Application of a Double-Talk Resilient DFT-Domain Adaptive Filter for Bin-wise Stepsize Controls to Adaptive Beamforming

W. Herbordt*, H. Buchner†, S. Nakamura*, and W. Kellermann†
*ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai Seikacho Soraku-gun, Kyoto 619-0288, Japan
Tel: +81-774-95-1354, Fax: +81-774-95-1308
E-mail: {wolfgang.herbordt, satoshi.nakamura}@atr.jp
†Telecommunications Laboratory, University Erlangen-Nuremberg
Cauerstraße 7, 91058 Erlangen, Germany
Tel: +49-9131-852-7669, Fax: +49-9131-852-8849
E-mail: {buchner,wk}@LNT.de

***Abstract*— In adaptive filtering, undetected noise bursts often disturb the adaptation and may lead to instabilities and divergence of the adaptive filter. The sensitivity against noise bursts increases with the convergence speed of the adaptive filter and limits the performance of signal processing methods where fast convergence is required. Typical applications which are sensitive against noise bursts are adaptive beamforming for audio signal acquisition or acoustic echo cancellation, where noise bursts are frequent due to undetected double-talk. In this paper, we apply double-talk resistant adaptive filtering [2] using a non-linear optimization criterion to adaptive beamforming in the discrete Fourier transform domain for bin-wise adaptation controls. We show the efficiency of double-talk resilient adaptive filtering for a generalized sidelobe canceller for speech and audio signal acquisition. The improved robustness leads to faster convergence, to higher noise-reduction, and to a better output signal quality in turn.**

I. Introduction

For applications as, e.g., acoustic echo cancellation or adaptive beamforming, convergence speed, tracking capability, computational complexity, and low delay are crucial factors for the choice of the adaptive algorithm. For acoustic echo cancellation, in [1], the need for robustness against double-talk bursts due to presence of local speakers (‘double-talk’) was pointed out and addressed by using a non-linear function of the error signal for the adaptation. In [2], robustness against double-talk bursts was obtained for a subband echo canceller by introducing robust statistics [3] into subband adaptive filtering by using a contaminated Gaussian model for the residual echo signal. In [4], the concept of robust statistics was used to derive double-talk robust versions of the normalized least mean-squares (NLMS) algorithm, of the proportionate NLMS (PNLMS) algorithm, and of the affine projection algorithm (APA). In [5], a robust recursive least-squares (RLS) algorithm is derived.

In recent years, discrete Fourier transform (DFT) domain adaptive algorithms (‘frequency-domain adaptive filters’ (FDAFs)) [6] became very attractive for acoustic echo cancellation since they (a) combine fast convergence with low computational complexity and (b) can be realized such that –for many applications– sufficiently high tracking capability and sufficiently low delay are obtained [7]. The FDAF generalizes well to the multi-channel (MC) case (MC-FDAF) [8], [9]. As for RLS algorithms, the convergence speed is independent of the condition number of the cross-correlation matrix of the input signals. This is especially important for highly auto- and cross-correlated input signals (as, e.g., for speech or music) in order to assure fast convergence. Additionally, DFT-domain realizations of acoustic echo cancellers allow for a DFT-bin-wise adaptation. This is especially advantageous for signals which are sparse in the time-frequency domain, since the stepsize of the adaptive algorithm can be adjusted for each DFT-bin individually. This leads to a more frequent adaptation and faster convergence of the adaptive filter [10].

For improving the robustness of this class of algorithms, a robust DFT-domain adaptive filter based on robust statistics and a non-linear least-squares error (LSE) criterion is derived and applied to acoustic echo cancellation in [11]. In contrast to the subband robust adaptive filter [2], where the non-linear cost function is applied to each subband (‘narrowband decomposition’), and the error signal in each subband is minimized individually, [11] minimizes the fullband error signal in the discrete time domain. However, due to the time-domain optimization criterion, [11] cannot be used in combination with a DFT-bin-wise stepsize control.

In this paper, we derive a double-talk resilient DFT-domain adaptive filter which uses a LSE cost function in the DFT domain so that DFT bin-wise stepsize controls can be used for double-talk robust algorithms. In order to apply this technique to multiple-input multiple-output (MIMO) systems, we formulate the algorithm for the multi-channel case and

This research was supported by a program with the National Institute of Information and Communications Technology (NICT), Japan, entitled “JAPAN TRUST International Research Coorporation Program”.

refer to it as multi-channel bin-wise robust FDAF (MC-BRFDAF) (Sect. II). Only the unconstrained case is considered for simplicity. The derivation is similar to [8], [9], [11].

