Robust Extended Multidelay Filter and Double-Talk Detector for Acoustic Echo Cancellation

Herbert Buchner, *Member, IEEE*, Jacob Benesty, *Senior Member, IEEE*, Tomas Gänsler, *Member, IEEE*, and Walter Kellermann, *Member, IEEE*

***Abstract*—We propose an integrated acoustic echo cancellation solution based on a novel class of efficient and robust adaptive algorithms in the frequency domain, the extended multidelay filter (EMDF). The approach is tailored to very long adaptive filters and highly auto-correlated input signals as they arise in wideband full-duplex audio applications. The EMDF algorithm allows an attractive tradeoff between the well-known multidelay filter and the recursive least-squares algorithm. It exhibits fast convergence, superior tracking capabilities of the signal statistics, and very low delay. The low computational complexity of the conventional frequency-domain adaptive algorithms can be maintained thanks to efficient fast realizations. We also show how this approach can be combined efficiently with a suitable double-talk detector (DTD). We consider a corresponding extension of a recently proposed DTD based on a normalized cross-correlation vector whose performance was shown to be superior compared to other DTDs based on the cross-correlation coefficient. Since the resulting DTD also has an EMDF structure it is easy to implement, and the fast realization also carries over to the DTD scheme. Moreover, as the robustness issue during double talk is particularly crucial for fast-converging algorithms, we apply the concept of robust statistics into our extended frequency-domain approach. Due to the robust generalization of the cost function leading to a so-called M-estimator, the algorithms become inherently less sensitive to outliers, i.e., short bursts that may be caused by inevitable detection failures of a DTD. The proposed structure is also well suited for an efficient generalization to the multichannel case.**

***Index Terms*—Adaptive filtering, doubletalk, echo cancellation, frequency domain, robust statistics.**

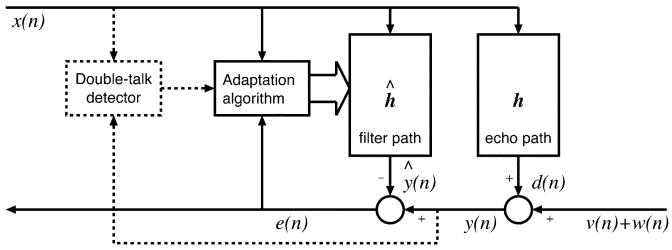


Fig. 1. Adaptive filter and DTD in the AEC application.

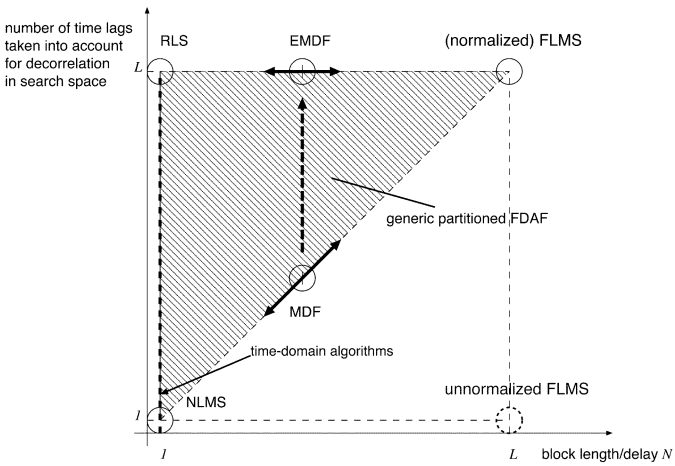


Fig. 2. Relation of EMDF to other algorithms.

I. Introduction

In ACOUSTIC echo cancellation (AEC), as illustrated in Fig. 1, two main problems have to be addressed: 1) system identification of the loudspeaker-room-microphone path in order to cancel the acoustic echo $d(n)$ and 2) reliable and fast control of the adaptation by a double-talk detector (DTD) in order to avoid a divergence of the adaptive system identification algorithm during presence of speech $v(n)$ in the receiving room in addition to the ambient noise $w(n)$.

Many signal processing applications call for adaptive filters with very long impulse responses. In AEC, thousands of FIR filter coefficients may be required to sufficiently model the echo path. Moreover, the input data are often very strongly correlated which causes slow convergence of most adaptation algorithms, such as the well-known normalized least-mean-square (NLMS) algorithm [1]. The requirements are particularly demanding for high-quality and/or multichannel audio reproduction so that more sophisticated algorithms taking into account the input signal correlations have to be used. In this sense, the recursive seast-squares (RLS) algorithm can be regarded as the optimum choice since it takes into account all relevant time lags for decorrelation in the search space of the filter coefficient vector of length L [1]. This is illustrated in Fig. 2, where these time-domain algorithms with sample-by-sample updates are represented by the vertical dashed line on the left. On the other hand, however, using fast-converging algorithms, such as the RLS or fast RLS (FRLS) is often considered to be impractical due to the associated high computational complexity and/or numerical instability problems.

An attractive solution to these problems is to use frequency-domain adaptive filters since, on the one hand, the computational complexity can be greatly reduced by exploiting the fast Fourier transform (FFT). On the other hand, the discrete Fourier transform (DFT) approximately decorrelates the input signals, which leads to very favorable convergence properties

Manuscript received August 17, 2004; revised May 30, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun C. Surendran.

H. Buchner and W. Kellermann are with the University of Erlangen-Nuremberg, Erlangen, Germany (e-mail: buchner@lnt.de; wk@lnt.de).

J. Benesty is with Université du Québec, INRS-EMT, Montréal, QC H5A 1K6 Canada (e-mail: benesty@emt.inrs.ca).

T. Gänsler is with Agere Systems, Allentown, PA 18109 USA (e-mail: gaensler@agere.com).

Digital Object Identifier 10.1109/TSA.2005.858559

of the adaptive algorithms. Frequency-domain methods rely on block-processing. The horizontal axis in Fig. 2 represents the block length N, i.e., the number of new samples used for each update. In early approaches, the block length was set to the number of filter taps (see FLMS in Fig. 2). The associated processing delay, equal to the block length, and the resulting difficulty to track time-varying statistics of nonstationary signals, are often considered to be a major handicap. Therefore, a more flexible structure was introduced, the multidelay filter (MDF) [2], where the filter length L is partitioned into shorter length-N sub-filters. While the processing delay can be significantly reduced with this structure, the major disadvantage of choosing a block length N that is much shorter than the filter length L is that the convergence speed is often severely degraded for strongly correlated signals. We attribute this degradation to the correlations between these shorter length-N blocks, which are not taken into account in the traditional MDF. Another related problem due to this approximation is the need of a relatively "severe" regularization which may additionally slow down the adaptation compared to the nonpartitioned version.

Extending the work in [3], we study in this paper an extended MDF (EMDF) and a fast implementation of it to overcome these problems. After Fig. 2, the objective is to get closer to the desired upper left region representing fast convergence and low delay without significantly increasing the computational load compared to the MDF. The EMDF algorithm in its baseline version [4] follows directly from a generic partitioned frequency-domain adaptive algorithm (shaded area in Fig. 2) which can be rigorously derived from an exponentially weighted least-squares criterion in the frequency-domain [5]. This generic frequency-domain framework has led to novel and efficient implementations of multichannel AEC systems by inherently taking all interchannel correlations into account. In a similar way, the EMDF algorithm also contains all interpartition correlations.

In addition to an improved echo cancellation performance, we also consider in this context the robust operation during double talk, i.e., during speaker activity in the receiving room.

The robustness issue during double talk is particularly crucial for fast-converging algorithms as inevitable failures of double-talk detection (Fig. 1) at the beginning or the end of utterances may then in turn cause fast divergence. Such detection failures causing outliers $\tilde{v}(n)$, i.e., short bursts leaking into the adaptation mechanism from the speech signal $v(n)$ may be seen as alteration of the statistics of the measurement error (in terms of system identification). Therefore, we apply the concept of *robust statistics* [10], [14] to the extended frequency-domain approach. According to [10], robustness signifies insensitivity to a certain amount of deviations from statistical modeling assumptions due to some fraction ϵ of outliers $\tilde{v}(n)$ with some arbitrary probability density function (pdf) $p_{\tilde{v}}(\tilde{v})$. In other words, the shape of the true underlying pdf $p_{w'}(w')$ of the remaining measurement error contributions w' affecting the coefficient updates after double-talk detection deviates from the assumed pdf $p_w(w)$ of the ambient noise (usually Gaussian). This leads to a super-gaussian pdf showing longer tails according to

$$p_{w'}(w') = (1-\epsilon)p_w(w) + \epsilon p_{\tilde{v}}(\tilde{v}). \tag{1}$$

To take this modified pdf into account, the cost function is generalized from a least-squares estimator to a maximum likelihood-type estimator, a so-called M-estimator [10]. Due to this generalization of the cost function, the algorithms may be designed to become inherently less sensitive to such outliers. Robust statistics has already been shown to be a very powerful tool to handle the double-talk situation in the echo cancellation application [11]–[13]. In this contribution, we show how this concept can be included in the generic partitioned frequency-domain algorithm for obtaining a robust EMDF realization.

One of the most widely used DTDs is the Geigel algorithm [6], which works fairly well when the echo return loss is well defined. In practice, however, this is generally not the case. The need for more sophisticated DTDs that do not depend on the path attenuation is obvious. Alternative methods for double-talk detection have been presented in, e.g., [7] and [8]. A DTD based on a normalized cross-correlation vector was proposed in [9], and it was shown that this DTD performs much better than the Geigel algorithm and other DTDs based on the cross-correlation coefficient. In this paper, we show how to extend the ideas of [9] to the EMDF algorithm. The resulting DTD has an EMDF structure which leads to an efficient implementation and is very well matched to the EMDF-based echo cancellation, as we shall confirm by experimental results.

To keep the formal presentation short and accessible, we concentrate on the single-channel EMDF algorithm in this paper; the generalization to the multichannel version is obtained analogously as in [5]. In contrast to interchannel correlations, the interpartition correlations in the EMDF result from a shift-structure of the data. This structure can be exploited to derive fast implementations. Using a fast implementation of the EMDF algorithm (FEMDF) the computational complexity can be kept on the same order as that of the classical MDF.

