation. In Sections 4 and 5, the SPFDAF will be compared to other LMS type adaptive filters in terms of computational complexity and convergence properties.

2. FREQUENCY-DOMAIN ADAPTIVE FILTERING

We summarize the FDAF and the PBFDAF for an effective length N of the echo canceler.

2.1. The FDAF Principle

The frequency-domain adaptive filter (FDAF) [1, 4] uses a DFT length of $M = 2N$. The DFT spectrum $E(\Omega_\ell, kR)$ at frame index $k \in \mathbb{Z}$ and frequency index $\Omega_\ell = 2\pi\ell/M$, $\ell = 0, 1, \dots, M-1$, is obtained from the windowed time domain signal $\tilde{e}(i)$

$$\begin{aligned} E(\Omega_\ell, kR) &\triangleq \text{DFT}\{\tilde{e}(i)\} \\ &\triangleq \text{DFT}\{e(kR - M + R + i)w_e(i)\} \\ &\triangleq \sum_{i=0}^{M-1} e(kR - M + R + i)w_e(i)e^{-j\Omega_\ell i} \end{aligned} \quad (2)$$

with frame shift $R \leq M/2 = N$. The rectangular window function applied to the signal $e(i)$ is defined as

$$w_e(i) = \begin{cases} 1 & \text{for } M/2 \leq i \leq M-1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The same notation as in (2) holds for the DFT coefficients $X(\Omega_\ell, kR)$ corresponding to the input signal $x(i)$ if we construct $\tilde{x}(i)$ with an extended window function

$$w_x(i) = \begin{cases} 1 & \text{for } 0 \leq i \leq M-1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Illustrated in Figure 2, the constrained FDAF updates the frequency-domain adaptive weights $W(\Omega_\ell, kR)$ according to

$$W(\Omega_\ell, (k+1)R) = W(\Omega_\ell, kR) \quad (5)$$

$$+ \text{DFT} \left\{ q(i) \cdot \text{IDFT} \left\{ \mu \frac{X^*(\Omega_\ell, kR) E(\Omega_\ell, kR)}{\Phi_{XX}(\Omega_\ell, kR)} \right\} \right\}$$

using the projection (i.e. constraining) window

$$q(i) = \begin{cases} 1 & \text{for } 0 \leq i \leq M/2 - 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

and the DFT spectrum $E(\Omega_\ell, kR)$ corresponding to the error signal ($M/2 \leq i \leq M-1$)

$$e(kR - M + R + i) = y(kR - M + R + i) - \text{IDFT} \{ X(\Omega_\ell, kR) W(\Omega_\ell, kR) \} \quad (7)$$

In (5), the normalization power spectral density (PSD) $\Phi_{XX}(\Omega_\ell, kR)$ is usually approximated by first order recursive smoothing of $|X(\Omega_\ell, kR)|^2$ and μ denotes the non-negative step-size.

The frame shift $R \leq N$ controls the exchange between signal delay and computational complexity. Conflicting demands with respect to delay and complexity are reduced in the PBFDAF.

2.2. The PBFDAF

N model taps of the echo canceler are uniformly divided into $L = 2N/M$ partitions of length $M/2$ and still M is the DFT length. The signal delay of the PBFDAF is equal to the frame shift which is set to $R = M/2$ for any N .

The adaptive weights corresponding to partition λ ,

$$W^{(\lambda)}(\Omega_\ell, kR), \quad 0 \leq \lambda \leq L-1, \quad (8)$$

are individually updated in the style of Equation (5)

$$W^{(\lambda)}(\Omega_\ell, (k+1)R) = W^{(\lambda)}(\Omega_\ell, kR) \quad (9)$$

$$+ \text{DFT} \left\{ q(i) \cdot \text{IDFT} \left\{ \mu^{(\lambda)} \frac{X^{*(\lambda)}(\Omega_\ell, kR) E(\Omega_\ell, kR)}{\Phi_{XX}(\Omega_\ell, kR)} \right\} \right\}$$

given delayed versions of the excitation spectrum

$$X^{(\lambda)}(\Omega_\ell, kR) = X(\Omega_\ell, (k-\lambda)R) \quad (10)$$

and $E(\Omega_\ell, kR)$ corresponding to the compound error signal ($M/2 \leq i \leq M-1$)

$$e(kR - R + i) = y(kR - R + i) - \text{IDFT} \left\{ \sum_{\lambda=0}^{L-1} X^{(\lambda)}(\Omega_\ell, kR) W^{(\lambda)}(\Omega_\ell, kR) \right\} \quad (11)$$