The MC-BRFDAF is then applied to adaptive beamforming for multi-channel speech enhancement using microphone arrays (Sect. III). In [12], it is shown with the example of a 'generalized sidelobe canceller' (GSC) [13] using an adaptive blocking matrix [14] that DFT-domain adaptive filtering can be efficiently applied to adaptive beamforming and that especially sparseness of the desired speech signal helps to solve the tracking problems of GSCs using adaptive realizations of blocking matrices. These DFT-domain GSCs are especially efficient to tackle the challenges of beamforming microphone arrays, such as robustness against reverberation and physical tolerances of the sensors, or time-variance of the desired signal and of interferers. In this paper, experiments with the GSC using the MC-BRFDAF show that robustness against double-talk can be greatly improved relative to the GSC using the MC-FDAF even with small-scale microphone arrays so that larger stepsizes can be chosen for the adaptation. This leads to faster convergence and to higher noise-reduction while preserving good output signal quality of the beamformer.

II. Double-talk resilient frequency-domain adaptive filter

In this section, we formulate the MC-BRFDAF for linear multiple-input single-output (MISO) filters. The generalization to the MIMO case is summarized at the end of this section. The derivation is analogously to [8], [9], [11].

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

A. Computation of the output signal using overlap-save

The output signal $e(k)$ of the adaptive MISO system with Q input channels is given by

$$e(k) = y_{\mathrm{ref}}(k) - \mathbf{x}^T(k)\mathbf{w}(k)\,, \tag{1}$$

where $y_{\mathrm{ref}}(k)$ is the reference signal. The MISO filter is described by the $QN \times 1$ vector $\mathbf{w}(k)$, which captures Q column vectors $\mathbf{w}_q(k)$ of length N with filter coefficients $w_{n,q}(k)$, $n = 0, 1, \ldots, N-1$:

$$\mathbf{w}(k) = \left(\mathbf{w}_0^T(k), \mathbf{w}_1^T(k), \ldots, \mathbf{w}_{Q-1}^T(k)\right)^T, \tag{2}$$
$$\mathbf{w}_q(k) = \left(w_{0,q}(k), w_{1,q}(k), \ldots, w_{N-1,q}(k)\right)^T. \tag{3}$$

The input signals of the adaptive filter are captured in the $QN \times 1$ vector $\mathbf{x}(k)$:

$$\mathbf{x}(k) = \left(\mathbf{x}_0^T(k), \mathbf{x}_1^T(k), \ldots, \mathbf{x}_{Q-1}^T(k)\right)^T, \tag{4}$$
$$\mathbf{x}_q(k) = \left(x_q(k), x_q(k-1), \ldots, x_q(k-N+1)\right)^T. \tag{5}$$

To calculate the output signal of the MISO system in the DFT domain using fast convolution and overlap-save, we form a block of N samples of the error signal $e(k)$ as

$$\mathbf{e}(k) = \mathbf{y}_{\mathrm{ref}}(k) - \mathbf{X}^T(k)\mathbf{w}(k)\,, \tag{6}$$

where

$$\mathbf{e}(k) = \left(e(k), e(k-1), \ldots, e(k-N+1)\right)^T, \tag{7}$$
$$\mathbf{y}_{\mathrm{ref}}(k) = \left(y_{\mathrm{ref}}(k), y_{\mathrm{ref}}(k-1), \ldots, y_{\mathrm{ref}}(k-N+1)\right)^T, \tag{8}$$
$$\mathbf{X}(k) = \left(\mathbf{x}(k), \mathbf{x}(k-1), \ldots, \mathbf{x}(k-N)\right)^T. \tag{9}$$

We define a block overlap factor $\alpha = N/R$, where R is the 'new' number of samples per block, and we replace in (6) the discrete time k by the block time r, which is related to k by $rR = k$. Then, the data matrix $\mathbf{X}(rR)$ is transformed into a block-diagonal matrix $\underline{\mathbf{X}}(r)$ of size $2NQ \times 2N$ in the DFT domain using the DFT matrix $\mathbf{F}_{2N\times 2N}$ of size $2N \times 2N$,

$$\underline{\mathbf{X}}(r) = \left(\underline{\mathbf{X}}_0(r), \underline{\mathbf{X}}_1(r), \ldots, \underline{\mathbf{X}}_{Q-1}(r)\right), \tag{10}$$
$$\underline{\mathbf{X}}_q(r) = \mathrm{diag}\left\{\mathbf{F}_{2N\times 2N}\begin{pmatrix} x_q(rR-N) \\ x_q(rR-N+1) \\ \vdots \\ x_q(rR+N-1)\end{pmatrix}\right\}, \tag{11}$$