II. Robust Generic Partitioned Frequency-Domain Adaptive Filtering

Here, we present a generic frequency-domain algorithm in its robust partitioned and constrained single-channel version providing the basis for the robust EMDF algorithm introduced in Section III.

A. Definitions and Notation

In this paper, we follow the same notation as in [5], where a detailed derivation and an analysis of the nonrobust generic algorithm can be found.

From Fig. 1, it can be seen that the error signal at time n between the output of the adaptive filter $\hat{y}(n)$ and the desired output signal $y(n)$ is given by

$$e(n) = y(n) - \sum_{\ell=0}^{L-1} x(n-\ell)\hat{h}_\ell \tag{2}$$

where $\hat{h}_\ell$ are the coefficients of the filter impulse response $\hat{h}$. By partitioning the impulse response $\hat{h}$ of length L into K segments of integer length $N = L/K$ as in [2], (2) can be written as

$$\begin{aligned} e(n) &= y(n) - \sum_{k=0}^{K-1}\sum_{\ell=0}^{N-1} x(n-Nk-\ell)\hat{h}_{Nk+\ell} \\ &= y(n) - \sum_{k=0}^{K-1} \mathbf{x}_k^T(n)\hat{\mathbf{h}}_k = y(n) - \mathbf{x}^T(n)\hat{\mathbf{h}} \end{aligned} \tag{3}$$

where

$$\mathbf{x}_k(n) = [x(n-Nk), x(n-Nk-1), \ldots, x(n-Nk-N+1)]^T \quad (4)$$

$$\hat{\mathbf{h}}_k = [\hat{h}_{Nk}, \hat{h}_{Nk+1}, \ldots, \hat{h}_{Nk+N-1}]^T \quad (5)$$

$$\mathbf{x}(n) = [\mathbf{x}_0^T(n), \mathbf{x}_1^T(n), \ldots, \mathbf{x}_{K-1}^T(n)]^T. \quad (6)$$

Superscript ^{T} denotes transposition of a vector or a matrix. The length-N vectors $\hat{\mathbf{h}}_k, k = 0, \ldots, K-1$ represent *sub-filters* of the partitioned tap-weight vector

$$\hat{\mathbf{h}} = [\hat{\mathbf{h}}_0^T, \ldots, \hat{\mathbf{h}}_{K-1}^T]^T. \quad (7)$$

We now define the block error signal of length N. Based on (3), we write

$$\mathbf{e}(m) = \mathbf{y}(m) - \sum_{k=0}^{K-1} \mathbf{U}_k^T(m)\hat{\mathbf{h}}_k \quad (8)$$

where m is the block time index, and

$$\mathbf{e}(m) = [e(mN), \ldots, e(mN+N-1)]^T \quad (9)$$

$$\mathbf{y}(m) = [y(mN), \ldots, y(mN+N-1)]^T \quad (10)$$

$$\mathbf{U}_k(m) = [\mathbf{x}_k(mN), \ldots, \mathbf{x}_k(mN+N-1)]. \quad (11)$$

To derive the frequency-domain algorithm, the block error signal, (8) is transformed by a DFT matrix to its frequency-domain counterpart. The matrices $\mathbf{U}_k(m), k = 0, \ldots, K-1$ are Toeplitz matrices of size $(N \times N)$. Since a Toeplitz matrix $\mathbf{U}_k(m)$ can be transformed, by doubling its size, to a circulant matrix of size $(2N \times 2N)$, and a circulant matrix can be diagonalized using the $(2N \times 2N)$-DFT matrix $\mathbf{F}_{2N}$ with elements $e^{-j2\pi\nu n/(2N)}(\nu, n = 0, \ldots, 2N-1)$, we have

$$\mathbf{U}_k^T(m) = \underbrace{[\mathbf{0}_{N\times N}, \mathbf{I}_{N\times N}]}_{=:\mathbf{W}_{N\times 2N}^{01}} \mathbf{F}_{2N}^{-1}\mathbf{X}_k(m)\mathbf{F}_{2N} \underbrace{[\mathbf{I}_{N\times N}, \mathbf{0}_{N\times N}]^T}_{=:\mathbf{W}_{2N\times N}^{10}} \quad (12)$$

with the diagonal matrices

$$\mathbf{X}_k(m) = \mathrm{diag}\{\mathbf{F}_{2N}[x(mN-Nk-N), \ldots, x(mN-Nk+N-1)]^T\}. \quad (13)$$

The superscript indices "01" and "10" of the window matrices $\mathbf{W}_{N\times 2N}^{01}$ and $\mathbf{W}_{2N\times N}^{10}$ describe the relative positions of the $N \times N$ identity matrix and the $N \times N$ zero matrix within the windows. This finally leads to the following block error signal:

$$\mathbf{e}(m) = \mathbf{y}(m) - \mathbf{W}_{N\times 2N}^{01}\mathbf{F}_{2N}^{-1}\mathbf{X}(m)\mathbf{G}_{2L\times L}^{10}\underline{\hat{\mathbf{h}}} \quad (14)$$

where

$$\mathbf{X}(m) = [\mathbf{X}_0(m), \mathbf{X}_1(m), \ldots, \mathbf{X}_{K-1}(m)] \quad (15)$$

$$\mathbf{G}_{2L\times L}^{10} = \mathrm{diag}\left\{\mathbf{G}_{2N\times N}^{10}, \ldots, \mathbf{G}_{2N\times N}^{10}\right\} \quad (16)$$

$$\mathbf{G}_{2N\times N}^{10} = \mathbf{F}_{2N}\mathbf{W}_{2N\times N}^{10}\mathbf{F}_N^{-1} \quad (17)$$

$$\underline{\hat{\mathbf{h}}} = \left[\underline{\hat{\mathbf{h}}}_0^T, \ldots, \underline{\hat{\mathbf{h}}}_{K-1}^T\right]^T \quad (18)$$

$$\underline{\hat{\mathbf{h}}}_k = \mathbf{F}_N\hat{\mathbf{h}}_k. \quad (19)$$

B. Optimization Criterion

We now formulate a criterion that is minimized with respect to the filter coefficients in the DFT domain using (14). Modeling the noise with a super-gaussian pdf to obtain an outlier-robust algorithm corresponds to a nonquadratic optimization criterion [10]–[13]. Following [10], we generalize the block-based weighted least-squares criterion (e.g., [5]) to a corresponding M-estimator

$$J(m, \underline{\hat{\mathbf{h}}}) = \sum_{i=0}^{m}\beta(i,m)\sum_{n=iN}^{iN+N-1}\rho\left[\frac{|e(n)|}{s}\right] = \sum_{i=0}^{m}\beta(i,m)\tilde{J}(i, \underline{\hat{\mathbf{h}}}) \quad (20)$$

where $\beta(i,m)$ is a weighting function defining different classes of algorithms [1], e.g., $\beta(i,m) = (1-\lambda)\lambda^{m-i}$ with the forgetting factor $0 < \lambda < 1$ to obtain an RLS-like algorithm. Note that $\rho[|e(n)|/s] = |e(n)|^2$ gives us the corresponding nonrobust algorithm. Moreover, the choice $\rho[(|e(n)|/s)] = -\log p_{w'}(e(n))$ immediately shows the close relation to the ordinary maximum likelihood estimator. In general, $\rho[\,\cdot\,]$ is a convex function and s is a real-valued positive scale factor for the ith block as discussed in [10], [12] (see also Section II-D). One of the main statements of the theory on robust statistics is that the resulting algorithm inherits robust properties as long as the nonlinear function $\rho[\,\cdot\,]$ has a bounded derivative [10]. We can easily verify that the condition of a bounded derivative is not fulfilled for the classical case $\rho[\,\cdot\,] = |\cdot|^2$.

A particularly simple, but efficient, choice of $\rho[\,\cdot\,]$ for robustness is given by the so-called Huber estimator [10]

$$\rho(|z|) = \begin{cases} \frac{|z|^2}{2} & \text{for } |z| \le k_0 \\ k_0|z| - \frac{k_0^2}{2}, & \text{for } |z| \ge k_0 \end{cases} \quad (21)$$

where k_0 is a constant controlling the robustness of the algorithm. Note that, indeed, according to Hampel's proposition in [14, pp. 117–119], the choice (21) gives the optimum equivariant robust estimator under the assumption of Gaussian background noise $p_w(w)$ in the error model (1).

C. Adaptation Algorithm

To obtain an adaptation algorithm from the optimization criterion (20), we need to recursively solve the corresponding normal equation (see [1] for the complex gradient)

$$\nabla J(m, \underline{\hat{\mathbf{h}}}) = 2\frac{\partial}{\partial \underline{\hat{\mathbf{h}}}^*}J(m, \underline{\hat{\mathbf{h}}}) = \mathbf{0}. \quad (22)$$

In contrast to the nonrobust case leading to an RLS solution, (22) is a nonlinear system of equations which we nevertheless would like to tackle in a similar way.

