

Multichannel Bin-Wise Robust Frequency-Domain Adaptive Filtering and Its Application to Adaptive Beamforming

Wolfgang Herbordt, Herbert Buchner, *Member, IEEE*, Satoshi Nakamura, *Member, IEEE*, and Walter Kellermann, *Senior Member, IEEE*

Abstract—Least-squares error (LSE) or mean-squared error (MSE) optimization criteria lead to adaptive filters that are highly sensitive to impulsive noise. The sensitivity to noise bursts increases with the convergence speed of the adaptation algorithm and limits the performance of signal processing algorithms, especially when fast convergence is required, as for example, in adaptive beamforming for speech and audio signal acquisition or acoustic echo cancellation. In these applications, noise bursts are frequently due to undetected double-talk. In this paper, we present impulsive noise robust multichannel frequency-domain adaptive filters (MC-FDAFs) based on outlier-robust M -estimation using a Newton algorithm and a discrete Newton algorithm, which are especially designed for frequency bin-wise adaptation control. Bin-wise adaptation and control in the frequency-domain enables the application of the outlier-robust MC-FDAFs to a generalized sidelobe canceler (GSC) using an adaptive blocking matrix for speech and audio signal acquisition. It is shown that the improved robustness leads to faster convergence and to higher interference suppression relative to nonrobust adaptation algorithms, especially during periods of strong interference.

Index Terms—Adaptive filtering, beamforming, doubletalk, frequency domain, generalized sidelobe canceler, microphone array, robust statistics.

I. INTRODUCTION

ADAPTIVE filters that are based on least-squares error (LSE) or mean-squared error (MSE) optimization criteria are highly sensitive to impulsive noise [1]. For applications such as acoustic echo cancellation or adaptive beamforming, this sensitivity to impulsive noise leads to a sensitivity to undetected double-talk bursts and limits the performance of the system. For acoustic echo cancellation, the need for robustness against double-talk bursts due to the presence of a local disturbance

was first discussed in [2]. An outlier-robust least mean-squares (LMS) algorithm using the median of the error signal for adaptation is proposed in [3]. Robust mixed-norm adaptive filters using a weighted combination of L_1 and L_2 -norm cost functions and a weighted combination of L_2 and L_4 -norm cost functions are proposed in [4], [5], and [6], respectively. Using robust maximum-likelihood type cost functions (so-called “robust M -estimates”) and an ϵ -contaminated Gaussian model [1] for the residual error signal of the adaptive filter, robustness against impulsive noise was obtained for a large variety of adaptation algorithms. Using Huber’s so-called “least-informative distribution” [1], an outlier-robust frequency-subband adaptive filter was proposed in [7]. Outlier-robust versions of the normalized LMS (NLMS) algorithm, the proportionate NLMS (PNLMS) algorithm, and the affine projection algorithm (APA) are derived in [8], and outlier-robust recursive least-squares (RLS) algorithms are proposed in [9] and [10]. A robust block-based adaptive filter based on sample matrix inversion is presented in [11]. Robust RLS algorithms using Hampel’s estimator [12] and a tanh-weighted cost function are derived in [13], [14], and [15], respectively. A robust LMS algorithm and a robust Kalman filter using linear combinations of order statistics (“ L -estimation”) [1] are derived in [16] and [17], respectively. Using a likelihood ratio test for detecting the presence of impulses and temporarily freezing the adaptation of the adaptive filter is proposed in [18].

In recent years, adaptation algorithms in the discrete Fourier transform (DFT) domain [*frequency-domain adaptive filters* (FDAFs)] [19]–[25] became very attractive for acoustic echo cancellation [26]–[32] and for adaptive beamforming [33]–[35] since they 1) combine fast convergence with low computational complexity; 2) can be realized such that sufficiently high tracking capability and sufficiently low delay are obtained; and 3) generalize well to the multichannel (MC) case (MC-FDAF) [27]–[29], [31]. Similar to RLS algorithms, the convergence speed of FDAFs is nearly independent of the condition number of the cross-correlation matrix of the input signals. In the context of adaptive beamforming, the improvement of the convergence speed of the MC-FDAF relative to single-channel FDAFs is illustrated in [31] for the GSC. This is especially important to assure fast convergence for highly auto- and cross-correlated input signals, such as speech or music. Furthermore, in combination with a DFT-bin-wise double-talk detector and/or a DFT-bin-wise step-size control, adaptive filters in the DFT domain allow for DFT-bin-wise adaptation. This is especially advantageous for signals that are sparse in the

Manuscript received January 5, 2006; revised October 23, 2006. This work was supported in part by the National Institute of Information and Communications Technology (NICT), Japan. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Davies.

W. Herbordt was with the Spoken Language Communication Research Laboratories, Advanced Telecommunications Research (ATR) International, Kyoto 619-0288, Japan. He is now with the Center of Competence Signal Processing, Rohde and Schwarz, 81671 Munich, Germany.

H. Buchner was with Multimedia Communication and Signal Processing, University Erlangen-Nuremberg, 91058 Erlangen, Germany. He is now with the Deutsche Telekom Laboratories, 10587 Berlin, Germany.

S. Nakamura is with the Spoken Language Communication Research Laboratories, Advanced Telecommunications Research (ATR) International, Kyoto 619-0288, Japan.

W. Kellermann is with Multimedia Communication and Signal Processing, University Erlangen-Nuremberg, 91058 Erlangen, Germany.

Digital Object Identifier 10.1109/TASL.2007.894530

time-frequency domain, since the step size can be adjusted for each frequency bin independently. This possibility of a bin-wise adaptation control also plays a major role in the present paper.

For better robustness against impulsive noise, such as double-talk bursts, an FDAF based on robust M -estimation and on a Newton-type adaptation algorithm is derived and applied to acoustic echo cancellation in [36], [37]. In contrast to the subband robust adaptive filter [7], where a nonlinear cost function is applied to each subband (“narrowband decomposition”), and the error signal is minimized in each subband individually, [36], [37] minimize the fullband error signal in the discrete-time domain. However, due to the time-domain optimization criterion, [36] and [37] cannot be directly used for DFT-bin-wise adaptation control, which is necessary for some applications of DFT-domain adaptive filtering, such as adaptive beamforming using GSCs with adaptive realizations of the blocking matrix [34].

To enable the usage of bin-wise adaptation control, we consider in this work an outlier-robust cost function in the DFT domain based on a weighting according to Huber [1]. We present an impulsive noise robust MC-FDAF for bin-wise adaptation based on the Newton algorithm [*multichannel bin-wise robust FDAF (MC-BRFDAF)*] and a discretized MC-BRFDAF using a so-called “discrete Newton algorithm” [38] for linear multiple-input multiple-output (MIMO) systems. The discretization of the Newton algorithm has the advantage of better stability relative to the MC-BRFDAF leading to improved robustness and slightly improved convergence speed relative to the nondiscretized MC-BRFDAF for multichannel interference cancellation problems.

Using the MC-BRFDAF and the discretized MC-BRFDAF, outlier-robust adaptive filtering is applied for the first time to adaptive beamforming for multichannel speech enhancement using microphone arrays. Using the example of a *generalized sidelobe canceler (GSC)* with an *adaptive blocking matrix* [39], it is shown in [34] that DFT-domain adaptive filtering in combination with bin-wise adaptation control can be efficiently applied to adaptive beamforming. It is further shown that the bin-wise adaptation is necessary to solve the tracking problems of GSCs with adaptive realizations of blocking matrices. Using fullband adaptation, the performance of the GSC is highly degraded especially during double-talk between the desired speaker and interference or noise. The DFT-domain GSCs using bin-wise adaptation are especially efficient for tackling the challenges of beamforming microphone arrays, such as robustness against reverberation and the physical tolerances of the sensors, or the nonstationarity of the desired signal and of interferers.

In this paper, experiments with the MC-BRFDAF and the discretized MC-BRFDAF show that robustness against double-talk can be greatly improved, relative to the GSC using the nonrobust MC-FDAF [34], even with small-scale microphone arrays, so that larger step sizes can be chosen for the adaptation. This leads to faster convergence and to higher noise reduction, while preserving the good output signal quality of the beamformer. These results can easily be generalized to all GSC realizations. Some parts of this paper are already published in [40] and [41].

This paper is organized as follows. The MC-BRFDAF and the discretized MC-BRFDAF are derived in Section II. In

Section III, the robust MC-FDAFs are applied to the GSC. Experimental results are given in Section IV.

II. IMPULSIVE NOISE ROBUST FREQUENCY-DOMAIN ADAPTIVE FILTER

In this section, we develop the aforementioned adaptation algorithms based on robust M -estimation for a linear MIMO system. The presentation is based on [27]–[29], [31], [37], and [40]. Section II-A introduces the notation and describes the filtering in the DFT domain using the overlap-save method. In Section II-B, the optimization criterion for the robust adaptive filter is defined. Section II-C derives the robust adaptive algorithms using the Newton algorithm and the discrete Newton algorithm and shows possible approximations. The computational complexity is investigated in Section II-D.

Lowercase and uppercase bold font represent vector and matrix quantities, respectively, and $[\cdot]^*$, $[\cdot]^T$, and $[\cdot]^H$ stand for complex conjugation, matrix or vector transposition, and conjugate transposition, respectively. Underlined quantities denote DFT-domain variables and k is the discrete time index.

A. Optimum Linear Filtering Using Overlap-Save

We start with formulating the adaptive filtering for a multiple-input single-output (MISO) system, and perform the generalization to MIMO systems in Section II-C3.