and $\mathbf{w}(rR)$ is written in the DFT domain as

$$\underline{\mathbf{w}}(r) = \mathrm{diag}\left\{\left(\mathbf{F}_{2N\times 2N}\mathbf{W}^{10}_{2N\times N}, \ldots, \mathbf{F}_{2N\times 2N}\mathbf{W}^{10}_{2N\times N}\right)\right\}\mathbf{w}(rR)\,, \tag{12}$$

where the windowing matrix

$$\mathbf{W}^{10}_{2N\times N} = \left(\mathbf{I}_{N\times N}, \mathbf{0}_{N\times N}\right)^T \tag{13}$$

appends N zeroes to the coefficient vectors $\mathbf{w}_q(rR)$ in order to prevent circular convolution. $\mathbf{I}_{N\times N}$ is an identity matrix of size $N \times N$, $\mathbf{0}_{N\times N}$ is a matrix of size $N \times N$ with zeroes. We obtain for (6) the expression

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

where the windowing matrix

$$\mathbf{W}^{01}_{N\times 2N} = \left(\mathbf{0}_{N\times N}, \mathbf{I}_{N\times N}\right) \tag{15}$$

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

B. Optimization criterion

For formulating the optimization criterion, we transform the block-error vector $\mathbf{e}(rR)$ into the DFT domain by multiplying (14) with $\mathbf{F}_{2N\times 2N}\mathbf{W}^{01}_{2N\times N}$ from the left, where $\mathbf{W}^{01}_{2N\times N} = (\mathbf{W}^{01}_{N\times 2N})^T$ [8]. We obtain

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

where

$$\underline{\mathbf{e}}(r) = \mathbf{F}_{2N\times 2N}\mathbf{W}^{01}_{2N\times N}\mathbf{e}(rR)\,, \tag{17}$$
$$\underline{\mathbf{y}}_{\mathrm{ref}}(r) = \mathbf{F}_{2N\times 2N}\mathbf{W}^{01}_{2N\times N}\mathbf{y}_{\mathrm{ref}}(rR)\,, \tag{18}$$
$$\mathbf{G}^{01}_{2N\times 2N} = \mathbf{F}_{2N\times 2N}\mathbf{W}^{01}_{2N\times N}\mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N\times 2N}\,. \tag{19}$$

The elements of $\underline{\mathbf{e}}(r)$ are denoted by $\underline{e}_n(r)$, $n = 0, 1, \ldots, 2N-1$.

We apply the Parseval theorem to the criteria after [2], [4], [11], and we define the cost function $\xi(r)$ in the DFT domain as

$$\xi(r) = \sum_{n=0}^{2N-1} \rho\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right), \tag{20}$$

where

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

The parameter k_0 is a constant. Since the scale of $|\underline{e}_n(r)|$ is generally unknown, the variable $s_n(r)$ is introduced into (20) in order to make $\rho(\cdot)$ scale invariant [3]. It should reflect the residual noise level at the system output [2], [4]. It may be seen that (20) corresponds to a LSE criterion [8] in the DFT domain with a quadratic error surface for $|\underline{e}_n(r)|/s_n(r) \le k_0$, while, for $|\underline{e}_n(r)|/s_n(r) > k_0$, the quadratic criterion is replaced by a 1-norm criterion. This choice of $\rho(\cdot)$ makes the estimator resilient against outliers, since the gradient of $\rho\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right)$ is reduced for $|\underline{e}_n(r)|/s_n(r) > k_0$ relative to a quadratic cost function. The choice of k_0 is a trade-off between convergence speed and robustness since the robustness of the algorithm increases with k_0 at the cost of decreasing convergence speed. The MC-FDAF [8] is obtained for $\xi(r) = \sum_{n=0}^{2N-1} |\underline{e}_n(r)|^2$, or, equivalently, for $k_0 \to \infty$, $s_0(r) = s_1(r) = \cdots = s_{2N-1}(r) = 1/\sqrt{2}$.

C. Adaptation algorithm

The cost function (20) is minimized w.r.t. $\underline{\mathbf{w}}(r)$ using an iterative Newton algorithm [15] of the form

$$\underline{\mathbf{w}}(r) = \underline{\mathbf{w}}(r-1) - \underline{\boldsymbol{\mu}}(r)\underline{\mathbf{\Lambda}}^{-1}(r)\nabla\xi(r)\,, \tag{22}$$

where $\nabla\xi(r) = 2\partial\xi(r)/\partial\underline{\mathbf{w}}^*(r)$ is the gradient of the cost function $\xi(r)$ w.r.t. $\underline{\mathbf{w}}(r)$ and where $\underline{\mathbf{\Lambda}}(r) = \mathcal{E}\{\nabla^2\xi(r)\} = 4\mathcal{E}\{\partial^2\xi(r)/\partial^2\underline{\mathbf{w}}^*(r)\}$ is the expected value of the Hessian of $\xi(r)$ w.r.t. $\underline{\mathbf{w}}(r)$. $\underline{\boldsymbol{\mu}}(r)$ is a diagonal matrix of size $2N \times 2N$ with stepsizes $\mu_n(r)$, $n = 0, 1, \ldots, 2N-1$, on the main diagonal for controling the adaptation in the frequency bins separately. The DFT-domain Newton step (22) is analogously to the Newton step in the discrete time domain in [5] and an extension of the DFT-domain Newton step in [11] to a bin-wise operation.