As shown in Appendix I, the complex adaptive Newton algorithm minimizes for $\beta(i,m) = (1-\lambda)\lambda^{m-i}$ the nonquadratic criterion (20) by using a recursion of the form

$$\underline{\hat{\mathbf{h}}}(m) = \underline{\hat{\mathbf{h}}}(m-1) - 2\mu(1-\lambda)\mathbf{S}_{\psi'}^{-1}(m)\nabla\tilde{J}[m, \underline{\hat{\mathbf{h}}}(m-1)] \quad (23)$$

where a relaxation parameter $0 \le \mu \le 1$ is introduced; $\nabla\tilde{J} = 2\partial/\partial\underline{\hat{\mathbf{h}}}^*\tilde{J}$ is the gradient of $\tilde{J}$ w.r.t. $\underline{\hat{\mathbf{h}}}$, and $\mathbf{S}_{\psi'}$

is an approximation of the expected value of the Hessian $\nabla\nabla^H\tilde{J} = 2\partial/\partial\hat{\underline{\mathbf{h}}}^*(\nabla\tilde{J})^H$ according to

$$\mathbf{S}_{\psi'}(m) = \lambda\mathbf{S}_{\psi'}(m-1)+(1-\lambda)\nabla\nabla^H\tilde{J}[m,\hat{\underline{\mathbf{h}}}(m-1)]. \quad (24)$$

Hence, to proceed with the generic frequency-domain derivation in generalization of [13], we need to calculate the gradient and Hessian of $\tilde{J}(m,\hat{\underline{\mathbf{h}}})$. To link the block formulation (14) with (20), we write

$$\begin{aligned} e^*(n) &= \mathbf{e}^H(m)\mathbf{1}_{n-mN} \\ &= \left[\mathbf{y}^H(m) - \hat{\underline{\mathbf{h}}}^H\left(\mathbf{G}^{10}_{2L\times L}\right)^H\mathbf{X}^H(m)\mathbf{F}^{-H}_{2N}\mathbf{W}^{01}_{2N\times N}\right] \\ &\quad\cdot\mathbf{1}_{n-mN}, \end{aligned} \quad (25)$$

where $n = mN,\ldots,mN+N-1$, $\cdot^{-H} = (\cdot^H)^{-1}$, and $\mathbf{1}_i$ is a length-N vector containing a 1 in position i and zeros in all other positions. Then, the gradient is found using the chain rule

$$\begin{aligned} \nabla\tilde{J} &= 2\frac{\partial}{\partial\hat{\underline{\mathbf{h}}}^*}\tilde{J} = 2\sum_{n=mN}^{mN+N-1}\frac{\partial}{\partial\hat{\underline{\mathbf{h}}}^*}\rho\left[\frac{|e(n)|}{s}\right] \\ &= 2\sum_{n=mN}^{mN+N-1}\frac{\partial}{\partial\hat{\underline{\mathbf{h}}}^*}[e^*(n)]\rho'\left[\frac{|e(n)|}{s}\right]\frac{\operatorname{sign}[e(n)]}{s}. \end{aligned}$$

Using (25), it follows:

$$\begin{aligned} \nabla\tilde{J} &= -\frac{1}{Ns}\left(\mathbf{G}^{10}_{2L\times L}\right)^H\mathbf{X}^H(m)\mathbf{F}_{2N}\mathbf{W}^{01}_{2N\times N}\boldsymbol{\psi}[\mathbf{e}(m)] \\ &= -\frac{1}{Ns}\left(\mathbf{G}^{10}_{2L\times L}\right)^H\mathbf{X}^H(m)\underline{\boldsymbol{\psi}}[\mathbf{e}(m)] \end{aligned} \quad (26)$$

where

$$\boldsymbol{\psi}[\mathbf{e}(m)] = \begin{bmatrix} \psi\left[\frac{|e(mN)|}{s}\right]\operatorname{sign}[e(mN)] \\ \vdots \\ \psi\left[\frac{|e(mN+N-1)|}{s}\right]\operatorname{sign}[e(mN+N-1)] \end{bmatrix}$$

$$\psi(|z|) = \rho'(|z|) \quad (27)$$

$$\underline{\boldsymbol{\psi}}[\mathbf{e}(m)] = \mathbf{F}_{2N}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \boldsymbol{\psi}[\mathbf{e}(m)]\end{bmatrix}. \quad (28)$$

Note that for the Huber estimator with (21), the derivative (27) results in a simple limiter

$$\psi(|z|) = \min\{|z|, k_0\}. \quad (29)$$

Combining (23) and (26), the robust frequency-domain update reads

$$\begin{aligned} \hat{\underline{\mathbf{h}}}(m) = \hat{\underline{\mathbf{h}}}(m-1) &+ \frac{2\mu(1-\lambda)}{Ns} \\ &\cdot\mathbf{S}^{-1}_{\psi'}(m)\left(\mathbf{G}^{10}_{2L\times L}\right)^H\mathbf{X}^H(m)\underline{\boldsymbol{\psi}}[\mathbf{e}(m)]. \end{aligned} \quad (30)$$

We now derive the Hessian of $\tilde{J}$ and the estimate $\mathbf{S}_{\psi'}$ from (24). The Hessian is expressed as

$$\begin{aligned} \nabla\nabla^H\tilde{J} &= 2\frac{\partial}{\partial\hat{\underline{\mathbf{h}}}^*}(\nabla\tilde{J})^H \\ &= -\frac{4}{s}\frac{\partial}{\partial\hat{\underline{\mathbf{h}}}^*}\boldsymbol{\psi}^H[\mathbf{e}(m)]\mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N}\mathbf{X}(m)\mathbf{G}^{10}_{2L\times L}. \end{aligned}$$

By applying again the chain rule, followed by recursive averaging using the forgetting factor λ according to (24), we finally obtain the estimate

$$\begin{aligned} \mathbf{S}_{\psi'}(m) = \frac{4(1-\lambda)}{s^2}\sum_{i=0}^{m}\lambda^{m-i}(\mathbf{G}^{10}_{2L\times L})^H\mathbf{X}^H(i)\tilde{\mathbf{G}}_{2N\times 2N}(i) \\ \cdot\mathbf{X}(i)\mathbf{G}^{10}_{2L\times L} \end{aligned} \quad (31)$$

where

$$\begin{aligned} &\tilde{\mathbf{G}}_{2N\times 2N}(m) \\ &\quad= \mathbf{F}^{-H}_{2N}\mathbf{W}^{01}_{2N\times N}\boldsymbol{\Psi}'[\mathbf{e}(m)]\mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N} \end{aligned} \quad (32)$$

$$\begin{aligned} &\boldsymbol{\Psi}'[\mathbf{e}(m)] \\ &\quad= \operatorname{diag}\left\{\psi'\left[\frac{|e(mN)|}{s}\right]\ldots\psi'\left[\frac{|e(mN+N-1)|}{s}\right]\right\} \end{aligned} \quad (33)$$

and ψ' is the derivative of ψ. For the special case of the Huber estimator, $\psi'(|z|)$ is 1 for $|z|\le k_0$ and 0 else.

Equations (14), (30), and (31) form the main equations of the generic adaptive algorithm. In the same way as shown in [5] and summarized in Appendix II, these equations can be reformulated in a practically more useful form involving only DFTs of length $2N$

$$\begin{aligned} \mathbf{S}_{\mathrm{d},\psi'}(m) &= \lambda\mathbf{S}_{\mathrm{d},\psi'}(m-1) \\ &\quad+\frac{4(1-\lambda)}{s^2}\mathbf{X}^H(m)\tilde{\mathbf{G}}_{2N\times 2N}(m)\mathbf{X}(m) \end{aligned} \quad (34)$$

$$\mathbf{K}_{\psi'}(m) = \mathbf{S}^{-1}_{\mathrm{d},\psi'}(m)\mathbf{X}^H(m) \quad (35)$$

$$\mathbf{e}(m) = \mathbf{y}(m) - \mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N}\mathbf{X}(m)\hat{\underline{\mathbf{h}}}_{2L}(m-1) \quad (36)$$

$$\begin{aligned} \hat{\underline{\mathbf{h}}}_{2L}(m) &= \hat{\underline{\mathbf{h}}}_{2L}(m-1) \\ &\quad+\frac{2\mu(1-\lambda)}{Ns}\mathbf{G}^{10}_{2L\times 2L}\mathbf{K}_{\psi'}(m)\underline{\boldsymbol{\psi}}[\mathbf{e}(m)] \end{aligned} \quad (37)$$

with the zero-padded coefficient vector

$$\hat{\underline{\mathbf{h}}}_{2L}(m) = \left[\hat{\underline{\mathbf{h}}}^T_{2N,0}(m),\ldots,\hat{\underline{\mathbf{h}}}^T_{2N,K-1}(m)\right]^T \quad (38)$$

$$\hat{\underline{\mathbf{h}}}_{2N,k}(m) = \mathbf{F}_{2N}\begin{bmatrix}\hat{\mathbf{h}}_k(m)\\ \mathbf{0}_{N\times 1}\end{bmatrix}. \quad (39)$$

Due to the formal similarity of (34)–(37) to the RLS algorithm [1] in the time domain, we call the matrix $\mathbf{K}_{\psi'}(m)$ the frequency-domain Kalman gain. The Kalman gain plays a key role in the following sections.

D. Scale Factor

The scaling factor s is a suitable estimate of the spread of the random errors. In practice, s needs to be obtained from the residual error, which in turn depends on $\hat{\underline{\mathbf{h}}}$. In the application to AEC, the scale factor should reflect the background noise level at the near end, be robust to short error bursts during double talk, and track long-term changes of the residual error due to echo path changes. An efficient solution to fulfill these three requirements along with a detailed derivation was first given in [12]. In block formulation for a block length N, this estimate reads

$$s(m+1) = \lambda_s s(m) + (1-\lambda_s)\frac{s(m)}{N\beta}\sum_{n=mN}^{mN+N-1}\psi\left[\frac{|e(n)|}{s(m)}\right] \quad (40)$$

where $s(0) = \sigma_x$ and β is a normalization constant, depending on k_0.

III. Robust Extended Multidelay Filter (EDMF)

The algorithm (34)–(37) is strictly equivalent to the (robust) RLS algorithm in the time domain for a block length $N = 1$. Unfortunately, the matrix $\mathbf{S}_{\mathrm{d},\psi'}(m)$ in (34) is not sparse, so the

above generic algorithm still has a high computational complexity due to the matrix inversion in (35).

In the case of the nonrobust frequency-domain method and the extended multidelay filter (EMDF) [4], [5], it has been shown that (34) and the associated Kalman gain (35) can very well be approximated as a matrix of diagonal submatrices which significantly reduces the complexity of the matrix inversion. To obtain a similarly efficient robust EMDF algorithm, we approximate the algorithm (34)–(37) in three steps.