A MISO system with Q input channels is described by a vector \mathbf{w} of size $NQ \times 1$, which captures Q column vectors \mathbf{w}_q , $q = 0, 1, \dots, Q-1$, of length N with filter coefficients $w_{q,n}$, $n = 0, 1, \dots, N-1$. That is

$$\mathbf{w} = [\mathbf{w}_0^T \mathbf{w}_1^T, \dots, \mathbf{w}_{Q-1}^T]^T, \quad (1)$$

$$\mathbf{w}_q = [w_{q,0} \ w_{q,1} \ \dots \ w_{q,N-1}]^T. \quad (2)$$

Accordingly, the input signals of the MISO system $x_q(k)$ are captured in the $NQ \times 1$ vector $\mathbf{x}(k)$ as follows:

$$\mathbf{x}(k) = [\mathbf{x}_0^T(k) \ \mathbf{x}_1^T(k) \ \dots \ \mathbf{x}_{Q-1}^T(k)]^T \quad (3)$$

$$\mathbf{x}_q(k) = [x_q(k) \ x_q(k-1) \ \dots \ x_q(k-N+1)]^T. \quad (4)$$

The output signal $e(k)$ of the adaptive MISO system can thus be written as

$$e(k) = y_{\text{ref}}(k) - \mathbf{x}^T(k)\mathbf{w} \quad (5)$$

where $y_{\text{ref}}(k)$ is the reference signal of the adaptive filter. To calculate the output signal of the MISO system in the DFT domain using fast convolution and overlap-save, we form blocks of N samples of $e(k)$, $y_{\text{ref}}(k)$, and $\mathbf{x}(k)$

$$\mathbf{e}(k) = [e(k) \ e(k-1) \ \dots \ e(k-N+1)]^T \quad (6)$$

$$\mathbf{y}_{\text{ref}}(k) = [y_{\text{ref}}(k) \ y_{\text{ref}}(k-1) \ \dots \ y_{\text{ref}}(k-N+1)]^T \quad (7)$$

$$\mathbf{X}(k) = [\mathbf{x}(k) \ \mathbf{x}(k-1) \ \dots \ \mathbf{x}(k-N+1)]^T \quad (8)$$

which yields for (5) the expression

$$\mathbf{e}(k) = \mathbf{y}_{\text{ref}}(k) - \mathbf{X}(k)\mathbf{w}. \quad (9)$$

Next, we transform (9) into the DFT domain using block overlapping. For this purpose, we define a block overlap factor $\alpha =$

N/R , where R is the new number of samples per block, and in (9) we replace the discrete time k by the block time r , which is related to k by $rR = k$. A block-diagonal data matrix $\underline{\mathbf{X}}(r)$ of size $2N \times 2NQ$ in the DFT domain is obtained by transforming blocks of length $2N$ of the filter input signals $x_q(k)$ into the DFT domain using the DFT matrix $\mathbf{F}_{2N \times 2N}$ of size $2N \times 2N$. That is

$$\underline{\mathbf{X}}_q(r) = \text{diag} \left\{ \mathbf{F}_{2N \times 2N} \begin{bmatrix} x_q(rR - N) \\ x_q(rR - N + 1) \\ \vdots \\ x_q(rR + N - 1) \end{bmatrix} \right\} \quad (10)$$

$$\underline{\mathbf{X}}(r) = [\underline{\mathbf{X}}_0(r) \ \underline{\mathbf{X}}_1(r) \ , \dots \ , \ \underline{\mathbf{X}}_{Q-1}(r)] . \quad (11)$$

The MISO system \mathbf{w} is transformed into the DFT domain as the $2NQ \times 1$ vector

$$\underline{\mathbf{w}} = \text{diag} \{ [\mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{10} \ , \dots \ , \ \mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{10}] \} \mathbf{w} \quad (12)$$

where the windowing matrix

$$\mathbf{W}_{2N \times N}^{10} = [\mathbf{I}_{N \times N} \ \mathbf{0}_{N \times N}]^T \quad (13)$$

appends N zeroes to the coefficient vectors \mathbf{w}_q in order to prevent circular convolution. Here, $\mathbf{I}_{N \times N}$ is the identity matrix of size $N \times N$, and $\mathbf{0}_{N \times N}$ is a matrix of size $N \times N$ with all zeroes. We obtain for (9) the expression

$$\mathbf{e}(rR) = \mathbf{y}_{\text{ref}}(rR) - \mathbf{W}_{N \times 2N}^{01} \mathbf{F}_{2N \times 2N}^{-1} \underline{\mathbf{X}}(r) \underline{\mathbf{w}} \quad (14)$$

where the windowing matrix

$$\mathbf{W}_{N \times 2N}^{01} = [\mathbf{0}_{N \times N} \ \mathbf{I}_{N \times N}] \quad (15)$$

extracts a block of N samples from $\mathbf{F}_{2N \times 2N}^{-1} \underline{\mathbf{X}}(r) \underline{\mathbf{w}}$. A block of R adaptive filter output samples is given by the last R samples of $\mathbf{e}(rR)$.

For deriving the DFT-domain adaptation algorithm, the block-error signal $\mathbf{e}(rR)$ is required in the DFT domain. For this reason, we transform (14) into the DFT domain by multiplying (14) with $\mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{01}$ from the left, where $\mathbf{W}_{2N \times N}^{01} = (\mathbf{W}_{N \times 2N}^{01})^T$. This yields the $2N \times 1$ vector

$$\underline{\mathbf{e}}(r) = \underline{\mathbf{y}}_{\text{ref}}(r) - \mathbf{G}_{2N \times 2N}^{01} \underline{\mathbf{X}}(r) \underline{\mathbf{w}} \quad (16)$$

where

$$\underline{\mathbf{e}}(r) = \mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{01} \mathbf{e}(rR) \quad (17)$$

$$\underline{\mathbf{y}}_{\text{ref}}(r) = \mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{01} \mathbf{y}_{\text{ref}}(rR) \quad (18)$$

$$\mathbf{G}_{2N \times 2N}^{01} = \mathbf{F}_{2N \times 2N} \mathbf{W}_{2N \times N}^{01} \mathbf{W}_{N \times 2N}^{01} \mathbf{F}_{2N \times 2N}^{-1} . \quad (19)$$

In the following, the elements of $\underline{\mathbf{e}}(r)$ are denoted by $\underline{e}_n(r)$, $n = 0, 1, \dots, 2N - 1$. The $\underline{e}_n(r)$ are assumed to have zero mean.

B. Outlier-Robust Cost Functions

To derive an outlier-robust recursive adaptation algorithm based on robust M -estimation, we start with a log-likelihood cost function $J(r, \underline{\mathbf{w}})$ in the DFT domain

$$J(r, \underline{\mathbf{w}}) = \sum_{i=0}^r \beta(i, r) \tilde{J}(i, \underline{\mathbf{w}}) \quad (20)$$

$$\tilde{J}(i, \underline{\mathbf{w}}) = - \sum_{n=0}^{2N-1} \log f \left(\frac{|\underline{e}_n(i)|}{\underline{g}_n(i)} \right) \quad (21)$$

which is to be minimized subject to the vector of filter coefficients $\underline{\mathbf{w}}$. The quantity $\beta(i, r)$ is a windowing function. The function $f(x)$ is the given probability density function of the normalized error signal with zero mean and unit variance. The divisor $\underline{g}_n(r)$ is a scale parameter for normalizing the zero-mean error signal $\underline{e}_n(r)$ to unity variance. Note that this cost function implies independence of the frequency bins of the error signal, which is used later for adapting the adaptive filter independently in the frequency bins. The cost function $J(r, \underline{\mathbf{w}})$ minimizes the error signal $\underline{e}_n(r)$, where the choice of the probability density function $f(x)$ defines the weighting of $|\underline{e}_n(i)|/\underline{g}_n(i)$.

The fundamental idea of outlier-robust M -estimation is to find a density $f(x)$ that yields an outlier-robust cost function $J(r, \underline{\mathbf{w}})$. In this paper, we consider the family of distributions that is symmetric to the origin as follows:

$$f(x) = A \exp[-\rho(|x|)] \quad (22)$$

where the constant A is chosen such that $\int_{-\infty}^{\infty} f(x) dx = 1$, and where $\rho(|x|)$ is a weighting function defining the shape of the probability density. The choice of an outlier-robust distribution $f(x)$ is thus reduced to the choice of the weighting rule $\rho(|x|)$. In the following, we consider Huber's least informative distribution. Other outlier-robust weighting rules are, for example, proposed in [12], [15], [42], and [43]. Our experiments showed that for our application, these weighting rules give similar results to Huber's least informative distribution, so they are not discussed in the following.

The least informative distribution can be found by maximizing the asymptotic variance (or minimizing the Fisher information) of the M -estimator over a given set \mathcal{P} of distributions [1]. In our scenario, we choose \mathcal{P} as the ϵ -contaminated normally distributed convex set \mathcal{P}_ϵ

$$\mathcal{P}_\epsilon = \{f(x) | f(x) = (1 - \epsilon)\phi(x) + \epsilon c(x) : c(x) \in \mathcal{C}\} \quad (23)$$

where $\phi(x)$ is the normal distribution with zero mean and unity variance, $c(x)$ is the contaminating distribution belonging to the set \mathcal{C} of distributions that are symmetric to the origin, and $\epsilon \in [0, 1]$ is the outlier probability.