1) Gradient of the cost function: Following [11], we write one element $\underline{e}_n^*(r)$ of the error vector $\underline{\mathbf{e}}^*(r)$ as

$$\underline{e}_n^*(r) = \underline{\mathbf{e}}^H(r)\mathbf{1}_{2N\times 1}^{(n)}\,, \tag{23}$$

where $\mathbf{1}_{2N\times 1}^{(n)}$ is a $2N \times 1$ vector with zeroes and with the n-th element equal to one. Using the chain rule, the gradient $\nabla\xi(r)$ is found as follows:

$$\begin{aligned}\nabla\xi(r) &= 2\frac{\partial\xi(r)}{\partial\underline{\mathbf{w}}^*(r)} = 2\sum_{n=0}^{2N-1}\frac{\partial}{\partial\underline{\mathbf{w}}^*(r)}\rho\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right) \\ &= 2\sum_{n=0}^{2N-1}\rho'\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right)\frac{1}{s_n(r)}\frac{\partial|\underline{e}_n(r)|}{\partial\underline{e}_n^*(r)}\frac{\partial\underline{e}_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)}\end{aligned} \tag{24}$$

where

$$\begin{aligned}\rho'(|z|) &= \min\{|z|, k_0\}\,, & (25)\\ \frac{\partial|z|}{\partial z^*} &= \frac{\partial\sqrt{zz^*}}{\partial z^*} = \frac{1}{2}\sqrt{\frac{z}{z^*}} = \frac{1}{2}\exp\{j\arg\{z\}\}, & (26)\\ \frac{\partial\underline{e}_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)} &= -\underline{\mathbf{X}}^H(r)\left(\mathbf{G}_{2N\times 2N}^{01}\right)^H\mathbf{1}_{2N\times 1}^{(n)}\,. & (27)\end{aligned}$$

Equation 24 may be written with the column vector

$$\underline{\boldsymbol{\psi}}(r) = \begin{pmatrix} -\frac{1}{2s_0}\rho'\left(\frac{|\underline{e}_0(r)|}{s_0(r)}\right)\exp\{j\arg\{\underline{e}_0(r)\}\} \\ -\frac{1}{2s_1}\rho'\left(\frac{|\underline{e}_1(r)|}{s_1}\right)\exp\{j\arg\{\underline{e}_1(r)\}\} \\ \vdots \\ -\frac{1}{2s_{2N-1}}\rho'\left(\frac{|\underline{e}_{2N-1}(r)|}{s_{2N-1}}\right)\exp\{j\arg\{\underline{e}_{2N-1}(r)\}\}\end{pmatrix} \tag{28}$$

of length $2N$ as

$$\nabla\xi(r) = 2\underline{\mathbf{X}}^H(r)\left(\mathbf{G}_{2N\times 2N}^{01}\right)^H\underline{\boldsymbol{\psi}}(r)\,. \tag{29}$$

2) Hessian of the cost function: The Hessian matrix $\nabla^2\xi(r)$ of size $2N \times 2N$ can be calculated from (29) using

$$\begin{aligned}\nabla^2\xi(r) &= 4\frac{\partial^2}{\partial^2\underline{\mathbf{w}}^*(r)}\xi(r) = 2\frac{\partial}{\partial\underline{\mathbf{w}}^*(r)}(\nabla\xi(r))^H \\ &= 4\frac{\partial}{\partial\underline{\mathbf{w}}^*(r)}\underline{\boldsymbol{\psi}}^H(r)\mathbf{G}_{2N\times 2N}^{01}\underline{\mathbf{X}}(r)\,.\end{aligned} \tag{30}$$