A. Approximation 1

To begin with, we approximate $\mathbf{\Psi}'[\mathbf{e}(m)]$ in (32) and (33) by the conservative choice (in terms of robustness)

$$\mathbf{\Psi}'[\mathbf{e}(m)] = \psi'_{\min}(m)\mathbf{I}_{N\times N} \tag{41}$$

where $\psi'_{\min}(m)$ is the minimum of ψ' within the current length-N block ($\psi' \in \{0,1\}$ for the Huber estimator). Moreover, as further explained below, we bound $\psi'_{\min}(m)$ by μ to prevent instability of the update (37) so that

$$\psi'_{\min}(m) = \max\left[\mu, \min_{0\le n\le N-1}\left\{\psi'\left[\frac{|e(mN+n)|}{s}\right]\right\}\right]. \tag{42}$$

Since $\mathbf{F}_{2N}^{-H}\mathbf{W}_{2N\times 2N}^{01}\mathbf{F}_{2N}^{-1} = \mathbf{G}_{2N\times 2N}^{01}/(2N)$ and $\mathbf{W}_{2N\times 2N}^{01} = \mathbf{W}_{2N\times N}^{01}\mathbf{W}_{N\times 2N}^{01}$, (32) is now approximated as

$$\tilde{\mathbf{G}}_{2N\times 2N}(m) = \frac{\psi'_{\min}(m)}{2N}\mathbf{G}_{2N\times 2N}^{01}. \tag{43}$$

B. Approximation 2

Now, as in the nonrobust EMDF [4], [5], matrix $\mathbf{G}_{2N\times 2N}^{01}$ may well be approximated by $\mathbf{G}_{2N\times 2N}^{01} = \mathbf{I}_{2N\times 2N}/2$ in (34) for sufficiently large N, which yields

$$\tilde{\mathbf{G}}_{2N\times 2N}(m) = \frac{\psi'_{\min}(m)}{4N}\mathbf{I}_{2N\times 2N}. \tag{44}$$

This approximation leads to a blockwise diagonal structure of matrix $\mathbf{S}_{\mathrm{d},\psi'}(m)$ in (34) and (35) with diagonal sub-matrices as illustrated in Fig. 3 for the example of five partitions. The classical MDF in its robust version is obtained by further approximating $\mathbf{S}_{\mathrm{d},\psi'}(m)$ by dropping the off-diagonal components, i.e., the interpartition correlations (grey diagonals in Fig. 3). This leads to the low computational complexity per output sample, which is linear in K. However, this additional MDF approximation often significantly degrades the convergence speed for highly correlated input data. Another related problem of the MDF due to this approximation is the need of a relatively high regularization compared to the nonpartitioned version which may additionally slow down the adaptation. The EMDF takes the interpartition correlations into account and, thus, provides a better approximation to the exact solution of the normal equation. However, a straightforward implementation leads to a computational complexity, which increases quadratically with the number K of partitions. Fast schemes, as discussed in Section V, provide a solution with a complexity that is comparable to that of the classical MDF.

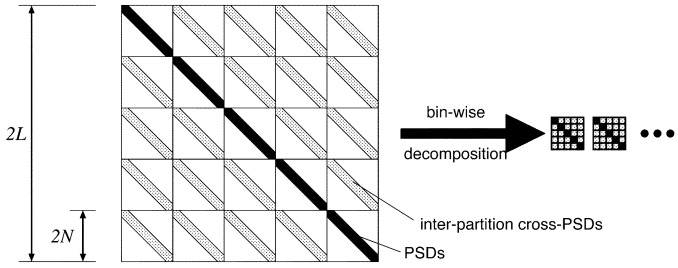


Fig. 3. Structure of matrix $\mathbf{S}'(m)$.

C. Approximation 3

In this paper, we further assume that $s(m)$ and $\psi'_{\min}(m)$ vary slowly over time so that (34) together with (44) can be expressed as

$$\mathbf{S}_{\mathrm{d},\psi'}(m) = \frac{\psi'_{\min}(m)}{Ns^2(m)}\mathbf{S}'(m) \tag{45}$$

where

$$\mathbf{S}'(m) = \lambda\mathbf{S}'(m-1) + (1-\lambda)\mathbf{X}^H(m)\mathbf{X}(m). \tag{46}$$

By introducing the ψ'-independent frequency-domain Kalman gain, as in the nonrobust case [4]

$$\mathbf{K}(m) = \mathbf{S}'^{-1}(m)\mathbf{X}^H(m) \tag{47}$$

so that

$$\mathbf{K}_{\psi}(m) = \frac{Ns^2(m)}{\psi'_{\min}(m)}\mathbf{K}(m) \tag{48}$$

we obtain a very efficient implementation of an integrated EMDF-based AEC system, as shown in the following sections. In particular, synergies in the corresponding DTD introduced in Section IV become possible. Moreover, as discussed in Section V, fast calculation schemes for both, the adaptive filtering and the double-talk detection may be devised.

Note also that the necessity of the bound for $\psi'_{\min}(m)$ in (42) is explained by its use in the denominator of (48).

IV. Extended Multidelay Double-Talk Detector

Double-talk detection (DTD) may be seen as a classification problem which is solved, in general, by forming a suitable (normalized) decision variable ξ, and comparing it to a certain threshold T [9]. The decision variable is generally related to a correlation or a coherence, and the threshold must be a known constant for best performance. Obviously, a fast adaptation of the filter coefficients requires particularly fast and reliable decisions of the DTD. Following the approach in Sections II and III, we propose here a corresponding extended multidelay double-talk detector (EMD DTD) that is well matched to the EMDF.

To simplify the derivation of the DTD, we neglect the effect of the background noise (i.e., $w = 0$) for the present. The noise will be included later. Thus, the microphone signal reads in the frequency domain

$$\underline{\mathbf{y}}(m) = \mathbf{G}_{2N\times 2N}^{01}\mathbf{X}(m)\underline{\mathbf{h}}_{2L} + \underline{\mathbf{v}}(m) \tag{49}$$

where

$$\underline{\mathbf{y}}(m) = \mathbf{F}_{2N} \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{y}(m) \end{bmatrix} \tag{50}$$

and $\underline{\mathbf{v}}(m)$ is defined in the same way in terms of the near-end speech signal $v(n)$.

To begin with, we consider the single-talk case $v = 0$. We have

$$\begin{aligned} \sigma_y^2 &= E\{\underline{\mathbf{y}}^H(m)\underline{\mathbf{y}}(m)\} \\ &= \underline{\mathbf{h}}_{2L}^H \mathbf{S} \underline{\mathbf{h}}_{2L}, \end{aligned} \tag{51}$$

where $E\{\cdot\}$ denotes the mathematical expectation, and

$$\mathbf{S} = E\{\mathbf{X}^H(m)\mathbf{G}_{2N\times 2N}^{01}\mathbf{X}(m)\}. \tag{52}$$

Thanks to (49) and (52), we have

$$\begin{aligned} E\{\mathbf{X}^H(m)\underline{\mathbf{y}}(m)\} &= \mathbf{S}\underline{\mathbf{h}}_{2L} \\ &=: \mathbf{s} \end{aligned} \tag{53}$$

and (51) can be re-written as

$$\begin{aligned} \sigma_y^2 &= \underline{\mathbf{h}}_{2L}^H \mathbf{s} \\ &= \sum_{k=0}^{K-1} \underline{\mathbf{h}}_k^H E\{\mathbf{X}_k^*(m)\underline{\mathbf{y}}(m)\} \\ &= \sum_{k=0}^{K-1} \underline{\mathbf{h}}_k^H \mathbf{s}_k \end{aligned} \tag{54}$$

with

$$\mathbf{s}_k = E\{\mathbf{X}_k^*(m)\underline{\mathbf{y}}(m)\}. \tag{55}$$

Now, in general, for $v \neq 0$

$$\sigma_y^2 = \underline{\mathbf{h}}_{2L}^H \mathbf{s} + \sigma_v^2 \tag{56}$$

where

$$\sigma_v^2 = E\{\underline{\mathbf{v}}^H(m)\underline{\mathbf{v}}(m)\}. \tag{57}$$

If we divide (54) by (56), we obtain the following decision variable to detect the presence of a near-end signal $v(n)$:

$$\xi^2 = \frac{\underline{\mathbf{h}}_{2L}^H \mathbf{s}}{\underline{\mathbf{h}}_{2L}^H \mathbf{s} + \sigma_v^2} = \frac{\underline{\mathbf{h}}_{2L}^H \mathbf{s}}{\sigma_y^2}. \tag{58}$$

We easily deduce from (58) that for $v = 0, \xi = 1$ and for $v \neq 0, \xi < 1$. Note also that ξ is not, in principle, sensitive to changes of the echo path when $v = 0$.