In [1], it is shown that the least informative distribution in \mathcal{P}_ϵ to estimate $\underline{\mathbf{w}}$ and scale $\underline{g}_n(r)$, in the sense that the asymptotic variance is maximized, can be obtained by

$$\rho_H(|x|) = \begin{cases} \frac{|x|^2}{2}, & \text{for } |x| \leq k_0 \\ k_0|x| - \frac{k_0^2}{2}, & \text{for } |x| > k_0 \end{cases} \quad (24)$$

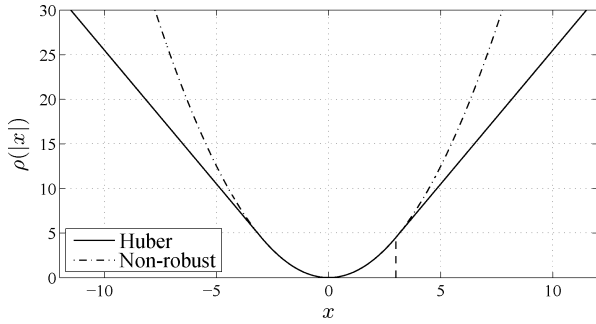


Fig. 1. Huber's outlier-robust weighting function $\rho_H(|x|)$ (Huber) and conventional quadratic cost function $\rho(|x|) = |x|^2$ (nonrobust).

where the constant k_0 depends on ϵ and is chosen such that $\int_{-\infty}^{\infty} f(x)dx = 1$. The subscript "H" in $\rho_H(|x|)$ stands for Huber. It may be seen that the least informative distribution is Gaussian in the center and Laplacian in the tails, leading to an L_2 -norm and L_1 -norm minimization problem, respectively. By correctly choosing the transition point k_0 , data samples that very likely correspond to outliers are thus linearly weighted, while non-outliers are quadratically weighted as for an ordinary LSE criterion. The transition depends on ϵ and decreases with increasing ϵ .

The nonlinear cost along with a conventional quadratic cost is illustrated in Fig. 1.

C. Robust Adaptation Algorithm

One way to minimize the cost function $J(r, \underline{\mathbf{w}})$ is the use of an iterative Newton algorithm [38], as in [8], [15], [36], [37], and [40]. As shown in [37] for a real-valued error signal, the nonquadratic cost $J(r, \underline{\mathbf{w}})$ for a complex-valued error signal can be minimized for $\beta(i, r) = (1 - \lambda)\lambda^{r-i}$ using a recursion of the form

$$\underline{\mathbf{w}}(r) = \underline{\mathbf{w}}(r-1) - \underline{\boldsymbol{\mu}}(r)\underline{\mathbf{A}}^{-1}(r)\nabla\tilde{J}[r, \underline{\mathbf{w}}(r-1)] \quad (25)$$

as a good approximation for computational efficiency, where

$$\nabla\tilde{J}(r, \underline{\mathbf{w}}) = 2\frac{\partial\tilde{J}(r, \underline{\mathbf{w}})}{\partial\underline{\mathbf{w}}^*} \quad (26)$$

is the gradient of the cost function $\tilde{J}(r, \underline{\mathbf{w}})$ w.r.t. $\underline{\mathbf{w}}$ and where $\underline{\boldsymbol{\mu}}(r)$ is a diagonal matrix of size $2NQ \times 2NQ$ with step sizes $\mu_u(r)$, $u = 0, 1, \dots, 2NQ - 1$ on the main diagonal for controlling the adaptation in the frequency bins separately. The matrix $\underline{\mathbf{A}}(r)$ is equal to the recursively averaged Hessian $\nabla\nabla^H\tilde{J}(r, \underline{\mathbf{w}}) = 2\partial/\partial\underline{\mathbf{w}}^*\nabla^H\tilde{J}(r, \underline{\mathbf{w}})$ evaluated for $\underline{\mathbf{w}}(r-1)$

$$\underline{\mathbf{A}}(r) = \lambda\underline{\mathbf{A}}(r-1) + (1 - \lambda)\nabla\nabla^H\tilde{J}[r, \underline{\mathbf{w}}(r-1)]. \quad (27)$$

The use of (25)–(27) leads to the MC-BRFDFAF [40]. To assure convergence, the Hessian matrix $\underline{\mathbf{A}}(r)$ must be positive-definite. While this condition is fulfilled for Huber's cost function (24), it is not necessarily the case for other robust cost functions, as, for example, Hampel's M -estimator [12]. A common solution to this problem is the use of the discrete Newton algorithm [38],

where the Hessian is approximated by a difference matrix, i.e., a matrix with finite differences as elements that approximate the differential. This approximation then yields the discretized MC-BRFDFAF. The experimental results in Section IV show that the approximation of the Hessian by a difference matrix not only solves the problem of the nonpositive-definite Hessian but also improves the robustness of the adaptation algorithm for Huber's M -estimator. For $\epsilon \rightarrow 0$, $k_0 \rightarrow \infty$, and $\underline{\mathbf{s}}_0(r) = \underline{\mathbf{s}}_1(r) = \dots = \underline{\mathbf{s}}_{2N-1}(r) = 1/\sqrt{2}$, we obtain the MC-FDAF after [27]–[29], [31] with RLS-like properties.

In Section II-C1, we determine the gradient $\nabla\tilde{J}(r, \underline{\mathbf{w}})$. Section II-C2 then derives the Hessian $\nabla\nabla^H\tilde{J}(r, \underline{\mathbf{w}})$ and introduces the difference approximation. Section II-C3 summarizes the adaptation algorithms and extends the MISO adaptation algorithms to the MIMO case. In Section II-C4, an approximation is given for reducing the computational complexity.

1) *Gradient of the Cost Function:* To determine the gradient $\nabla\tilde{J}(r, \underline{\mathbf{w}})$ in (26), we write one element $\underline{\mathbf{e}}_n^*(r)$ of the error vector $\underline{\mathbf{e}}^*(r)$ as an inner product

$$\underline{\mathbf{e}}_n^*(r) = \underline{\mathbf{e}}^H(r)\mathbf{1}_{2N \times 1}^{(n)} \quad (28)$$

where $\mathbf{1}_{2N \times 1}^{(n)}$ is a $2N \times 1$ vector of zeroes except for the n th element which is equal to 1. Using the chain rule, we can calculate the gradient $\nabla\tilde{J}(r, \underline{\mathbf{w}})$ as follows:

$$\begin{aligned} \nabla\tilde{J}(r, \underline{\mathbf{w}}) &= \\ &= 2\frac{\partial\tilde{J}(r, \underline{\mathbf{w}})}{\partial\underline{\mathbf{w}}^*} = 2\sum_{n=0}^{2N-1}\frac{\partial}{\partial\underline{\mathbf{w}}^*}\rho\left[\frac{|\underline{\mathbf{e}}_n(r)|}{\underline{\mathbf{s}}_n(r)}\right] \\ &= 2\sum_{n=0}^{2N-1}\rho'\left[\frac{|\underline{\mathbf{e}}_n(r)|}{\underline{\mathbf{s}}_n(r)}\right]\frac{1}{\underline{\mathbf{s}}_n(r)}\cdot\frac{\partial|\underline{\mathbf{e}}_n(r)|}{\partial\underline{\mathbf{e}}_n^*(r)}\cdot\frac{\partial\underline{\mathbf{e}}_n^*(r)}{\partial\underline{\mathbf{w}}^*} \end{aligned} \quad (29)$$

where $\rho'(x)$ is the derivative of $\rho(x)$ and where

$$\frac{\partial|x|}{\partial x^*} = \frac{\partial\sqrt{xx^*}}{\partial x^*} = \frac{1}{2}\sqrt{\frac{x}{x^*}} = \frac{1}{2}\exp\{j\arg(x)\} \quad (30)$$

$$\frac{\partial\underline{\mathbf{e}}_n^*(r)}{\partial\underline{\mathbf{w}}^*} = -\underline{\mathbf{X}}^H(r)(\mathbf{G}_{2N \times 2N}^{01})^H\mathbf{1}_{2N \times 1}^{(n)}. \quad (31)$$

Substituting (30) and (31) into (29) and defining a vector $\underline{\boldsymbol{\psi}}(r)$ of size $2N \times 1$ with elements

$$\underline{\boldsymbol{\psi}}_n(r) = -\frac{1}{2\underline{\mathbf{s}}_n(r)}\rho'\left[\frac{|\underline{\mathbf{e}}_n(r)|}{\underline{\mathbf{s}}_n(r)}\right]\exp\{j\arg[\underline{\mathbf{e}}_n(r)]\} \quad (32)$$

$n = 0, 1, \dots, 2N - 1$, we obtain

$$\nabla\tilde{J}(r, \underline{\mathbf{w}}) = 2\underline{\mathbf{X}}^H(r)(\mathbf{G}_{2N \times 2N}^{01})^H\underline{\boldsymbol{\psi}}(r). \quad (33)$$

2) *Hessian of the Cost Function and Discretization:* The Hessian matrix $\nabla\nabla^H\tilde{J}(r, \underline{\mathbf{w}})$ of size $2N \times 2N$ can be calculated from (33) using

$$\begin{aligned} \nabla\nabla^H\tilde{J}(r, \underline{\mathbf{w}}) &= 4\frac{\partial^2}{\partial^2\underline{\mathbf{w}}^*}\tilde{J}(r, \underline{\mathbf{w}}) = 2\frac{\partial}{\partial\underline{\mathbf{w}}^*}[\nabla J(r, \underline{\mathbf{w}})]^H \\ &= 4\frac{\partial}{\partial\underline{\mathbf{w}}^*}\underline{\boldsymbol{\psi}}^H(r)\mathbf{G}_{2N \times 2N}^{01}\underline{\mathbf{X}}(r). \end{aligned} \quad (34)$$