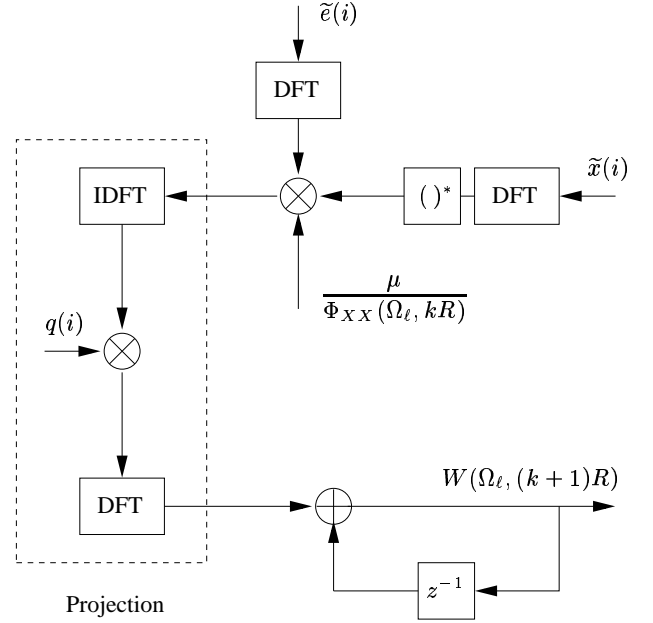


Fig. 2. Adaptation loop of the FDAF according to Equation (5). The signal flow comprises correlation, normalization, constraining (projection), and the actual update.

The computational complexity of the PBFDAF is mainly given by the projection (compare Figure 2) required in each partition λ . That motivates the optimization of this module and results in the SPFDAF.

3. THE SOFT-PARTITIONED FREQUENCY-DOMAIN ADAPTIVE FILTER

In the PBFDAF, each set of coefficients $W^{(\lambda)}(\Omega_\ell, kR)$, $0 \leq \lambda \leq L-1$, corresponds to a unique rectangular section of model taps. The reason is the rectangular window function $q(i)$ as defined in (6). This constraint will be relaxed in the SPFDAF by the modified window function $q_{\text{soft}}(i)$ to be used instead of $q(i)$.

3.1. Cosine Projection Window

We propose a non-rectangular projection window $q_{\text{soft}}(i)$, for example

$$q_{\text{soft}}(i) = \begin{cases} 0.5(1 - \cos(2\pi i/M)) & \text{for } 0 \leq i \leq M/2 - 1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

which completely eliminates time-domain aliasing produced by the cyclic correlation with the DFT. The non-zero extent of the cosine window in (12) is thus limited and aligned to the support region of the previous rectangular window. For a DFT length of $M = 256$ both windows are compared to each other in Figure 3.

In order to provide a uniform support of the echo canceler, the projection window $q_{\text{soft}}(i)$ must be overlapped with 50% as shown in Figure 4. That results in $L_{\text{soft}} = 2L = 4N/M$ partitions to realize an echo canceler with N discrete FIR taps.

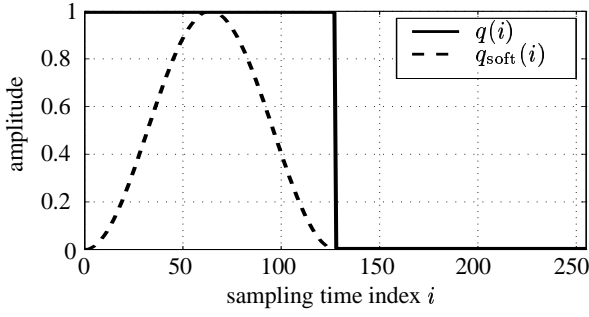


Fig. 3. Our cosine projection window $q_{\text{soft}}(i)$ for $M = 256$.