In practice, ξ is *estimated* as follows:

$$\xi^2(m) = \frac{\sum_{k=0}^{K-1} \hat{\underline{\mathbf{h}}}_{\mathrm{b},k}^H(m)\mathbf{s}_k(m)}{\sigma_y^2(m)}. \tag{59}$$

Here, the estimation of the echo path for the decision variable is performed by a separate background EMDF $\hat{\underline{\mathbf{h}}}_{\mathrm{b},k}, k = 0, 1, \ldots, K-1$. However, it is important in practice that the statistics of the signal $y(n)$ (containing both the echo and the near-end signal during double-talk) is tracked fast enough by the background filter, i.e., faster than the statistics of the update of the foreground filter. Hence, the forgetting factor λ_{b} ($0 \ll \lambda_{\mathrm{b}} < 1$) of the background filter is chosen smaller

TABLE I
ROBUST EMDF AND EMD DTD

Definitions:

$$\begin{aligned} \mathbf{W}_1 &= \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix} \\ \mathbf{G}_1 &= \mathbf{F}_{2N}\mathbf{W}_1\mathbf{F}_{2N}^{-1} \\ \mathbf{W}_2 &= \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \end{bmatrix} \\ \tilde{\mathbf{G}}_2 &= \mathbf{F}_{2N}\mathbf{W}_2\mathbf{F}_{2N}^{-1} \\ \mathbf{G}_2 &= \mathrm{diag}\{\tilde{\mathbf{G}}_2, \ldots, \tilde{\mathbf{G}}_2\} \end{aligned}$$

Input:

$$\begin{aligned} \mathbf{X}_k(m) &= \mathrm{diag}\{\mathbf{F}_{2N}[x(mN - Ni - N), \ldots \\ &\quad \ldots, x(mN - Ni + N - 1)]^T\}, \\ &\quad k = 0, \ldots, K-1 \\ \mathbf{X}(m) &= [\mathbf{X}_0(m), \mathbf{X}_1(m), \ldots, \mathbf{X}_{K-1}(m)] \end{aligned}$$

Kalman gain *(see Table II for a fast version)*:

$$\begin{aligned} \mathbf{S}'(m) &= \lambda\mathbf{S}'(m-1) + (1-\lambda)\mathbf{X}^H(m)\mathbf{X}(m) \\ \mathbf{K}(m) &= \mathbf{S}'^{-1}(m)\mathbf{X}^H(m) \end{aligned}$$

Double-talk detector (background filter):

$$\begin{aligned} \underline{\mathbf{e}}_{\mathrm{b}}(m) &= \underline{\mathbf{y}}(m) - \mathbf{G}_1\mathbf{X}(m)\hat{\underline{\mathbf{h}}}_{\mathrm{b}}(m-1) \\ \hat{\underline{\mathbf{h}}}_{\mathrm{b}}(m) &= \hat{\underline{\mathbf{h}}}_{\mathrm{b}}(m-1) + 2(1-\lambda_{\mathrm{b}})\mathbf{G}_2\mathbf{K}(m)\underline{\mathbf{e}}_{\mathrm{b}}(m) \\ \mathbf{s}_k(m) &= \lambda_{\mathrm{b}}\mathbf{s}_k(m-1) + (1-\lambda_{\mathrm{b}})\mathbf{X}_k^*(m)\underline{\mathbf{y}}(m), \\ &\quad k = 0, \ldots, K-1 \\ \sigma_y^2(m) &= \lambda_{\mathrm{b}}\sigma_y^2(m-1) + (1-\lambda_{\mathrm{b}})\underline{\mathbf{y}}^H(m)\underline{\mathbf{y}}(m), \\ \xi^2(m) &= \frac{\sum_{k=0}^{K-1} \hat{\underline{\mathbf{h}}}_{\mathrm{b},k}^H(m)\mathbf{s}_k(m)}{\sigma_y^2(m)} \\ \xi(m) &< T \Rightarrow \text{double-talk, } \mu = 0 \\ \xi(m) &\geq T \Rightarrow \text{no double-talk, } \mu' = \mu(1-\lambda) \end{aligned}$$

Echo canceller (foreground filter):

$$\begin{aligned} \mathbf{e}(m) &= \mathbf{y}(m) - \mathbf{W}_{N\times 2N}^{01}\mathbf{F}_{2N}^{-1}\mathbf{X}(m)\hat{\underline{\mathbf{h}}}(m-1) \\ \psi'_{\min}(m) &= \max\left[\mu, \min_{0\leq n\leq N-1}\left\{\psi'\left[\frac{|e(mN+n)|}{s}\right]\right\}\right] \\ \hat{\underline{\mathbf{h}}}(m) &= \hat{\underline{\mathbf{h}}}(m-1) + \frac{\mu' s(m)}{\psi'_{\min}(m)}\mathbf{G}_2\mathbf{K}(m)\underline{\psi}[\mathbf{e}(m)] \\ s(m+1) &= \lambda_s s(m) \\ &\quad +(1-\lambda_s)\frac{s(m)}{N\beta}\sum_{n=mN}^{mN+N-1}\psi\left[\frac{|e(n)|}{s(m)}\right] \end{aligned}$$

than the forgetting factor λ used for the system identification by the foreground EMDF algorithm. This way, the DTD alerts the foreground filter before it diverges by freezing its adaptation during double-talk. Furthermore, for practical reasons, even though not mathematically stringent, we can use the same spectral matrix $\mathbf{S}'(m)$, and, thus, the same Kalman gain $\mathbf{K}(m)$ for the foreground and background filters. Table I summarizes the combination of the robust EMDF echo canceller and the EMD DTD. Fast calculation schemes for the jointly used frequency-domain Kalman gain are discussed next in Section V.

V. Fast Implementations of the Robust EMDF

When reducing the computational complexity of the robust EMDF algorithm, it is important to keep in mind that the data among the partitions are not independent. Due to the formal similarity of (36), (37), (46), and (47), with the RLS algorithm in the time domain [1], [16], corresponding *fast* implementations of the Kalman gain (47) can be expected. In [4], it is shown that all fast calculation schemes known for the RLS can actually be applied to the EMDF after a slight modification. The key to fast RLS realizations is the shift-structure of the input signal vector [1], [16]. In the case of the EMDF, there is a corresponding shift-structure among the partitions (in each frequency bin $\nu, \nu = 0, \ldots, 2N-1$). Exploiting this property, the complexity increases only linearly with the number of partitions K (instead of quadratically as with the ordinary EMDF algorithm), and the overall complexitiy is on the same order as for the classical MDF. Note that this scheme may be seen as a *divide-and-conquer* method, similarly to the classical subband structure [17]. However, the main difference is that the EMDF is based on a wideband optimization, as reflected by the constraint matrices in (34)–(37). As an example, a fast EMDF algorithm based on the so-called fast transversal filter (FTF) structure [16] is given in Table II. The FTF can be derived by using the *a priori* Kalman gain $\mathbf{K}^{(\nu)}(m) = (\mathbf{S}^{(\nu)})^{-1}(m-1)\mathbf{X}^{(\nu)H}(m)$, where $\mathbf{X}^{(\nu)}(m)$ is a length K row *vector*. This *a priori* Kalman gain for each block of N output samples can be computed recursively by $5KN$ (complex) multiplications (see Section VI for a more detailed discussion of the computational complexity). "Stabilized" versions of FTF (with L, respectively KN more multiplications) exist in the literature, but with nonstationary signals like speech, they are not much more stable than their nonstabilized counterparts. A simple remedy is to re-initialize the predictor-based variables when instability is detected with the use of the maximum likelihood variable $\varphi^{(\nu)}$ which is an inherent variable of the fast algorithm [16].

TABLE II
Frequency-Domain Kalman Gain by an FTF-Based Fast EMDF Implementation

			Arithm. OPs
$\mathbf{X}^{(\nu)}(m)$	$\leftarrow$	$\mathbf{X}_k(m),\ \nu = 0, \ldots, 2N-1,$ $k = 0, \ldots, K-1$	
$\underline{e}_a^{(\nu)}(m)$	$=$	$X_0^{(\nu)*}(m)$ $-\underline{\mathbf{a}}^{(\nu)H}(m-1)\mathbf{X}^{(\nu)H}(m-1)$	$8K$
$\varphi_1^{(\nu)}(m)$	$=$	$\varphi^{(\nu)}(m-1) + \frac{\lvert\underline{e}_a^{(\nu)}(m)\rvert^2}{E_a^{(\nu)}(m-1)}$	5
$\begin{bmatrix} \mathbf{t}^{(\nu)}(m) \\ M^{(\nu)}(m) \end{bmatrix}$	$=$	$\begin{bmatrix} 0 \\ \mathbf{K}_1^{(\nu)}(m-1) \end{bmatrix}$ $+\begin{bmatrix} 1 \\ -\underline{\mathbf{a}}^{(\nu)}(m-1) \end{bmatrix}\frac{\underline{e}_a^{(\nu)}(m)}{E_a^{(\nu)}(m-1)}$	$8K+2$
$E_a^{(\nu)}(m)$	$=$	$\lambda\left(E_a^{(\nu)}(m-1) + \frac{\lvert\underline{e}_a^{(\nu)}(m)\rvert^2}{\varphi^{(\nu)}(m-1)}\right)$	3
$\underline{\mathbf{a}}^{(\nu)}(m)$	$=$	$\underline{\mathbf{a}}^{(\nu)}(m-1)$ $+\mathbf{K}_1^{(\nu)}(m-1)\frac{\underline{e}_a^{(\nu)*}(m)}{\varphi^{(\nu)}(m-1)}$	$8K+2$
$\underline{e}_b^{(\nu)}(m)$	$=$	$E_b^{(\nu)}(m-1)M^{(\nu)}(m)$	2
$\mathbf{K}_1^{(\nu)}(m)$	$=$	$\mathbf{t}^{(\nu)}(m) + \underline{\mathbf{b}}^{(\nu)}(m-1)M^{(\nu)}(m)$	$8K$
$\varphi^{(\nu)}(m)$	$=$	$\varphi_1^{(\nu)}(m) - \underline{e}_b^{(\nu)*}(m)M^{(\nu)}(m)$	7
$E_b^{(\nu)}(m)$	$=$	$\lambda\left(E_b^{(\nu)}(m-1) + \frac{\lvert\underline{e}_b^{(\nu)}(m)\rvert^2}{\varphi^{(\nu)}(m)}\right)$	2
$\underline{\mathbf{b}}^{(\nu)}(m)$	$=$	$\underline{\mathbf{b}}^{(\nu)}(m-1) + \mathbf{K}_1^{(\nu)}(m)\frac{\underline{e}_b^{(\nu)*}(m)}{\varphi^{(\nu)}(m)}$	$8K$
$\mathbf{K}^{(\nu)}(m)$	$=$	$\frac{\mathbf{K}_1^{(\nu)}(m)}{\varphi^{(\nu)}(m)}$	1
$\mathbf{K}(m)$	$\leftarrow$	$\mathbf{K}^{(\nu)}(m),\ \nu = 0, \ldots, 2N-1$	
			Total: $40K+24$

VI. Computational Complexity

We study the computational complexity in terms of arithmetic operations, i.e., the number of real multiplications, real additions, real subtractions, and real divisions. Thereby, each complex multiplication is realized by four real multiplications and two real additions, and each complex addition is realized by two real additions. Moreover, if a Fourier transform of length N is computed using the FFT routine devised by [18], it requires $(N/2)\log_2[N] - (5N/4)$ real multiplications and $(3N/2)\log_2[N] - (N/4) - 4$ additions, giving a total of

$$2N\log_2[N] - \frac{3N}{2} - 4 \tag{60}$$

operations.