The derivative $\partial \underline{\psi}^H(r)/\partial \underline{\mathbf{w}}^*$ in (34) can then be calculated by applying the chain rule as follows:

$$\frac{\partial \underline{\psi}^*(r)}{\partial \underline{\mathbf{w}}^*} = \frac{\partial \underline{\psi}^*(r)}{\partial \underline{\epsilon}_n^*(r)} \cdot \frac{\partial \underline{\epsilon}_n^*(r)}{\partial \underline{\mathbf{w}}^*} \quad (35)$$

where $\partial \underline{\epsilon}_n^*(r)/\partial \underline{\mathbf{w}}^*$ is given by (31). For the MC-BRFDAF, we expand the derivative w.r.t. $\underline{\epsilon}_n^*(r)$ [40], that is

$$\begin{aligned} \frac{\partial \underline{\psi}^*(r)}{\partial \underline{\epsilon}_n^*(r)} &= -\frac{1}{4} \left\{ \frac{1}{\underline{s}_n^2(r)} \rho'' \left[\frac{|\underline{\epsilon}_n(r)|}{\underline{s}_n(r)} \right] + \frac{1}{\underline{s}_n(r) |\underline{\epsilon}_n(r)|} \rho' \left[\frac{|\underline{\epsilon}_n(r)|}{\underline{s}_n(r)} \right] \right\} \\ &=: -\underline{\gamma}_n(r). \end{aligned} \quad (36)$$

With the $\underline{\gamma}_n(r)$, $n = 0, 1, \dots, 2N - 1$, captured in a diagonal matrix $\underline{\Gamma}(r)$ of size $2N \times 2N$, the Hessian matrix $\nabla \nabla^H \tilde{J}(r, \underline{\mathbf{w}})$ can be written as

$$\nabla \nabla^H \tilde{J}(r, \underline{\mathbf{w}}) = 4 \underline{\mathbf{X}}^H(r) (\mathbf{G}_{2N \times 2N}^{01})^H \underline{\Gamma}(r) \mathbf{G}_{2N \times 2N}^{01} \underline{\mathbf{X}}(r). \quad (37)$$

To obtain the discretized MC-BRFDAF using a discrete Newton algorithm, we approximate the derivatives $\partial \underline{\psi}^*(r)/\partial \underline{\epsilon}_n^*(r)$ at the origin by first-order difference equations. That is,

$$\left. \frac{\partial \underline{\psi}^*(r)}{\partial \underline{\epsilon}_n^*(r)} \right|_{\underline{\epsilon}_n^*(r)=0} \approx \frac{\underline{\psi}_n^*(r) \Big|_{\underline{\epsilon}_n^*(r)=\underline{\epsilon}_n^*(r)} - \underline{\psi}_n^*(r) \Big|_{\underline{\epsilon}_n^*(r)=0}}{\underline{\epsilon}_n^*(r)} =: -\underline{q}_n(r). \quad (38)$$

Noting that $\underline{\psi}_n^*(r) \Big|_{\underline{\epsilon}_n^*(r)=0} = 0$ and substituting $\underline{\epsilon}_n^*(r) := \underline{e}_n^*(r)$, we obtain the approximation

$$\underline{\gamma}_n(r) \approx \underline{q}_n(r) = -\frac{\underline{\psi}_n^*(r)}{\underline{e}_n^*(r)} = \frac{1}{2 \underline{s}_n(r) |\underline{\epsilon}_n(r)|} \rho' \left[\frac{|\underline{\epsilon}_n(r)|}{\underline{s}_n(r)} \right]. \quad (39)$$

The $\underline{q}_n(r)$, $n = 0, 1, \dots, 2N - 1$, are captured in a diagonal matrix $\underline{\mathbf{Q}}(r)$ of size $2N \times 2N$ so that the discrete approximation of the Hessian matrix $\nabla \nabla^H \tilde{J}(r, \underline{\mathbf{w}})$ is finally obtained as

$$\nabla \nabla^H \tilde{J}(r, \underline{\mathbf{w}}) \approx 4 \underline{\mathbf{X}}^H(r) (\mathbf{G}_{2N \times 2N}^{01})^H \underline{\mathbf{Q}}(r) \mathbf{G}_{2N \times 2N}^{01} \underline{\mathbf{X}}(r). \quad (40)$$

3) *Generalization*: A generalization of the adaptation algorithm to MIMO systems

$$\underline{\mathbf{W}} = [\underline{\mathbf{w}}_0 \ \underline{\mathbf{w}}_1 \ \dots \ \underline{\mathbf{w}}_{P-1}] \quad (41)$$

with Q input channels and P output channels is straightforward by repeating the algorithm for all P output channels. In summary, one iteration of the discrete MC-BRFDAF can be written for the MIMO case as follows:

$$\mathbf{e}_p(rR) = \mathbf{y}_{\text{ref},p}(rR) - \mathbf{W}_{N \times 2N}^{01} \mathbf{F}_{2N \times 2N}^{-1} \underline{\mathbf{X}}(r) \underline{\mathbf{w}}_p(r) \quad (42)$$

$$\begin{aligned} \underline{\mathbf{A}}_p(r) &= \lambda \underline{\mathbf{A}}_p(r-1) + 4(1-\lambda) \\ &\quad \times \underline{\mathbf{X}}^H(r) (\mathbf{G}_{2N \times 2N}^{01})^H \underline{\mathbf{Q}}_p(r) \mathbf{G}_{2N \times 2N}^{01} \underline{\mathbf{X}}(r) \end{aligned} \quad (43)$$

$$\begin{aligned} \underline{\mathbf{w}}_p(r) &= \underline{\mathbf{w}}_p(r-1) + 2(1-\lambda) \\ &\quad \times \underline{\boldsymbol{\mu}}_p(r) \underline{\mathbf{A}}_p^{-1}(r) \underline{\mathbf{X}}^H(r) \mathbf{G}_{2N \times 2N}^{01} \underline{\boldsymbol{\psi}}_p(r). \end{aligned} \quad (44)$$

The MC-BRFDAF is obtained when replacing in (43) $\underline{\mathbf{Q}}_p(r)$ by $\underline{\Gamma}_p(r)$. The adaptation step (44) requires the inversion of the large Hessian matrix $\underline{\mathbf{A}}_p(r)$ of size $2NQ \times 2NQ$. For efficiently

calculating $\underline{\mathbf{A}}_p^{-1}(r)$, the blockwise-diagonal structure of $\underline{\mathbf{A}}_p(r)$ after introducing the approximation in Section II-C4 can be used to transform the $2NQ \times 2NQ$ matrix $\underline{\mathbf{A}}_p(r)$ into $2N$ matrices of size $Q \times Q$. This reduces the complexity of the inversion of a $2NQ \times 2NQ$ to $2N$ matrices of size $Q \times Q$ [29], [31]. In addition to [37, Eq. (30)], the update (44) allows for bin-wise operation with a bin-dependent step size and a bin-dependent scale parameter. To estimate the scale parameter $\underline{s}_n(r)$, we use the outlier-robust M -estimator for scale as presented in [8].

4) *Approximation*: The calculation of the averaged Hessian matrix $\underline{\mathbf{A}}_p(r)$ (43) requires the computation of $4Q$ DFTs. To limit the computational complexity, an approximation of $\underline{\mathbf{A}}_p(r)$ is thus desirable for practical systems. Following [27]–[29], and [31], we use the approximation $\mathbf{G}_{2N \times 2N}^{01} \approx \mathbf{I}_{2N \times 2N}/2$ for sufficiently large N , which leads to

$$\underline{\mathbf{A}}_p'(r) = \lambda \underline{\mathbf{A}}_p'(r-1) + \frac{1-\lambda}{4} \underline{\mathbf{X}}^H(r) \underline{\mathbf{Q}}(r) \underline{\mathbf{X}}(r) \quad (45)$$

and, when replacing $\underline{\mathbf{A}}_p(r)$ by $\underline{\mathbf{A}}_p'(r)$ in (44), to the unconstrained MC-FDAF [27]–[29], [31].

D. Computational Complexity

We compare the computational complexity of the discretized MC-BRFDAF with that of the MC-FDAF in terms of real-valued arithmetic operations per sample (OPS), i.e., the number of real additions and real multiplications. Due to the minor complexity difference between the MC-BRFDAF and the discrete MC-BRFDAF in the computation of (36) and (39), we illustrate only the complexity of the discretized MC-BRFDAF defined by (42) and (44), using the approximation of (45). The matrix inversion in (44) is carried out by applying the matrix inversion lemma to (45) [31]. For the non-robust MC-FDAF, the matrix $(\mathbf{G}_{2N \times 2N}^{01})^H \underline{\mathbf{Q}}_p(r)$ is replaced by the identity matrix $\mathbf{I}_{2N \times 2N}$. Each complex addition is realized by two real additions, and each complex multiplication is realized by two real additions and four real multiplications. The symmetry of the DFT coefficients of real-valued sequences is exploited. The Fourier transform of length $2N$ is carried out by $4N \log_2(2N) - 3N - 4$ real-valued arithmetic operations [44]. Fig. 2 shows the computational complexity OPS of the discretized MC-BRFDAF (solid lines) and the MC-FDAF (dashed lines) as a function of the number of filter taps N for (a) $P = 1$, $Q = 1, 4, 8$ and (b) $Q = 1$, $P = 1, 4, 8$. For MISO systems ($P = 1$), the computational complexity of the discretized MC-BRFDAF relative to that of the MC-FDAF decreases with an increase of the number of input channels Q , while for single-input multiple-output (SIMO) systems ($Q = 1$), the complexity difference increases with an increase in the number of output channels P due to the necessity to calculate the inverse of the CPSD matrix $\underline{\mathbf{A}}_p(r)$ for each of the output channels p .

E. Convergence Behavior in Impulsive Noise

To illustrate the convergence of the outlier-robust adaptation algorithms, we consider a single-input single-output (SISO) system identification task, which appears for example in monophonic acoustic echo cancellation, with white noise input for $N = 2048$, block overlap factor $\alpha = 2$, $k_0 = 0.3$, and the

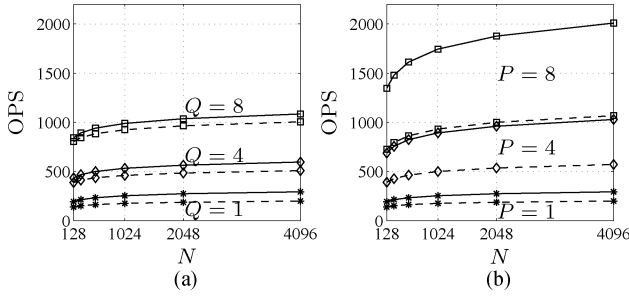


Fig. 2. Comparison of the number of real-valued arithmetic operations (sum of additions and multiplications) per sample (OPS) of the discretized MC-BRFDAF (solid lines) and the nonrobust MC-FDAF (dashed lines) for $\alpha = 1$, (a) as a function of Q for $P = 1$ (MISO system) and (b) as a function of P for $Q = 1$ (SIMO system).