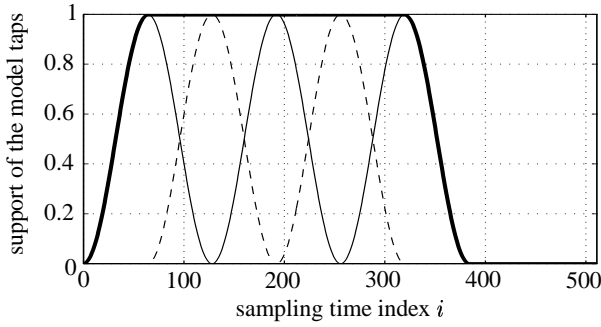


Fig. 4. Overlapping cosine projection windows and total region of support of the echo canceler (bold) with the SPFDAF. $M = 256$ and $L_{\text{soft}} = 5$.

3.2. Cyclic Projection Filter

The replacement of $q(i)$ with $q_{\text{soft}}(i)$ encourages the translation of the projection in Figure 2 into a cyclic convolution in the DFT domain. The changes can be observed from the new system in Figure 5.

The advantage of this procedure becomes obvious if we consider the DFT coefficients of our soft window $Q_{\text{soft}}(\Omega_\ell) = \text{DFT}\{q_{\text{soft}}(i)\}$ illustrated by Figure 6. Apparently, the cyclic convolution of $\Delta W(\Omega_\ell, kR)$ and $Q_{\text{soft}}(\Omega_\ell)$ can be approximated very efficiently with a small number of coefficients.

An example approximation $\hat{Q}_{\text{soft}}(\Omega_\ell) \approx Q_{\text{soft}}(\Omega_\ell)$ which still provides 40 dB attenuation of aliasing components is given by the following coefficients obtained by truncation of the exact coefficients ($M = 256$):

$$\hat{Q}_{\text{soft}}(\Omega_\ell) = \begin{cases} 64 & \text{for } \ell = 0 \\ -j \cdot 54.3249 & \text{for } \ell = 1 \\ -32 & \text{for } \ell = 2 \\ j \cdot 10.8650 & \text{for } \ell = 3 \\ 0 & \text{for } \ell = 4, 5, \dots, \frac{M}{2} + 1 \\ \hat{Q}_{\text{soft}}^*(\Omega_{M-\ell}) & \text{otherwise} \end{cases} \quad (13)$$

The approximation in (13) is the basis for any further evaluation of the SPFDAF in this paper.

3.3. SPFDAF Algorithm

The SPFDAF algorithm is summarized as follows: The adaptive weights corresponding to partition λ

$$W^{(\lambda)}(\Omega_\ell, kR), \quad 0 \leq \lambda \leq L_{\text{soft}} - 1, \quad (14)$$

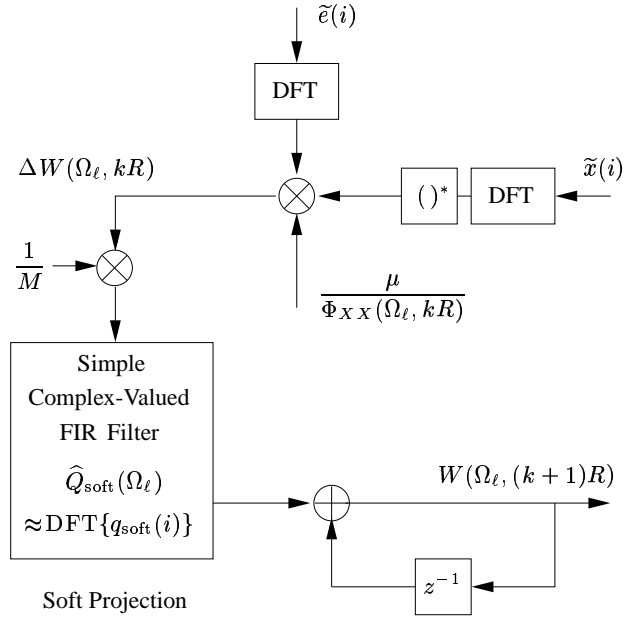


Fig. 5. Adaptation loop of the SPFDAF (partition index λ omitted) with an efficient filter-realization of the soft window $q_{\text{soft}}(i)$.