Denoting in (28) the n-th element of $\underline{\boldsymbol{\psi}}(r)$ by $\psi_n(r)$, we can calculate the n-th element of $\partial\underline{\boldsymbol{\psi}}^H(r)/\partial\underline{\mathbf{w}}^*(r)$ in (30) by applying the chain rule as follows:

$$\begin{aligned}&\frac{\partial\psi_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)} = \\ &-\frac{1}{2s_n^2(r)}\left(\rho''\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right)\frac{\partial|\underline{e}_n(r)|}{\partial\underline{e}_n^*(r)}\exp\{-j\arg\{\underline{e}_n(r)\}\}\right. \\ &\left.+\,\rho'\left(\frac{|\underline{e}_n(r)|}{s_n(r)}\right)\frac{\partial\exp\{-j\arg\{\underline{e}_n(r)\}\}}{\partial\underline{e}_n^*(r)}\right)\frac{\partial\underline{e}_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)}\,,\end{aligned} \tag{31}$$

where

$$\begin{aligned}\rho''(|z|) &= \begin{cases}1 & |z| \le k_0\\ 0 & |z| > k_0\end{cases}\,, & (32)\\ \frac{\partial\exp\{-j\arg\{z\}\}}{\partial z^*} &= \frac{\partial}{\partial z^*}\sqrt{\frac{z^*}{z}} = \frac{1}{2|z|}\,. & (33)\end{aligned}$$

Introducing (25), (26), (32), and (33) into (31), we obtain

$$\begin{aligned}\frac{\partial\psi_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)} &= -\frac{\partial\underline{e}_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)}\cdot\frac{1}{s_n^2(r)}\begin{cases}\frac{1}{2} & \text{for } \frac{|\underline{e}_n(r)|}{s_n(r)} \le k_0\\ \frac{k_0 s_n(r)}{4|\underline{e}_n(r)|} & \text{for } \frac{|\underline{e}_n(r)|}{s_n(r)} > k_0\end{cases} \\ &= -\frac{\partial\underline{e}_n^*(r)}{\partial\underline{\mathbf{w}}^*(r)}\cdot\gamma_n(r)\,.\end{aligned} \tag{34}$$

Forming a diagonal $2N \times 2N$ matrix $\underline{\mathbf{\Gamma}}(r)$ with $\gamma_n(r)$, $n = 0, 1, \ldots, 2N-1$, on the main diagonal,

$$\underline{\mathbf{\Gamma}}(r) = \operatorname{diag}\{(\gamma_0(r), \gamma_1(r), \ldots, \gamma_{2N-1}(r))\}\,, \tag{35}$$

and introducing (27) and (34) into (30), we obtain

$$\nabla^2\xi(r) = 4\underline{\mathbf{X}}^H(r)\left(\mathbf{G}_{2N\times 2N}^{01}\right)^H\underline{\mathbf{\Gamma}}(r)\mathbf{G}_{2N\times 2N}^{01}\underline{\mathbf{X}}(r)\,. \tag{36}$$

An estimate of the expected value $\underline{\mathbf{\Lambda}}(r) = \mathcal{E}\{\nabla^2\xi(r)\}$ is obtained by recursively averaging $\nabla^2\xi(r)$ with a forgetting factor $0 < \lambda < 1$ as

$$\hat{\underline{\mathbf{\Lambda}}}(r) = \lambda\hat{\underline{\mathbf{\Lambda}}}(r-1) + (1-\lambda)\nabla^2\xi(r) \tag{37}$$

[5], [11]. This choice of recursive estimate of $\underline{\mathbf{\Lambda}}(r)$ leads to the MC-FDAF algorithm with RLS-like properties for $k_0 \to \infty$, $s_0(r) = s_1(r) = \cdots = s_{2N-1}(r) = 1/\sqrt{2}$. (See (42).)

3) Approximations: Since the Newton-type adaptation step (22) requires the inverse of the $2NQ \times 2NQ$ matrix $\nabla^2\xi(r)$, an approximation of (36) for reducing the computational complexity may be necessary for practical systems. Following [8], [9], we can approximate $\mathbf{G}^{01}_{2N\times 2N} \approx \frac{1}{2}\mathbf{I}_{2N\times 2N}$ for sufficiently large N, which leads to

$$\nabla^2\xi(r) \approx \underline{\mathbf{X}}^H(r)\underline{\mathbf{\Gamma}}(r)\underline{\mathbf{X}}(r)\,. \tag{38}$$

For calculating $\hat{\underline{\mathbf{\Lambda}}}^{-1}(r)$ using (37), the block-diagonal structure of $\nabla^2\xi(r)$ can be used to transform the $2NQ \times 2NQ$ matrix $\nabla^2\xi(r)$ 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 $Q \times Q$ [9].