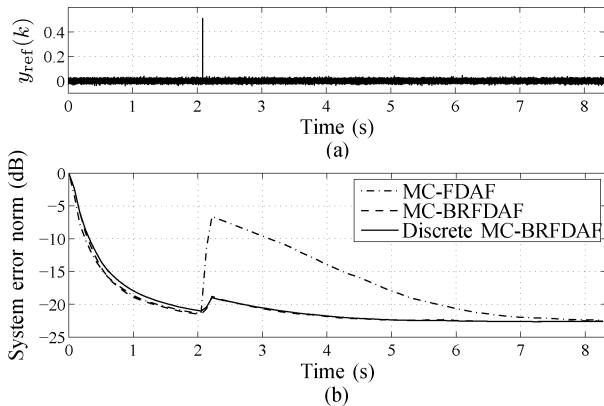


Fig. 3. Comparison between the MC-FDAF, the MC-BRFDAF, and the discrete MC-BRFDAF for a SISO system identification task (acoustic echo cancellation) with white noise input and presence of an impulse of 1-ms duration ($N = 2048$, $\alpha = 2$, sampling rate $f_s = 12$ kHz). (a) Reference signal. (b) System error norm.

sampling rate $f_s = 12$ kHz. The reference signal $y_{\text{ref}}(k)$ is disturbed by an impulse with a duration of 1 ms at time $t = 2.1$ s [Fig. 3(a)]. For the outlier-robust adaptation algorithms, the parameter k_0 is optimized for maximum robustness against the impulsive noise while preserving maximum convergence between $t = 0$ and $t = 2.1$ s. The unknown system \mathbf{w}' is a measured room impulse response of an acoustic environment with reverberation time $T_{60} = 250$ ms truncated after 4096 taps. Fig. 3(b) shows the coefficient error norm $\|\mathbf{w}' - \mathbf{w}(k)\|^2 / \|\mathbf{w}'\|^2$ obtained by the MC-FDAF, the MC-BRFDAF, and the discrete MC-BRFDAF. The convergence for the outlier-robust adaptation algorithms is significantly more robust than for the nonrobust adaptation algorithm.

III. APPLICATION TO A ROBUST GENERALIZED SIDELobe CANCELER (RGSC) WITH DFT-BIN-WISE DOUBLE-TALK DETECTION

To illustrate the robustness, we apply the proposed algorithms to multichannel speech enhancement with a DFT-domain realization of a GSC [34] using an adaptive blocking matrix [39], which is referred to as robust GSC (RGSC) in the following. (“Robust” here means robustness against desired signal cancellation in reverberant conditions and the ability to suppress interference with nonstationary second-order statistics.)

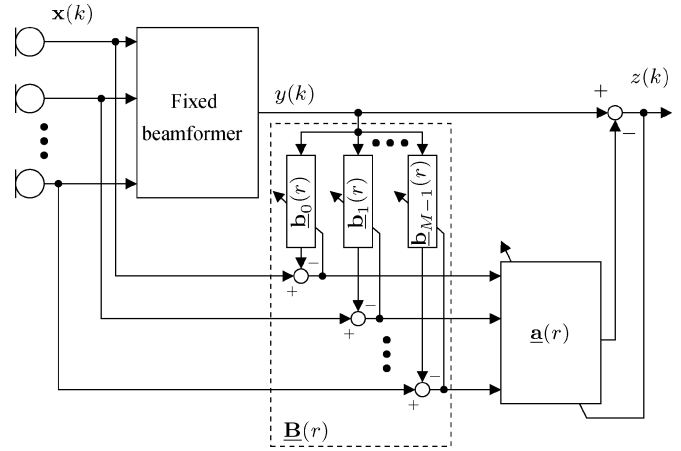


Fig. 4. RGSC with adaptive blocking matrix after [39].

In Section III-A, an overview of the RGSC is given. Section III-B describes the motivation for choosing an impulsive noise robust MC-FDAF for the adaptation.

A. Overview of the RGSC

According to Fig. 4, the RGSC consists of three blocks: a fixed beamformer, an adaptive blocking matrix $\mathbf{B}(r)$, and an interference canceler $\mathbf{a}(r)$. The blocking matrix $\mathbf{B}(r)$ and the interference canceler $\mathbf{a}(r)$ are realized by systematically applying the robust adaptation algorithms.

1) *Fixed Beamformer*: The fixed beamformer forms the reference path of the RGSC. It steers the sensor array to the position of the desired source and enhances the desired signal relative to the interference. For scenarios with little movement of the desired source relative to the steering direction, the fixed beamformer can be designed with a sufficiently broad mainlobe so that the desired signal is not attenuated; otherwise, automatic steering techniques may be necessary. Without loss of generality, we use a uniformly weighted delay-and-sum beamformer in our case for simplicity.

2) *Adaptive Blocking Matrix*: The blocking matrix $\mathbf{B}(r)$ is a spatial SIMO filter which suppresses the desired signal and passes interference such that the output of $\mathbf{B}(r)$ provides references for the interference. The adaptive blocking matrix is realized by adaptive filters with the output signal of the fixed beamformer as reference for the desired signal. In contrast to a fixed blocking matrix \mathbf{B} , which does not perfectly suppress the desired signal whenever there is a mismatch between the spatial filtering of \mathbf{B} and the actual wave field of the desired signal, an adaptive blocking matrix can track changes of the wave field of the desired signal. This is especially important in time-varying reverberant environments, where the desired signal continuously leaks through the fixed blocking matrix and leads to reduced robustness against cancellation of the desired signal.

3) *Adaptive Interference Canceler*: The interference canceler $\mathbf{a}(r)$ is realized by an adaptive MISO filter and subtracts the residual interference from the output of the fixed beamformer using the blocking matrix output as reference for the interference.

4) *DFT-Bin-Wise Adaptation Control*: The RGSC requires a carefully designed adaptation control. The adaptive blocking matrix needs adaptation control for the following reason.

- 1) Due to the limited interference suppression of the fixed beamformer, the output of the fixed beamformer not only contains the desired signal but also the attenuated interference. This estimate of the desired signal is used as input signals of the adaptive filters of the blocking matrix. Continuous adaptation of the filters of the blocking matrix during presence of the desired signal and the interference would thus not only cancel the desired signal but also the interference. However, interference components that are suppressed by the blocking matrix cannot be canceled by the interference canceler, and, thus, leak to the output of the RGSC [34], [39], [45]. To prevent suppression of the interference by the blocking matrix, the blocking matrix should therefore only be adapted when the signal-to-interference ratio (SIR) is high.

The interference canceler requires adaptation control for the following two reasons.

- 2a) Typically, the blocking matrix cannot completely suppress the desired signal so that the blocking matrix output still contains desired signal components. These desired signal components would be confused with the interference and canceled by the interference canceler if the interference is adapted during presence of the interference. To prevent cancellation and distortion of the desired signal, the interference canceler should therefore only be adapted if the SIR is low [39], [45]–[47].
- 2b) Assuming now that the blocking matrix totally suppresses the desired signal, any interference that the fixed beamformer still contains should be suppressed by the interference canceler. (In practice, the suppression of the desired signal by the blocking matrix is between 15 and 25 dB.) However, the desired signal at the fixed beamformer output can act as a noise source which is uncorrelated with the input signal of the adaptive filters. The presence of this uncorrelated noise source disturbs the adaptation and may lead to instability of the adaptive filters just as in a system identification task.

To exploit the sparseness of the desired signal and of interference in the time-frequency domain and, thus, to allow a higher tracking capability, the decision “adaptation of $\underline{\mathbf{B}}(r)$ ” or “adaptation of $\underline{\mathbf{a}}(r)$ ” should be made in separate frequency bins. It is shown in [34] that such a bin-wise decision is even necessary to obtain maximum performance in time-varying acoustic conditions.

An activity detector for “desired signal only” [adaptation of $\underline{\mathbf{B}}(r)$], “interference only” [adaptation of $\underline{\mathbf{a}}(r)$], and “double-talk” (no adaptation) which operates in separate DFT bins is presented in [34]. Using the directivity of a fixed beamformer, the activity detector forms a biased estimate $\Upsilon_n(r)$ of the ratio of the power spectral densities (PSDs) $SIR_n(r) = S_{dd,n}(r)/S_{ii,n}(r)$ of the desired signal, $S_{dd,n}(r)$, and of the interference, $S_{ii,n}(r)$, and tracks the maxima and the minima of $\Upsilon_n(r)$ in the n th bin. The blocking matrix and the

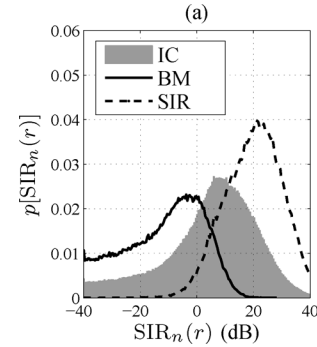


Fig. 5. Histogram of $SIR_n(r)$ averaged over frequency bins n and over block time r : for all samples (SIR), for samples for which the blocking matrix is adapted (BM), and for samples for which the interference canceler is adapted (IC).

interference canceler are adapted whenever $\Upsilon_n(r)$ is close to maximum or minimum, respectively.