A. EMDF

Table III shows the computational complexity of the algorithm steps of the conventional MDF and FEMDF. For the detailed analysis of the FTF-based fast Kalman gain computation for FEMDF, we refer to Table II.

Moreover, we consider here as references the well-known real-valued NLMS [1] and fast RLS (FRLS, real FTF in fullband version analogously to Table II) algorithms. The number of real operations per output sample for the different algorithms is summarized in Table IV. The illustration in Fig. 4 clearly shows the effectiveness of the proposed approach from the complexity point of view. Thereby, the two uppermost curves on the logarithmic scale correspond to FRLS and NLMS, respectively. The complexities of MDF and FEMDF are shown exemplarily for various block lengths $N = 64, N = 512, N = 1024$. Note

TABLE III
COMPLEXITY ANALYSIS FOR MDF/EMDF

			Arithmetic OPs
Input:			
$\mathbf{X}_k(m)$	$=$	$\mathrm{diag}\{\mathbf{F}[x(mN-Ni-N),\dots$	
		$\dots,x(mN-Ni+N-1)]^T\}$	
	$=$	$\mathbf{X}_0(m-k)$,	
		$k=0,\dots,K-1$	$4N\log_2[2N]-3N-4$
$\mathbf{X}(m)$	$=$	$[\mathbf{X}_0(m),\dots,\mathbf{X}_{K-1}(m)]$	
Kalman gain:			*MDF approx.:*
$\mathbf{S}'(m)$	$=$	$\lambda\mathbf{S}'(m-1)$	
		$+(1-\lambda)\mathbf{X}^H(m)\mathbf{X}(m)$	$6N$
$\mathbf{K}(m)$	$=$	$\mathbf{S}'^{-1}(m)\mathbf{X}^H(m)$	$2L$
			FEMDF
			after Table II:
			$40L+24N$
Filtering:			
$\mathbf{e}(m)$	$=$	$\mathbf{y}(m)$	
		$-\mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N}\mathbf{X}(m)\hat{\underline{\mathbf{h}}}(m-1)$	$6L-2N$
			$+4N\log_2[2N]-4$
$\underline{\mathbf{e}}(m)$	$=$	$\mathbf{F}_{2N}\left[\mathbf{e}^T(m)\ \mathbf{0}_{1\times N}\right]^T$	$4N\log_2[2N]-3N-4$
$\hat{\underline{\mathbf{h}}}(m)$	$=$	$\hat{\underline{\mathbf{h}}}(m-1)+\mu'\mathbf{G}_2\mathbf{K}(m)\underline{\mathbf{e}}(m)$	$2N+2L-8K$
			$+8L\log_2[2N]$

TABLE IV
NUMBER OF REAL OPERATIONS PER OUTPUT SAMPLE FOR DIFFERENT ALGORITHMS

Alg.	**Operations per sample**
MDF	$10K-8\frac{K}{N}-\frac{12}{N}+4(2K+3)\log_2[2N]$
FEMDF (FTF)	$48K-8\frac{K}{N}+18-\frac{12}{N}+4(2K+3)\log_2[2N]$
NLMS	$4L+7$
FRLS (FTF)	$14L+15$

that the complexities of FEMDF and MDF are on the same low order.

B. EMD DTD and Robustness Enhancement

To analyze the complexity of the integrated system after Table I, we build upon the results from above.

To begin with, we consider the robustness enhancement of the foreground filter. It consists of an additional real scalar factor

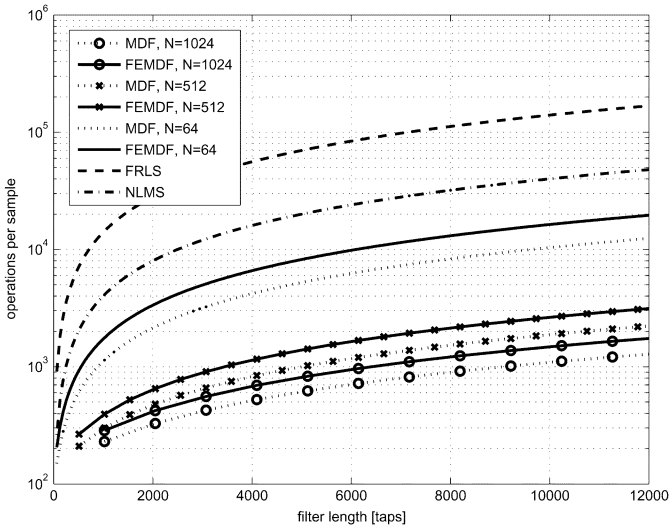


Fig. 4. Illustration of computational complexities for various block lengths.

TABLE V
COMPLEXITY ANALYSIS FOR EMD DTD

Step	**Arithm. OPs per length-N block**
$\underline{\mathbf{e}}_\mathrm{b}$, $\hat{\underline{\mathbf{h}}}_\mathrm{b}$ (as in Table III)	$8L-3N-8K-8$
	$+8(L+N)\log_2[2N]$
$\mathbf{s}_k$, $k=0,\dots,K-1$	$10KN$
σ_y^2	$4N+2$
ξ^2	$4KN+K$

requiring 2 ops per block. Assuming a Huber estimator, i.e., $\psi[\cdot]$ is simply a limiter after (29), the calculation of $\psi'_{\min}$ and the estimation of $s(m)$ require another $3N+4$ ops per block, giving a total of only three additional operations per sample.

As indicated above, the frequency-domain Kalman gain can be used jointly for DTD and foreground filter. By exploiting this synergy, the additional operations for the EMD DTD come from four steps, as summarized in Table V.

This gives a total of $22KN+N-7K-6+8N(K+1)\log_2[2N]$ additional operations per block, or

$$22K+1-\frac{7K}{N}-\frac{6}{N}+8(K+1)\log_2[2N]$$

additional operations per sample. It can easily by verified that the overall complexity of the system is still well below the complexity of the classical NLMS algorithm. For example, $N=64, K=50, L=KN=3200$ as used for simulations below, gives 3951 operations per sample in addition to the 5296 operations per sample needed for the filter update. In contrast, the classical NLMS update without DTD requires 12 007 operations for these parameter settings. Moreover, a corresponding DTD in the time domain would further increase that complexity by the same order due to the algorithmic similarity of the cancellation part and the background filter part.

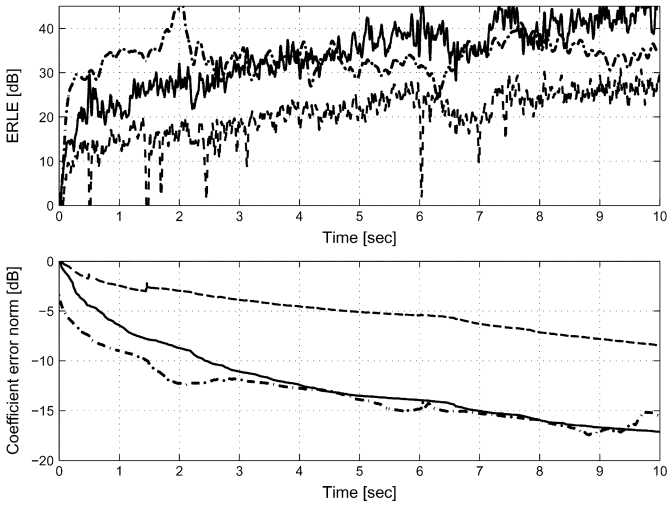


Fig. 5. Comparison between classical (dashed lines) MDF, (dash-dot) fast RLS, and (solid lines) EMDF for filter length $L = 50 \cdot 64 = 3200$.

VII. Experimental Results

In this section, we apply the single-channel robust EMDF algorithm for single-channel AEC with a block length $N = 64$ for each partition, and a high sampling rate of 48 kHz. In general, very high filter orders are needed for this sampling rate. In order to compare the FEMDF performance also with the FRLS, two different acoustic scenarios using real recordings are considered. First, we consider a low reverberation which can be modeled by $K = 50$ partitions, i.e., a filter length of $L = 50 \cdot 64 = 3200$. Moreover, in the second case, a more realistic environment with an impulse response of length $L = 150 \cdot 64 = 9600$ is considered. As an input signal, we chose classical music (*Air* by Bach) in both cases. The signal sequence is highly auto-correlated (tonal sounds, which are known as worst case for the adaptation).

A. Convergence Speed During Single-Talk

To begin with, we compare the performance of different algorithms for $L = 3200$ without doubletalk (i.e., $v(n) = 0$) in Fig. 5. In this case, an echo-to-background noise ratio (EBR) of 45 dB on the microphone was chosen. The dashed and dash-dotted lines in Fig. 5 show the *echo return loss enhancement* $\mathrm{ERLE}(n) = \overline{d^2(n)}/\overline{(e(n) - w(n))^2}$ and the coefficient error norm $\|\mathbf{h} - \hat{\mathbf{h}}(n)\|^2/\|\boldsymbol{h}\|^2$ achieved by the conventional MDF and the fast RLS, respectively, as the extreme cases for a given maximum block size (Fig. 2). For the solid lines, the same data and the same parameters are used with the EMDF algorithm. It is important to note that the regularization is adjusted in each case to ensure stable convergence. Several simulations have confirmed that the EMDF shows a significantly more stable behavior than the classical MDF due to the more accurate approximation to the exact recursive solution of the normal equation while the complexity is kept low as demonstrated in Section VI. In Fig. 6, the corresponding simulation results for filter length $L = 9600$ and $K = 150$ partitions are shown. Here, the FRLS is not considered due to stability reasons. In contrast, the FEMDF still shows a very reliable convergence behavior in this case.