The adaptive algorithm for a MISO system is finally given by (17), (22), (29), (37), and (38). A generalization of the adaptive algorithm to MIMO systems

$$\underline{\mathbf{W}}(r) = \big(\underline{\mathbf{w}}_0(r),\, \underline{\mathbf{w}}_1(r),\, \ldots,\, \underline{\mathbf{w}}_{P-1}(r)\big) \tag{39}$$

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 adaptive algorithm can be written for the MIMO case as follows:

$$\begin{aligned}
\mathbf{e}_p(rR) &= \mathbf{y}_{\mathrm{ref},p}(rR) - \mathbf{W}^{01}_{N\times 2N}\mathbf{F}^{-1}_{2N\times 2N}\underline{\mathbf{X}}(r)\underline{\mathbf{w}}_p(r) &(40)\\
\hat{\underline{\mathbf{\Lambda}}}_p(r) &= \lambda\hat{\underline{\mathbf{\Lambda}}}_p(r-1) + (1-\lambda)\underline{\mathbf{X}}^H(r)\underline{\mathbf{\Gamma}}_p(r)\underline{\mathbf{X}}(r)\,, &(41)\\
\underline{\mathbf{w}}_p(r) &= \underline{\mathbf{w}}_p(r-1) - \\
&\quad -2\underline{\boldsymbol{\mu}}_p(r)\hat{\underline{\mathbf{\Lambda}}}_p^{-1}(r)\underline{\mathbf{X}}^H(r)\mathbf{G}^{01}_{2N\times 2N}\underline{\boldsymbol{\psi}}_p(r)\,. &(42)
\end{aligned}$$

Note that, in contrast to the MC-FDAF, we have to calculate the inverse of the weighted cross-power spectral density matrix $\hat{\underline{\mathbf{\Lambda}}}_p(r)$ for all P output channels due to the dependency on $\underline{\mathbf{\Gamma}}_p(r)$. The MC-FDAF is obtained for $\underline{\mathbf{\Gamma}}_p(r) = \mathbf{I}_{2N\times 2N}$ and $\mathbf{G}^{01}_{2N\times 2N}\underline{\boldsymbol{\psi}}_p(r) = \underline{\mathbf{e}}_p(r)$. In addition to [11, (29)], we may notice that the update equation (42) allows for a bin-wise operation with a bin-dependent stepsize $\mu_n(r)$ and a bin-dependent scale parameter $s_n(r)$. Moreover, the derivation based on the cost function in the DFT domain (20) does not require the approximation [11, (31)] of the weighting matrix which is equivalent to $\underline{\mathbf{\Gamma}}(r)$ for obtaining an efficient realization of the algorithm.

III. Application to adaptive beamforming with DFT-bin-wise double-talk detection

For verification, we apply the proposed algorithm to multichannel speech enhancement using a DFT-domain realization of a GSC [12].

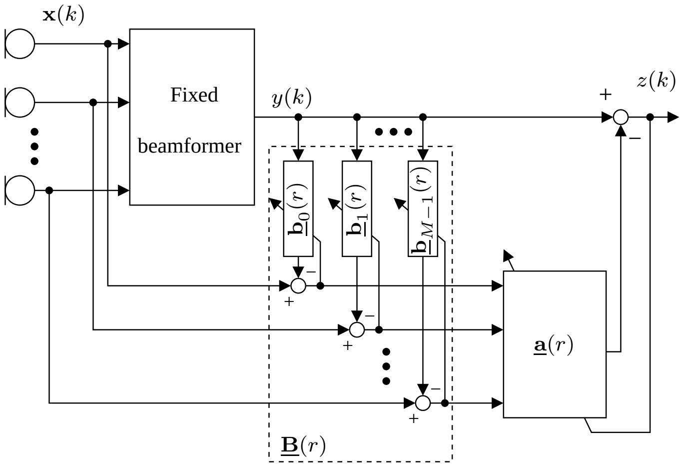

Fig. 1. GSC with an adaptive blocking matrix after [14].

A. Overview of the GSC using an adaptive blocking matrix

The GSC using an adaptive blocking matrix is depicted in Fig. 1. We identify three blocks, the fixed beamformer, the adaptive blocking matrix $\underline{\mathbf{B}}(r)$, and the interference canceller $\underline{\mathbf{a}}(r)$. The blocking matrix $\underline{\mathbf{B}}(r)$ and the interference canceller $\underline{\mathbf{a}}(r)$ are realized by systematically applying the MC-BRFDAF.

1) Fixed beamformer: The fixed beamformer steers the sensor array to the position of the desired source and enhances the desired signal relative to the interference. The fixed beamformer forms the reference path of the GSC. Usually, the fixed beamformer is designed to allow movements of the desired source within a given area so that the desired signal is not attenuated. Although all known beamformer designs can be used for realizing the fixed beamformer, we use a simple uniformly weighted beamformer for simplicity. Since, especially for small-scale microphone arrays ($M \leq 8$), the interference suppression of fixed beamformers is not sufficient for many applications, the adaptive sidelobe cancelling path –consisting of adaptive blocking matrix and interference canceller– is required.