B. Motivation for the Usage of Outlier-Robust Adaptive Filtering

A statistical illustration of the adaptation control is depicted in Fig. 5 for male-desired speech and orchestra music played back by a loudspeaker. The experimental setup corresponds to that of Section IV-A. Fig. 5 shows the histograms of $SIR_n(r)$ for which the classifier detects “desired signal only” (dashed line) and “interference only” (black line) along with the histogram of $SIR_n(r)$ at the sensors (gray shaded surface). The histograms are averaged over DFT bins n and block-time r for a signal segment of length 20 s with average $SIR = 3$ dB. Although the centers of the histograms for “desired signal only” and “interference only” are located at high $SIR_n(r) = 20$ dB and at low $SIR_n(r) = -3$ dB, respectively, “desired signal only” and “interference only” may be wrongly detected for low $SIR_n(r)$ and high $SIR_n(r)$, respectively. Another illustration of the problem of outliers of the RGSC is given in [40].

These outliers—potentially leading to divergence of the adaptive filters—may be prevented by reducing the step sizes of the adaptive filters, or by reducing the adaptation thresholds so that the adaptive filters are less likely to be adapted during double-talk. However, both of these options reduce the tracking capability and, thus, the interference suppression of the RGSC. To avoid this tradeoff, we apply the outlier-robust adaptive filters to the blocking matrix and to the interference canceler. From Fig. 4, it may be seen that the blocking matrix corresponds to a single-input multiple-output system with $Q = 1$ and $P = M$ and that the interference canceler corresponds to a MISO system with $Q = M$ and $P = 1$. Identifying the adaptive filters $\underline{\mathbf{b}}_m(r)$ of the blocking matrix in (44) with $\underline{\mathbf{w}}_p(r)$, $p = m = 0, 1, \dots, M - 1$, and identifying the adaptive filter $\underline{\mathbf{a}}(r)$ with $\underline{\mathbf{w}}_p(r)$, $p = 0$, we can systematically use the outlier-robust adaptive filters for the adaptation of the RGSC. The step sizes $\underline{\boldsymbol{\mu}}(r)$ are determined by the adaptation control and switched between 0 and frequency-independent constant values $\mu_b(1 - \lambda)$ and $\mu_a(1 - \lambda)$ for the blocking matrix and the interference canceler, respectively, for disabling and enabling the adaptation.

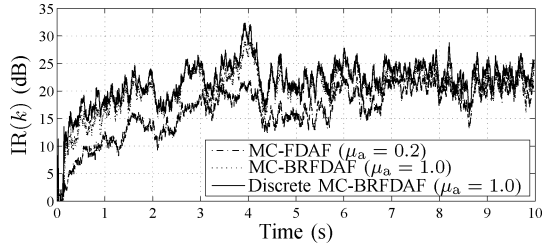


Fig. 6. Comparison of the RGSC interference rejection (IR) using MC-FDAF, MC-BRFDAF, and the discretized MC-BRFDAF in the presence of interference only (orchestra music). ($f_s = 12$ kHz, $N = 256$, $R = 64$, $\lambda = 0.97$, $k_0 = 1.0$. MC-FDAF: $\mu_b = 0.7$. MC-BRFDAF, discretized MC-BRFDAF: $\mu_b = 1.3$.)

IV. EXPERIMENTAL RESULTS

We applied the MC-BRFDAF and the discretized MC-BRFDAF to the adaptation of the blocking matrix and of the interference canceler and compared the performance with the RGSC using MC-FDAFs. The bin-wise scale parameter $\underline{s}_n(r)$ was replaced by a bin-independent scale parameter $s(r)$ since the dependency on DFT bins did not improve the performance relative to a bin-independent scale estimate in our experiments. We studied two extreme cases, the presence of interference only in the absence of desired signal (Section IV-A), and “continuous” double-talk consisting of desired speech and various interference signals (Section IV-B).

A. Comparison Between MC-FDAF, MC-BRFDAF, and Discretized MC-BRFDAF

Here, we study the transient behavior of the interference canceler after initialization using the discretized MC-BRFDAF, the MC-FDAF, and the MC-BRFDAF without the presence of a desired signal, documenting the convergence speed of the adaptive filters in the undisturbed case.

A microphone array with a 12-cm aperture and $M = 4$ elements was placed in an office room with reverberation time $T_{60} = 250$ ms. The desired source signal (male speech) arrived from the broadside direction from a distance of 60 cm. The interference (orchestra music played back by a loudspeaker) arrived from one of the endfire directions from a distance of 1.2 m. The microphone signals were obtained by measuring the room impulse responses and convolving the clean source signals with the corresponding impulse responses. The frequency range was 200 Hz–6 kHz. The parameters are optimized by simulations to simultaneously maximize the interference suppression and minimize the desired signal distortion after convergence of the adaptive filters.

In this experiment, the blocking matrix was first adapted until convergence in the presence of desired signal only. Then, the interference canceler was adapted in the presence of interference only. The transient behavior of the interference canceler is illustrated in Fig. 6, showing the interference rejection $[\text{IR}(k)]$ of the RGSC as a function of time. The interference suppression was measured as a first-order recursively averaged estimate of the variance of the interference at the beamformer output relative to the variance of the interference at the sensors. The parameters, shown in the caption, are optimized individually

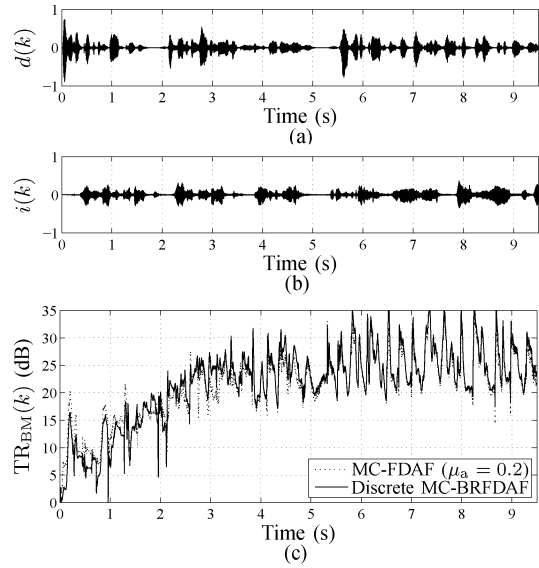


Fig. 7. Comparison of the RGSC target rejection (TR) using MC-FDAF and the discretized MC-BRFDAF for continuous double-talk of speech signals ($f_s = 12$ kHz, $N = 256$, $R = 64$, $\lambda = 0.97$, $k_0 = 1.0$. MC-FDAF: $\mu_b = 0.7$, $\mu_a = 0.2$. Discretized MC-BRFDAF: $\mu_b = 1.3$, $\mu_a = 1.0$). (a) Desired speech. (b) Interfering speech at the $M/2$ th microphone. (c) Suppression of the desired signal $\text{TR}_{\text{BM}}(k)$ by the blocking matrix.

for the discretized MC-BRFDAF, the MC-FDAF, and the MC-BRFDAF for maximum convergence speed and maximum interference suppression after convergence during a double-talk situation. It can be noticed that the step size for the MC-FDAF ($\mu_a = 0.2$) is considerably smaller than those for the robust adaptation algorithms ($\mu_a = 1.0$) due to the sensitivity of the MC-FDAF to double-talk detection errors during adaptation. This smaller step size necessarily leads to a slower convergence of the interference canceler. Due to the improved robustness, the discretized MC-BRFDAF performed slightly better than the MC-BRFDAF. When the same step size μ_a is chosen for all adaptation algorithms, the convergence speed of the MC-FDAF is slightly higher than that of the discretized MC-BRFDAF since the improved robustness is obtained by reducing the gradient of the cost function, which leads to a reduced convergence speed without the presence of outliers.

B. Adaptation During Double-Talk of Desired Signal With Interference Present

In this section, the performance of the RGSC is studied during double-talk of desired speech in the presence of interference. In Section IV-B1, the transient behavior is discussed. Section IV-B2 investigates the steady-state performance after convergence of the adaptive filters.

1) *Transient Behavior:* The transient behavior of the RGSC is now studied using speech as the desired signal and as interference [Fig. 7(a) and (b)]. Experimental results for orchestra music as used in Section IV-A and for car noise can be found in [40] and in [41], respectively. Only the MC-FDAF and the discretized MC-BRFDAF are considered here since the convergence of the MC-BRFDAF is very similar to that of the discretized MC-BRFDAF. The parameters are the same as in Section IV-A.

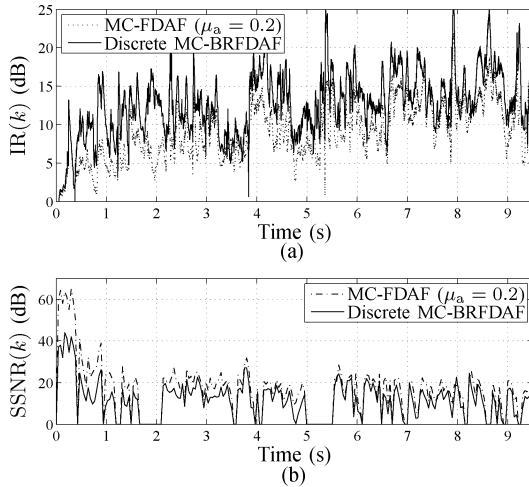


Fig. 8. Comparison of the RGSC using MC-FDAF and the discretized MC-BRFDAF for continuous double-talk of speech signals, parameters as in Fig. 7. (a) Suppression of the interference $IR(k)$ by the RGSC. (b) Distortion of the desired signal, $SSNR(k)$.