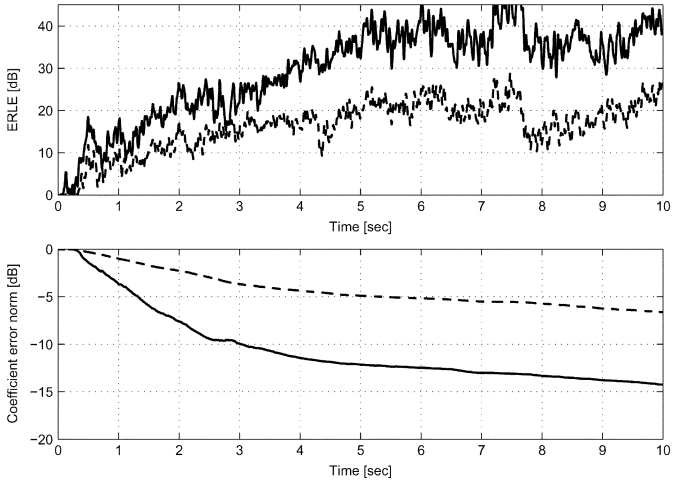


Fig. 6. Comparison between classical (dashed lines) MDF and (solid lines) EMDF for filter length $L = 150 \cdot 64 = 9600$.

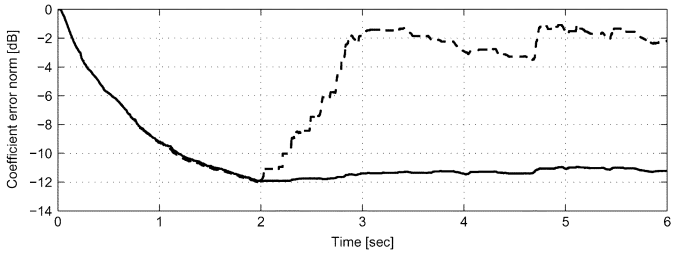


Fig. 7. Robustness during double-talk: (dashed line) nonrobust EMDF and (solid line) robust EMDF. Near-end speech after 2 s.

B. Robustness During Double-Talk

In Fig. 7, we compare the coefficient error norm of the robust and nonrobust EMDF algorithms in the double-talk case combined with the EMD DTD from Section IV. For simplicity, only the first acoustic scenario is considered here. The speech signal on the near end sets in after 2 s. While the convergence in the single-talk case (from 0 to 2 s) is unaffected by the robustness enhancement, the divergence due to inevitable detection failures is effectively avoided.

C. EMD DTD

In Section IV, we have seen that the EMDF and the EMD DTD can be very efficiently linked from an algorithmic point of view. Fig. 8 illustrates that this also holds in terms of detection performance. Again, the speech signal on the near end sets in after 2 s. The robust EMDF has been applied in the foreground in both cases shown in Fig. 8. While the DTD based on the conventional MDF (i.e., Table I with MDF as background filter) cannot keep track of the foreground filter, so that several false alarms (0 to 2 s) hinder convergence and divergence occurs after 2 s, the EMD DTD is perfectly matched to the foreground EMDF.

VIII. On the Extension to the Multichannel Case

In this section, we briefly outline the efficient applicability of the EMDF concept to the multichannel case, especially to multiple strongly cross-correlated loudspeaker channels.

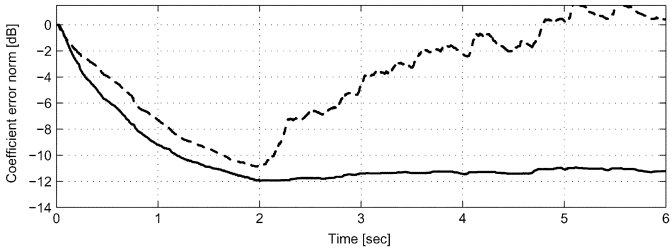


Fig. 8. DTD performance: robust EMDF combined with (dashed line) MD DTD and robust EMDF combined with (solid line) EMD DTD in the double-talk case. Near-end speech after 2 s.

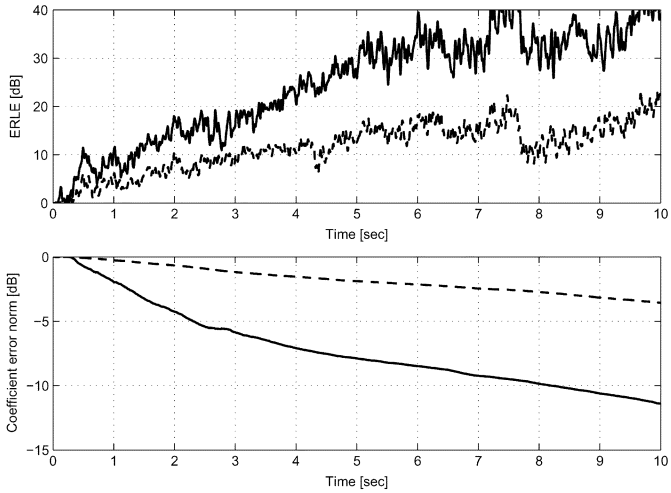


Fig. 9. Comparison between (dashed lines) two-channel MDF (solid lines) and two-channel EMDF for filter length $L = 150 \cdot 64 = 9600$.

The key for efficient multichannel AEC is to take into account not only the auto-correlations in matrix $\mathbf{S}_{\mathrm{d},\psi'}$, but also all cross-correlations between the input channels [19]. As shown in [5], the multichannel extension of frequency-domain adaptive filtering inherently exhibits this property and is, thus, very suitable for stereophonic AEC or even five-channel surround sound applications. By concatenating the coefficient vectors of the individual channels in [5] to an overall coefficient vector $\hat{\underline{\mathbf{h}}}(m)$, as done above for multiple partitions, we can directly link the multipartition equations (34)–(37) to the multichannel equations (4.57)–(4.60) of [5] as these sets of equations exhibit the same structure. As a generalization, we can treat any combination of filter decompositions leading to such a segmentation of the coefficient vector, which also includes, e.g., nonlinear AEC based on the class of polynomial filters [20]. In any case, a combination of different types of segmentations leads to a nesting of partitions within the matrix $\mathbf{S}_{\mathrm{d},\psi'}$ for which the EMDF may favorably be applied.

In the case of multiple reproduction channels, the EMDF is particularly attractive compared to the standard MDF as the MDF not only approximates the auto-correlation matrices according to Fig. 3, but also all cross-correlation matrices which may cause severe degradation of the convergence speed and misalignment. In contrast, the EMDF fully accounts for both, the auto-correlations and cross-correlations. Fig. 9 shows the results with the two-channel MDF and the two-channel FEMDF for the stereo version of the music signal as above and the same environment as used for Fig. 6.

IX. Conclusion

We presented a system for AEC based on a novel class of robust frequency-domain adaptive filtering algorithms. This class combines several very desirable properties of RLS and conventional FDAF and MDF, particularly for very long adaptive filters and high sampling rates. Due to the rigorous derivation of the new algorithm with a block size $N \leq L$, we found a natural way of efficiently taking all cross-correlations between the partitions into account. As shown by way of simulations, the algorithm can lead to a significant improvement of the convergence speed over the MDF, i.e., the conventional frequency-domain algorithm with partitioned blocks. Moreover, we have shown that the proposed approach also carries over to a very efficient treatment of the double-talk problem by integrating the concept of robust statistics and introducing an EMDF-based DTD that is very well matched to the new system. Moreover, we proposed fast calculation schemes for the frequency-domain Kalman gain that give a low-complexity EMDF solution, where the computational complexity can be kept on the same order as that of the conventional MDF. The concept can be efficiently extended to the multichannel case.

Appendix I
Newton Update (23)

Based on a second-order Taylor expansion of (22), a numerical off-line optimization for $\hat{\underline{\mathbf{h}}}$ is given for any choice of $\beta(i,m)$ by the well-known iterative Newton-Raphson algorithm (e.g., [15]) which reads in complex form

$$\hat{\underline{\mathbf{h}}}(m) = \hat{\underline{\mathbf{h}}}(m-1) - 2\mathbf{S}_{\psi'}^{-1}(m)\nabla J[m, \hat{\underline{\mathbf{h}}}(m-1)] \tag{61}$$

with the gradient of $J(m, \hat{\underline{\mathbf{h}}})$

$$\nabla J[m, \hat{\underline{\mathbf{h}}}(m-1)] = \sum_{i=0}^{m} \beta(i,m) \nabla \tilde{J}(i, \hat{\underline{\mathbf{h}}}) \tag{62}$$

and the Hessian matrix

$$\begin{aligned}\mathbf{S}_{\psi'}(m) &= \nabla\nabla^H J(m, \hat{\underline{\mathbf{h}}}) = 2\partial/\partial \hat{\underline{\mathbf{h}}}^* (\nabla J(m, \hat{\underline{\mathbf{h}}}))^H \\ &= \sum_{i=0}^{m} \beta(i,m) \nabla\nabla^H \tilde{J}(i, \hat{\underline{\mathbf{h}}}).\end{aligned} \tag{63}$$

Several types of robust recursive algorithms can be derived from these equations for certain choices of $\beta(i,m)$ and certain approximations. To obtain a rank-one (RLS-like) adaptive Newton algorithm, we choose $\beta(i,m) = (1-\lambda)\lambda^{m-i}$ with the forgetting factor λ, and use the following two approximations [15, p. 329]:

$$\nabla J(m-1, \hat{\underline{\mathbf{h}}}(m-1)) = \mathbf{0} \tag{64}$$

$$\nabla\nabla^H J(m-1, \hat{\underline{\mathbf{h}}}(m-1)) = \nabla\nabla^H J(m-1, \hat{\underline{\mathbf{h}}}(m-2)). \tag{65}$$

In (64), we assume that $\hat{\underline{\mathbf{h}}}(m-1)$ is the minimum point of $\nabla J(m-1, \hat{\underline{\mathbf{h}}})$ to simplify (62) to

$$\begin{aligned}\nabla J(m, \hat{\underline{\mathbf{h}}}) &= \beta(m,m) \nabla \tilde{J}(m, \hat{\underline{\mathbf{h}}}) \\ &= (1-\lambda) \nabla \tilde{J}(m, \hat{\underline{\mathbf{h}}}).\end{aligned} \tag{66}$$

The approximation (65) means that the Hessian varies slowly with $\hat{\underline{\mathbf{h}}}$ which allows a recursive calculation of (63) with the forgetting factor λ. Equation (61) with (66) and (63) with (65) give directly (23) and (24), respectively.