2) Adaptive blocking matrix: The blocking matrix $\underline{\mathbf{B}}(r)$ is a spatial filter which suppresses the desired signal and passes interference such that the output of $\underline{\mathbf{B}}(r)$ is a reference for the interference. In contrast to fixed blocking matrices $\underline{\mathbf{B}}$, which do not perfectly suppress the desired signal whenever there is a mismatch between the spatial filtering of $\underline{\mathbf{B}}$ and the actual wave field of the desired signal, adaptive blocking matrices can track changes of the wave field of the desired signal. This is especially important in time-varying reverberant environments, where fixed blocking matrices continuously let through desired signal components. Using multi-channel adaptive filtering, an adaptive blocking matrix can be realized by using the output signal of the fixed beamformer as a reference and by subtracting this reference from each channel of the sidelobe cancelling path using adaptive filters [14].

3) Interference canceller: Using the output signal of the blocking matrix as a reference for the interference, the interference canceller $\underline{\mathbf{a}}(r)$ adaptively subtracts the residual interference from the reference path using adaptive filtering.

B. Adaptation control

The fixed beamformer cannot produce an estimate of the desired signal that is free of interference. Therefore, the blocking matrix should only be adapted when the signal-to-interference ratio (SIR) is high in order to prevent suppression of interference by the blocking matrix. Interference components that are suppressed by the blocking matrix cannot be cancelled by the interference canceller, and, thus, leak to the output of the GSC. Generally, the blocking matrix does not produce an estimate of the interference which is completely free of the desired signal. Therefore, the interference canceller should only be adapted if the SIR is low in order to prevent cancellation and distortion of the desired signal. Higher tracking capability is obviously obtained when the decision 'adaptation of $\underline{\mathbf{B}}(r)$' or 'adaptation of $\underline{\mathbf{a}}(r)$' is made in separate frequency bins and not for the fullband signals since sparseness of the desired signal and of interference in the time-frequency domain can be exploited. This motivates the usage of frequency-domain adaptive filtering for the realization of the GSC [12].

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 [12]. Using the directivity of a fixed beamformer, The activity detector forms a biased estimate $\Upsilon(r,n)$ of the ratio of the power spectral densities (PSDs) $SIR(r,n) = S_{dd}(r,n)/S_{ii}(r,n)$ of the desired signal, $S_{dd}(r,n)$, and of the interference, $S_{ii}(r,n)$, and tracks the maxima and the minima of $\Upsilon(r,n)$. The blocking matrix and the interference canceller are adapted whenever $\Upsilon(r,n)$ is maximum or minimum, respectively.

C. Motivation for the usage of double-talk robust FDAFs

In Fig. 2, a typical example for the behavior of the adaptation control for male desired speech and orchestra music played by a loudspeaker is shown. The experimental setup corresponds to that of Sect. III-D. Figure 2a and Fig. 2b show the desired signal and the interference signal recorded at the $M/2$-th microphone. In Fig. 2c, the ratio $SIR(r,n)$ of recursively averaged estimates of the PSDs of the desired signal and of the interference is depicted as a function of frequency n (in kHz) and block time r. In Fig. 2d, the decision based on $SIR(r,n)$ is shown. Blocking matrix (BM) and interference canceller (IC) are adapted for $10\log_{10} SIR(r,n) \geq 15\,\mathrm{dB}$ and for $10\log_{10} SIR(r,n) \leq -15\,\mathrm{dB}$, respectively.[1] Figure 2e illustrates the decision of the adaptation control using $\Upsilon(r,n)$. It may be seen that the adaptation control in Fig. 2e does not always detect activity of the desired signal and of interference correctly. The blocking matrix and the interference canceller may thus be adapted during double-talk, which leads to outliers in the adaptive filters. These outliers –and potential