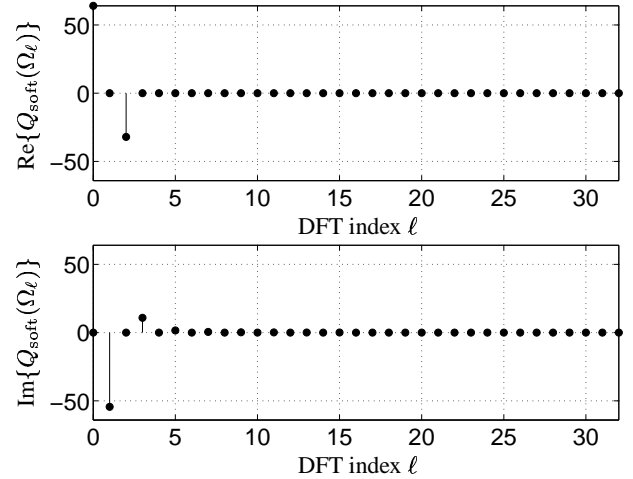


Fig. 6. DFT coefficients of the cosine projection window $q_{\text{soft}}(i)$ for $M = 256$ and $\ell = 0, 1, \dots, M/8$.

are updated in the same way as in Equation (9), given the delayed versions of the excitation spectrum

$$X^{(\lambda)}(\Omega_\ell, kR) = X(\Omega_\ell, (k - \frac{\lambda}{2})R). \quad (15)$$

The projection window $q_{\text{soft}}(i)$ is used instead of $q(i)$ and is approximated as a short filter $\hat{Q}_{\text{soft}}(\Omega_\ell)$ in the DFT domain, compare Figure 5. The DFT coefficients $E(\Omega_\ell, kR)$ correspond to the compound error signal obtained from Equation (11) where the number of partitions L must be replaced with $L_{\text{soft}} = 4N/M$. In this way, the SPFDAF has the same number of model taps N and yields the same algorithmic delay of R samples as the FDAF or PBFDAF.

4. COMPUTATIONAL COMPLEXITY

We compare the computational complexity of the LMS type algorithms NLMS, FDAF, PBFDAF, and SPFDAF, given the number of model taps N of the echo canceler.

The complexity is measured in terms of operations per output sample. A single operation shall be a real-valued multiply/accumulate instruction which is typically performed within one cycle on a digital signal processor. We further assume that a division can also be accomplished within one operation, possibly at the cost of reduced accuracy. According to [5], we fix the number of operations required for an FFT of length M to $3M \log M$, although the FFT complexity is still dependent on the processor.

The results of this study are plotted in Figure 7, where we assumed a DFT length of $M = 2R = 256$ for the partitioned block approaches (PBFDAF, SPFDAF) and a frame shift of $R = 128$ for the FDAF. That yields an equal algorithmic delay (except for the NLMS which has zero delay). The FDAF is considered in an "efficient" unconstrained realization.

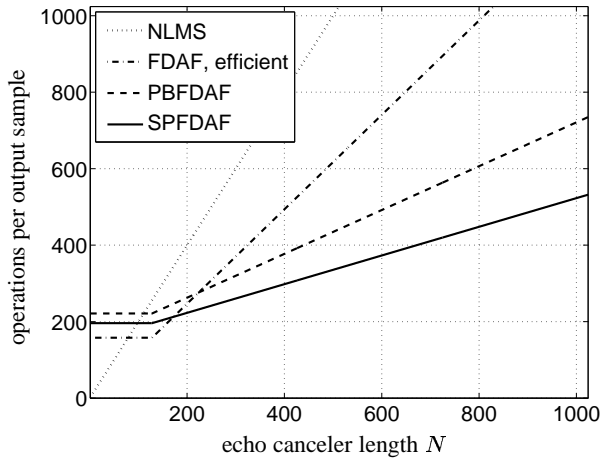


Fig. 7. Multiply/accumulate operations per output sample for different adaptive filters with an equal frame shift (delay) of $R = 128$ samples (except NLMS with zero delay).