APPENDIX II
PRACTICAL REFORMULATION (34)–(37) OF THE GENERIC ALGORITHM

Here, we summarize the steps leading from (30), (31), and (14) to the practically more useful form (34)–(37). The advantages of this equivalent formulation are that it involves exclusively DFTs of length $2N$, and that the relation to some known algorithms, such as the conventional MDF can be established.

To begin with, we introduce the zero-padded coefficient vector $\hat{\underline{\mathbf{h}}}_{2L}(m)$, as defined in (38) and (39). Its relation to the length-L vector $\hat{\underline{\mathbf{h}}}(m)$ can be expressed conveniently using (16) as

$$\hat{\underline{\mathbf{h}}}_{2L}(m) = \mathbf{G}_{2L\times L}^{10}\hat{\underline{\mathbf{h}}}(m). \tag{67}$$

Using this relation, (36) follows immediately from (14). Next, the coefficient update (30) is premultiplied by $\mathbf{G}_{2L\times L}^{10}$ on both sides so that we obtain

$$\hat{\underline{\mathbf{h}}}_{2L}(m) = \hat{\underline{\mathbf{h}}}_{2L}(m-1) + \frac{2\mu(1-\lambda)}{Ns} \cdot \mathbf{G}_{2L\times L}^{10}\mathbf{S}_{\psi'}^{-1}(m)(\mathbf{G}_{2L\times L}^{10})^H\mathbf{X}^H(m)\underline{\boldsymbol{\psi}}[\mathbf{e}(m)]. \tag{68}$$

This update equation can be simplified by introducing the matrix $\mathbf{S}_{\mathrm{d},\psi'}$, defined in (34). The relation of $\mathbf{S}_{\mathrm{d},\psi'}(m)$ to $\mathbf{S}_{\psi'}(m)$ from (31) is obviously given by

$$\mathbf{S}_{\psi'}(m) = \left(\mathbf{G}_{2L\times L}^{10}\right)^H \mathbf{S}_{\mathrm{d},\psi'}(m)\mathbf{G}_{2L\times L}^{10}. \tag{69}$$

Finally, in order to obtain the coefficient update (37) from (68), we verify the following relation between the inverses of the two matrices $\mathbf{S}_{\mathrm{d},\psi'}$ and $\mathbf{S}_{\psi'}$

$$\mathbf{G}_{2L\times 2L}^{10}\mathbf{S}_{\mathrm{d},\psi'}^{-1}(m) = \mathbf{G}_{2L\times L}^{10}\mathbf{S}_{\psi'}^{-1}(m)\left(\mathbf{G}_{2L\times L}^{10}\right)^H \tag{70}$$

where

$$\begin{aligned}\mathbf{G}_{2L\times 2L}^{10} &= \mathrm{diag}\left\{\mathbf{G}_{2N\times 2N}^{10},\ldots,\mathbf{G}_{2N\times 2N}^{10}\right\}\\ \mathbf{G}_{2N\times 2N}^{10} &= \mathbf{F}_{2N}\mathbf{W}_{2N\times 2N}^{10}\mathbf{F}_N^{-1}\\ &= \mathbf{G}_{2N\times N}^{10}(\mathbf{G}_{2N\times N}^{10})^H\\ \mathbf{W}_{2N\times 2N}^{10} &= \begin{bmatrix}\mathbf{I}_{N\times N} & \mathbf{0}_{N\times N}\\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N}\end{bmatrix}.\end{aligned}$$

The relation (70) can now be verified by post-multiplying both sides by $\mathbf{S}_{\mathrm{d},\psi'}(m)\mathbf{G}_{2L\times L}^{10}$ and noting that $\mathbf{G}_{2L\times 2L}^{10}\mathbf{G}_{2L\times L}^{10} = \mathbf{G}_{2L\times L}^{10}$.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1996.

[2] J.-S. Soo and K. K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 2, pp. 373–376, Feb. 1990.

[3] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, "An outlier-robust extended multidelay filter with application to acoustic echo cancellation," in *Proc. Int. Workshop Acoustic Echo and Noise Control*, Sep. 2003, pp. 19–22.

[4] H. Buchner, J. Benesty, and W. Kellermann, "An extended multidelay filter: Fast low-delay algorithms for very high-order adaptive systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Apr. 2003, pp. 385–388.

[5] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive algorithms with application to acoustic echo cancellation," in *Adaptive Signal Processing: Application to Real-World Problems*, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer, Feb. 2003, pp. 95–128.

[6] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.

[7] H. Ye and B.-X. Wu, "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. Commun.*, vol. 39, no. 11, pp. 1542–1545, Nov. 1991.

[8] T. Gänsler, M. Hansson, C.-J. Ivarsson, and G. Salomonsson, "A double-talk detector based on coherence," *IEEE Trans. Commun.*, vol. 44, no. 11, pp. 1421–1427, Nov. 1996.

[9] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.

[10] P. J. Huber, *Robust Statistics*. New York, NY: Wiley, 1981.

[11] T. Gänsler, "A double-talk resistant subband echo canceller," *Signal Process.*, vol. 65, pp. 89–101, Feb. 1998.

[12] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[13] J. Benesty *et al.*, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer, 2001.

[14] F. Hampel *et al.*, *Robust Statistics*. New York: Wiley, 1986.

[15] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice Hall, 1989.

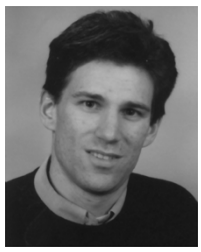
[16] M. G. Bellanger, *Adaptive Digital Filters and Signal Analysis*. New York: Marcel Dekker, 1987.

[17] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Apr. 1988, pp. 2570–2573.

[18] H. V. Sorensen, D. L. Jones, M. T. Heideman, and C. S. Burrus, "Real-valued fast Fourier transform algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 3, pp. 849–863, Jun. 1987.

[19] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[20] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Process. Mag.*, vol. 8, no. 7, pp. 10–26, Jul. 1991.



Herbert Buchner (S'01-M'04) received the Dipl.-Ing. (FH) degree in electrical engineering from the University of Applied Sciences, Regensburg, Germany, in 1997, and the Dipl.-Ing. Univ. degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 2000.

He is a member of the research staff at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg. In 1995, he was a Visiting Researcher at the Colorado Optoelectronic Computing Systems Center (OCS), Boulder/Ft. Collins, where he worked in the field of microwave technology. From 1996 to 1997, he did research at the Cyber Space Labs (former Human Interface Labs) of the R&D division of Nippon Telegraph and Telephone Corporation (NTT), Tokyo, Japan, working on adaptive filtering for teleconferencing. In 1997 and 1998, he was with the Driver Information Systems Department, Siemens Automotive, Regensburg, Germany. His current areas of interest include efficient multichannel algorithms for adaptive digital filtering and their applications for acoustic human-machine interfaces, such as multichannel acoustic echo cancellation, beamforming, blind source separation, and dereverberation.

Mr. Buchner received the VDI award in 1998 for his Dipl.-Ing. (FH) thesis from the Verein Deutscher Ingenieure and a best student paper award in 2001.

Jacob Benesty (M'98–SM'04) was born in 1963. He received the M.S. degree in microwaves from the Pierre and Marie Curie University, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During the Ph.D. program (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris, France.

From January 1994 to July 1995, he was with Telecom Paris University, working on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In May 2003, he joined the University of Quebec, INRS-EMT, Montréal, QC, Canada, as an Associate Professor. He co-authored the book *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001). He is also a Co-Editor/Co-Author of the books *Speech Enhancement* (Springer-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia communication Systems* (Kluwer, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunication* (Kluwer, 2000). His research interests are in acoustic signal processing and multimedia communications.

Dr. Benesty received the 2001 Best Paper Award from the IEEE Signal Processing Society. He is a member of the editorial board of the *EURASIP Journal on Applied Signal Processing* and was the Co-Chair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Tomas Gänsler (S'91–M'96) was born in Sweden in 1966. He received his M.S. degree in electrical engineering and the Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1990 and 1996, respectively.

From 1997 to September 1999, he was an Assistant Professor at Lund University. In 1998, he was with Bell Labs, Lucent Technologies, as a Consultant, and in October 1999, as a Member of Technical Staff. Since 2001, he has been with Agere Systems, Allentown, PA, a spinoff from Lucent Technologies' Microelectronics Group. He co-authored the books *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001) and *Acoustic Signal Processing for Telecommunication* (Kluwer, 2000). His research interests include robust estimation, adaptive filtering, mono/multichannel echo cancellation, and subband signal processing.

Walter Kellermann (M'89) received the Dipl.-Ing. (univ.) degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1983, and the Dr.-Ing. degree from the Technical University Darmstadt, Darmstadt, Germany, in 1988.

He is a Professor for communications at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg. From 1989 to 1990, he was a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany. From 1993 to 1999, he was a Professor at the Fachhochschule Regensburg, Regensburg, Germany, where he became a Director of the Institute of Applied Research in 1997. In 1999, he co-founded DSP Solutions, a consulting firm in digital signal processing, and he joined the University Erlangen-Nuremberg as a Professor and Head of the Audio Research Laboratory. He has authored or co-authored six book chapters and more than 60 refereed papers in journals and conference proceedings. His current research interests include speech signal processing, array signal processing, adaptive filtering, and its applications to acoustic human/machine interfaces.

Dr. Kellermann has served as a Guest Editor for various journals, as an Associate Editor and Guest Editor for the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING from 2000 to 2004, and presently serves as Associate Editor for the *EURASIP Journal on Applied Signal Processing*. He was the General Chair of the 5th International Workshop on Microphone and Signal Processing Arrays in 2003 and he served as General Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2005.