The suppression of the desired signal [target signal rejection, $TR_{BM}(k)$] by the blocking matrix as a function of time is depicted in Fig. 7(c). Here, $TR_{BM}(k)$ is measured analogously to $IR(k)$. The convergence of the blocking matrix is only slightly improved for the discretized MC-BRFDAF. The smaller difference of optimum step sizes (MC-FDAF: $\mu_b = 0.7$; discretized MC-BRFDAF: $\mu_b = 1.3$) relative to the interference canceler (MC-FDAF: $\mu_a = 0.2$; discretized MC-BRFDAF: $\mu_a = 1.0$) results from the inherent robustness of the adaptive blocking matrix against signals from outside of the mainlobe of the fixed beamformer [34, pp.123–125].

Fig. 8(a) and (b) shows the interference suppression $IR(k)$ and the distortion $SSNR(k)$ of the desired signal by the RGSC, respectively. Here, $SSNR(k)$ is the segmental SNR between the output of the fixed beamformer and the output of the RGSC for the desired signal only with data blocks of length 512, and ideally $SSNR(k) \rightarrow \infty$. (Note that the interference canceler dereverberates the desired signal to some extent. This effect reduces $SSNR(k)$, but is neglected in this context.) It can be seen in Fig. 8(a) that the interference canceler converges noticeably faster when using the discretized MC-BRFDAF since a larger step size can be chosen due to the improved robustness against double-talk. The improved interference suppression yields a slight increase of the speech distortion $SSNR(k)$ [Fig. 8(b)].

In Fig. 9(a) and (b), the $IR(k)$ and the $SSNR(k)$ using the MC-FDAF and the discretized MC-BRFDAF with the same step sizes $\mu_b = 1.3$ and $\mu_a = 1.0$ are illustrated, respectively. When using the MC-FDAF, instability cannot be efficiently prevented, leading to reduced interference suppression [Fig. 9(a), e.g., $t = 4.8 \dots 5.2$ s] and increased speech distortion [Fig. 9(b), $t > 0.5$ s].

2) *Steady-State Performance*: Finally, the interference suppression and the speech distortion are studied after convergence of the adaptive filters as a function of the SIR at the sensors for speech and car noise as interference. Other experimental setup is the same as in Section IV-A. To study the behavior of the adap-

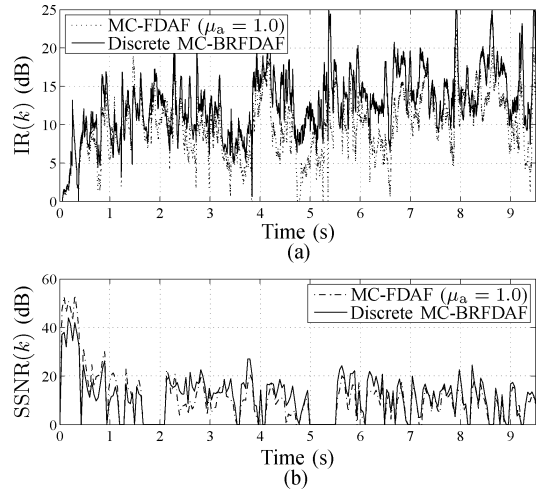


Fig. 9. Comparison of the RGSC using MC-FDAF and the discretized MC-BRFDAF for continuous double-talk of speech signals, parameters as in Fig. 7 except with the same step size $\mu_a = 1.0$ for both MC-FDAF and the discretized MC-BRFDAF. (a) Suppression of the interference $IR(k)$ by the RGSC. (b) Distortion of the desired signal, $SSNR(k)$.

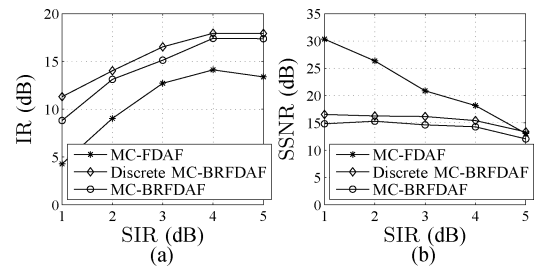


Fig. 10. Comparison of the steady-state performance of the RGSC using the MC-FDAF, the MC-BRFDAF, and the discretized MC-BRFDAF for continuous double-talk with speech as interference (parameters as in Fig. 7). (a) Suppression of the interference IR by the RGSC. (b) Distortion of the desired signal measured by the segmental SNR $SSNR$ between the FBF output and the RGSC output.

tive filters after convergence, we used as desired speech signals a subset of 30 utterances of the TIDigits database [48]. For speech as interference, the reverberation time of the environment was $T_{60} = 250$ ms. For car noise as interference, the desired source signal is convolved with impulse responses with $T_{60} = 50$ ms obtained by the image method [49] to simulate the passenger cabin of a car. The car noise inside of the passenger cabin of a moving car is recorded with a microphone array mounted on one of the sun visors.

The results for speech as interference are depicted in Fig. 10 and in Table I. In Fig. 10(a), it can be seen that the RGSC using outlier-robust adaptive filters yields improved interference suppression relative to the nonrobust adaptive filter as already indicated in the previous experiments. However, interference suppression is increased at the cost of reduced $SSNR$ [Fig. 10(b)]. The discretized MC-BRFDAF yields slightly better performance of the RGSC than the MC-BRFDAF.

The results for car noise as interference are depicted in Fig. 11 and in Table II. In contrast to Fig. 10, the improvement of the interference suppression of the RGSC by using outlier-robust adaptive filters is restricted to high and medium $SIR \geq 10$ dB. The discretized MC-BRFDAF performs slightly better than the MC-BRFDAF.

TABLE I

INTERFERENCE SUPPRESSION (IN dB) OF THE MC-BRFDAF AND OF THE DISCRETIZED MC-BRFDAF FOR CONTINUOUS DOUBLE-TALK WITH SPEECH AS INTERFERENCE. SEGMENTAL SNR $SSNR(k)$ VALUES (IN dB) ARE FOR THE DESIRED SIGNAL BETWEEN THE FBF OUTPUT AND THE RGSC OUTPUT

SIR (dB)	30	20	10	0	-10
Discrete MC-BRFDAF					
IR (dB)	11.31	14.02	16.52	17.94	17.92
SSNR (dB)	16.54	16.26	16.15	15.41	13.34
MC-BRFDAF					
IR (dB)	8.82	13.12	15.12	17.42	17.38
SSNR (dB)	14.82	15.30	14.62	14.27	12.07

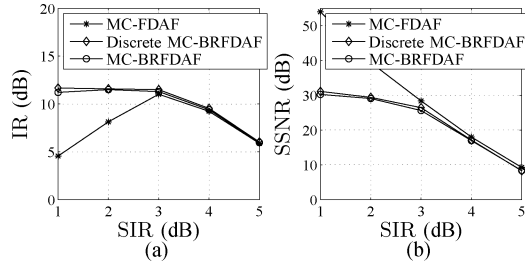


Fig. 11. Comparison of the steady-state performance of the RGSC using the MC-FDAF, the MC-BRFDAF, and the discretized MC-BRFDAF for continuous double-talk with car noise as interference (parameters as in Fig. 7). (a) Suppression of the interference IR by the RGSC. (b) Distortion of the desired signal, SSNR.

TABLE II

COMPARISON OF THE INTERFERENCE SUPPRESSION (IN dB) OF THE MC-BRFDAF WITH THE DISCRETIZED MC-BRFDAF FOR CONTINUOUS DOUBLE-TALK WITH CAR NOISE AS INTERFERENCE. SEGMENTAL SNR $SSNR$ VALUES (IN dB) ARE FOR THE DESIRED SIGNAL BETWEEN THE FBF OUTPUT AND THE RGSC OUTPUT

SIR (dB)	30	20	10	0	-10
Discrete MC-BRFDAF					
IR (dB)	11.67	11.58	11.50	9.54	6.03
SSNR (dB)	31.13	29.42	26.45	17.13	8.30
MC-BRFDAF					
IR (dB)	11.22	11.50	11.30	9.40	5.97
SSNR (dB)	30.26	29.08	25.63	17.00	8.36

The experimental results showed that the improvement in interference suppression using outlier-robust adaptive filters came along with a decrease in the quality of the desired speech signal. To evaluate this effect, we applied the RGSC as a front-end for a large-vocabulary automatic speech recognition (ASR) system which was trained on clean speech [50]. The results showed that the gain in word accuracy of the MC-BRFDAF relative to the MC-FDAF corresponds to the improvement of the SIR at the ASR input due to outlier-robust adaptive filtering. However, an influence of the reduction of the SSNR on the speech recognition performance could not be observed. This effect may be explained as follows. First, the improvement of the SIR has more influence on the speech recognition performance than the decrease of the SSNR w.r.t. the desired signal alone, especially since the SSNR is greater than 13 dB for the investigated range of the SIR. Second, informal listening tests showed that the degradation of the SSNR at the output of the RGSC could not be perceived in the mixture of speech and interference. Further details can be found in [50].