5. CONVERGENCE PROPERTIES

By simulation we investigate the rate of convergence of different adaptive filters: NLMS (with short decorrelation filters), FDAF, PBFDAF, and SPFDAF. The rate of convergence is evaluated in terms of Echo Return Loss Enhancement (ERLE) over time. The results are presented in Figure 8.

The acoustic echo was produced artificially by means of a measured (but fixed) car impulse response of 512 taps at 8kHz. For Figure 8 we used a stationary but correlated excitation signal $x(i)$ with "formants" and harmonic fine structure ("pitch"). The PSD of the speech-like test signal is plotted in Figure 9. The acoustic echo at the local microphone was mixed with car background noise such that the global echo-to-noise ratio evaluates to 10 dB. Simulations with real speech have been consistent with Figure 8.

The length of the echo canceler was chosen as $N = 512$ and the algorithmic delay was set to $R = 128$ samples in every algorithm (except the NLMS which has zero delay). For the PBFDAF that means $L = 4$ partitions whereas the SPFDAF uses $L_{\text{soft}} = 2L = 8$ overlapping sections of coefficients.

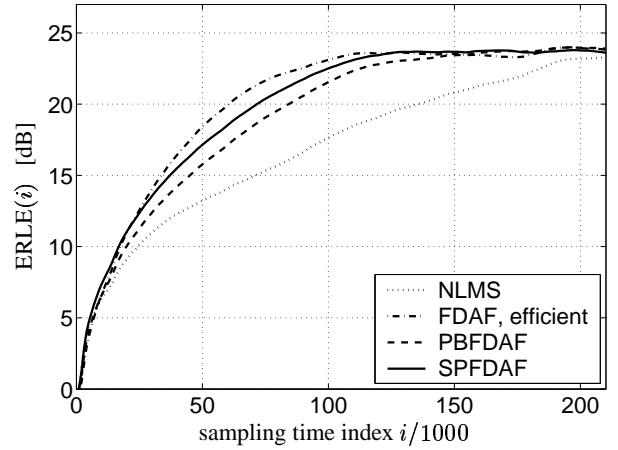


Fig. 8. Rate of convergence for different adaptive filters with $N = 512$ and frame shift $R = 128$ corresponding to Figure 7.

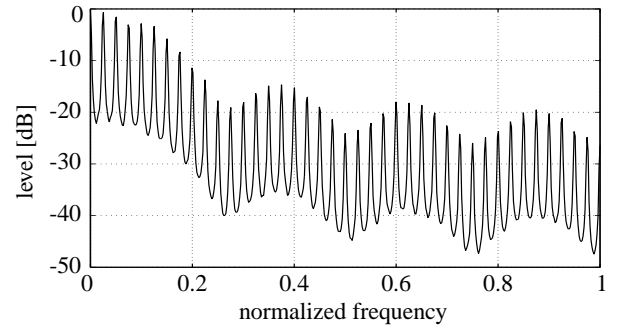


Fig. 9. PSD of the correlated (speech-like) excitation signal $x(i)$.

6. CONCLUSION

We have introduced a *soft-partitioned frequency-domain adaptive filter (SPFDAF)*. This technique is of interest in applications relying on very high order adaptive filters. In a comparison of LMS type approaches, the SPFDAF converges faster than the PBFDAF and at the same time consumes less processing power.

The SPFDAF applies a soft projection window to section the model taps of an echo canceler. In contrast to [3], the limited extent of our window function does not support the aliasing components produced by the DFT. This is essential for a high signal quality.

7. 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] P. C. W. Sommen, *Adaptive Filtering Methods*, Ph.D. thesis, TU Eindhoven, 1992.
- [3] R. Derks, G. Egelmeers, P. C. W. Sommen, "New constraining method for partitioned block frequency-domain adaptive filters," *IEEE Trans. Signal Processing*, vol. 50, pp. 2177–2186, 2002.
- [4] S. Haykin, "Adaptive Filter Theory," Prentice Hall, 1996.
- [5] A. V. Oppenheim, R. W. Schaffer, "Discrete-Time Signal Processing," Prentice Hall, 1989.