V. CONCLUSION

We have derived outlier-robust multichannel frequency-domain adaptive filters for bin-wise adaptation controls based on a Newton algorithm and on a discrete Newton algorithm. Huber's least informative distribution was considered as a nonlinear weighting rule for the error signal of the adaptation algorithm. For MISO systems, the computational complexity of both robust adaptive filters is slightly increased relative to the MC-FDAF. For SIMO systems, the complexity relative to the MC-FDAF increases with the number of output channels. The robust adaptive filters were then applied to the RGSC. After studying the necessity of outlier-robustness for this type of adaptive beamformer, experimental studies showed that the convergence speed and the interference suppression of the RGSC could be considerably improved by using the discretized MC-BRFDAF relative to the MC-FDAF. Furthermore, a slight improvement could be observed for the discrete MC-BRFDAF relative to the MC-BRFDAF algorithm. The application of the RGSC to a hands-free large-vocabulary ASR showed that at low SIRs, the gain in word accuracy of the ASR system due to outlier-robust adaptive filtering at low SIRs corresponds to the SIR gain at the ASR input.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- [1] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [2] M. M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, vol. XLVI, no. 3, pp. 497–510, Mar. 1967.
- [3] P. M. Clarkson and T. I. Haweel, "Median LMS algorithm," *Electron. Lett.*, vol. 25, no. 8, pp. 520–522, Apr. 1989.
- [4] J. A. Chambers, O. Tanrikulu, and A. G. Constantinides, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, no. 19, pp. 1574–1575, Sep. 1994.
- [5] D. P. Mandic, E. V. Papoulis, and C. G. Bourkis, "A normalized mixed-norm adaptive filtering algorithm robust under impulsive noise interference," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 1, pp. 333–336.
- [6] J. A. Chambers and A. Avlonitis, "A robust mixed-norm (RMN) adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 1, pp. 46–48, Feb. 1997.
- [7] T. Gänslér, "A double-talk resistant subband echo canceller," *Signal Process.*, vol. 65, no. 1, pp. 89–101, Feb. 1998.
- [8] T. Gänslér, 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.
- [9] J. F. Weng and S. H. Leung, "Adaptive nonlinear RLS algorithm for robust filtering in impulsive noise," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1997, vol. 4, pp. 2337–2340.
- [10] S.-C. Chan and Y.-X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.
- [11] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.
- [12] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.
- [13] Y. Zou, S. C. Chan, and T. S. Ng, "A robust M-estimate adaptive filter for impulsive noise suppression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1999, vol. 4, pp. 1765–1768.
- [14] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulsive noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.

- [15] J. Benesty and T. Gänslér, "A robust fast converging least-squares adaptive algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. 6, pp. 3785–3788.
- [16] T. I. Haweel and P. M. Clarkson, "A class of order statistic LMS algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 1, pp. 44–53, Jan. 1992.
- [17] R. Settineri, M. Najim, and D. Ottaviani, "Order statistic fast Kalman filter," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1996, vol. 2, pp. 116–119.
- [18] B. Sayadi and S. Marcos, "A robustification method of the adaptive filtering algorithms in impulsive noise environments based on the likelihood ratio test," in *Proc. IEEE Int. Symp. Control, Commun., Signal Process.*, 2004, pp. 685–688.
- [19] M. Dentino, J. McCool, and B. Widrow, "Adaptive filtering in the frequency domain," *Proc. IEEE*, vol. 66, no. 12, pp. 1658–1659, Dec. 1978.
- [20] E. R. Ferrara, "Fast implementation of LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 4, pp. 474–475, Aug. 1980.
- [21] D. Mansour and A. H. Gray, "Unconstrained frequency-domain adaptive filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 5, pp. 726–734, Oct. 1982.
- [22] 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.
- [23] J. Benesty and P. Duhamel, "A fast exact least mean square adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 2904–2920, Dec. 1992.
- [24] J. Prado and E. Moulines, "Frequency-domain adaptive filtering with applications to acoustic echo cancellation," *Ann. Telecommun.*, vol. 49, no. 7, pp. 414–428, Jul. 1994.
- [25] E. Moulines, O. A. Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: Structure and convergence analysis," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 14–28, Jan. 1995.
- [26] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Process.*, vol. 80, no. 9, pp. 1733–1745, Sep. 2000.
- [27] J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2000, vol. 2, pp. 789–797.
- [28] J. Benesty and D. R. Morgan, "Multi-channel frequency-domain adaptive filtering," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Boston, MA: Kluwer, 2000, ch. 7, pp. 121–133.
- [29] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive filtering with application to multichannel acoustic echo cancellation," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer, 2003, ch. 4.
- [30] G. Enzner and P. Vary, "Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2003, pp. 43–46.
- [31] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: Efficient realization and application to hands-free speech communication," *Signal Process.*, vol. 85, no. 3, pp. 549–570, Mar. 2005.
- [32] A. W. Khong and P. A. Naylor, "Frequency domain adaptive algorithms for stereophonic acoustic echo cancellation employing tap selection," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2005, pp. 141–144.
- [33] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [34] W. Herboldt, *Sound Capture for Human/Machine Interfaces: Practical Aspects of Microphone Array Signal Processing*. Heidelberg, Germany: Springer, 2005.
- [35] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Heidelberg, Germany: Springer, 2005, ch. 9, pp. 199–228.
- [36] H. Buchner, J. Benesty, T. Gänslér, and W. Kellermann, "An outlier robust extended multidelay filter with application to acoustic echo cancellation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2003, pp. 19–22.
- [37] H. Buchner, J. Benesty, T. Hänslér, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 5, pp. 1633–1644, Sep. 2006.
- [38] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [39] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
- [40] W. Herboldt, H. Buchner, S. Nakamura, and W. Kellermann, "Application of a double-talk resilient DFT-domain adaptive filter for bin-wise stepsize controls to adaptive beamforming," in *Proc. IEEE-EURASIP Workshop Nonlinear Signal Image Process.*, May 2005, pp. 175–181.
- [41] W. Herboldt, H. Buchner, S. Nakamura, and W. Kellermann, "Outlier-robust DFT-domain adaptive filtering for bin-wise stepsize controls, and its application to a generalized sidelobe canceller," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2005, pp. 113–116.
- [42] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, *Robust Estimates of Location: Survey and Advances*. Princeton, NJ: Princeton Univ. Press, 1972.
- [43] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, no. 2, pp. 147–185, 1974.
- [44] H. V. Sorensen, D. L. Jones, M. T. Heidemann, and C. S. Burrus, "Real-valued fast Fourier transform algorithms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 6, pp. 849–863, Jun. 1987.
- [45] D. V. Comperolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1990, vol. 2, pp. 833–836.
- [46] S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 514–518, Nov. 1993.
- [47] J. V. Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *J. Acoust. Soc. Amer.*, vol. 103, no. 6, pp. 3621–3626, Jun. 1998.
- [48] R. G. Leonard, "A database for speaker independent digit recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1984, vol. 3, pp. 42.11.1–42.11.4.
- [49] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 954–950, Apr. 1979.
- [50] W. Herboldt, T. Horiuchi, M. Fujimoto, T. Jitsuhiro, and S. Nakamura, "Hands-free speech recognition and communication on PDAs using microphone array technology," in *Proc. IEEE Workshop Automatic Speech Recognition Understanding*, 2005, pp. 302–307.



Wolfgang Herboldt received the Dipl.-Ing. degree and the Dr.-Ing. degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1999 and 2004, respectively.

From 1999 to 2004, he was a Member of Research Staff at the chair of Multimedia Communications and Signal Processing, University Erlangen-Nuremberg. From 2000 to 2005, he was also a Visiting Researcher at Intel Corporation, Hillsboro, OR, and at Intel China Research Center, Beijing, China.

From 2004 to 2006, he was a Researcher at ATR Spoken Language Communication Research Laboratories, Kyoto, Japan. Since 2006, he has been with the Center of Competence Signal Processing, Rohde and Schwarz, Munich, Germany. His current research interests include robust speech recognition, array signal processing, multichannel adaptive filtering, and their applications to acoustic human/machine interfaces.

Dr. Herboldt received the Best Student Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, in 2000, and the Best Doctoral Thesis Award from the Technical Faculty of the University of Erlangen-Nuremberg in 2004.



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

He is a Research Scientist at the Deutsche Telekom Laboratories, Berlin, Germany. From 2000 to 2006, he was a Member of Research Staff at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Germany. 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), R&D Division, 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. 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, source localization, and dereverberation. He has authored or coauthored over 50 book chapters, journal articles, and conference papers in his field.

Dr. Buchner received the VDI Award in 1998 for his Dipl.-Ing. (FH) thesis from the Verein Deutscher Ingenieure, and he is author or coauthor of three best conference paper awards.



Satoshi Nakamura (M'89) received the B.S. degree in electronic engineering from Kyoto Institute of Technology, Kyoto, Japan, in 1981 and the Ph.D. degree in information science from Kyoto University in 1992.

In 1981–1993, he worked with the Central Research Laboratory, Sharp Corporation, Nara, Japan. From 1986 to 1989, he worked with ATR Interpreting Telephony Research Laboratories. From 1994 to 2000, he worked as an Associate Professor of the Graduate School of Information Science at the

Nara Institute of Science and Technology. In 1996, he was a Visiting Research Professor of the CAIP Center, Rutgers University, New Brunswick, NJ. In 2000, he moved to the ATR Spoken Language Translation Research Laboratories to serve as the Department Head for acoustics and speech research. He is currently the director at ATR Spoken Language Communication Laboratories, Kyoto, Japan. Since 2006, he has also been the Group Leader of the Spoken Language Communication Group, National Institute of Information and Communications Technology, Japan. He also serves as an Honorary Professor at the University of Karlsruhe, Karlsruhe, Germany, since 2004. His current research interests include speech recognition, speech translation, spoken dialogue systems, stochastic modeling of speech, and microphone arrays.

Dr. Nakamura received the Awaya Award from the Acoustical Society of Japan in 1992 and the Interaction2001 best paper Award from the Information Processing Society of Japan in 2001. He is a member of IEICE, ASJ, and PSJ. He served as an Editor of IEICE Transactions on Information and Systems from 2000 to 2002.



Walter Kellermann (SM'06) 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, where he also became director of the Institute of Applied Research in 1997. In 1999, he cofounded 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 authored or coauthored seven book chapters and more than 100 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 served as a Guest Editor to 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 to the *EURASIP Journal on Applied Signal Processing* and the *EURASIP Journal on Signal Processing*. He was the General Chair of the Fifth International Workshop on Microphone Arrays in 2003 and the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2005. He was appointed to serve as a Distinguished Lecturer of the IEEE Signal Processing Society from 2007 to 2008.