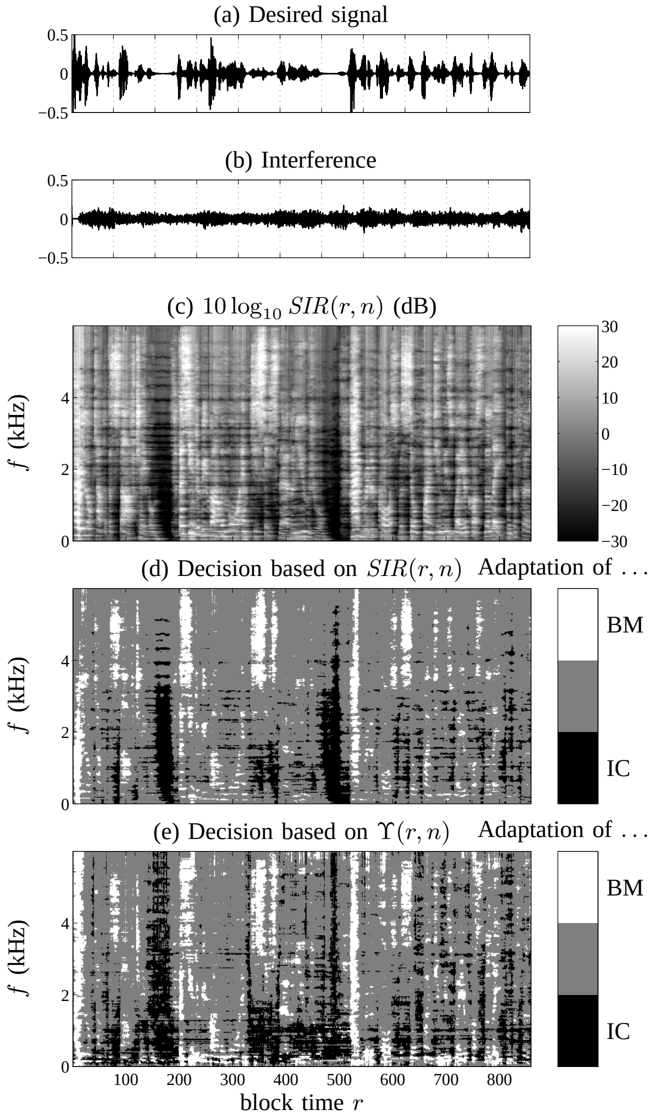

Fig. 2. Typical behavior of the adaptation control for the GSC for male desired speech and orchestra music played back by a loudspeaker.

divergence of the adaptive filters– may be prevented (a) by reducing the stepsizes of the adaptive filters or (b) by reducing the adaptation thresholds so that the adaptive filters are less likely adapted during double-talk.

However, both options reduce the tracking capability and, thus, the interference suppression of the GSC. For avoiding this trade-off, we apply the MC-BRFDAF to the blocking matrix and to the interference canceller. From Fig. 1, 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 canceller 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 (42) with $\underline{\mathbf{w}}_p(r)$, $p = m = 0, 1, \ldots, M-1$, and identifying the adaptive filter $\underline{\mathbf{a}}(r)$ with $\underline{\mathbf{w}}_p(r)$, $p = 0$, we can systematically use the MC-BRFDAF for the adaptation of the GSC. The stepsizes $\underline{\boldsymbol{\mu}}(r)$ are determined by the adaptation control and switched between

[1] Experiments with this adaptation control showed that maximum interference suppression and minimum distortion of the desired signal is obtained with these thresholds.

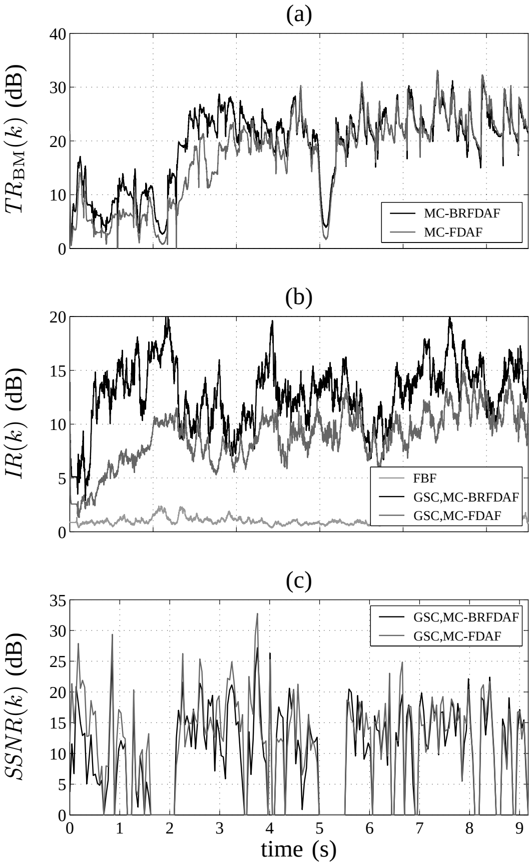

Fig. 3. Comparison of the GSC using MC-FDAF and MC-BRFDAF for 'continuous' double-talk: **(a)** Suppression of the desired signal $TR_{\mathrm{BM}}(k)$ by the blocking matrix, **(b)** suppression of the interference $IR(k)$ by the GSC and by the fixed beamformer (FBF), and **(c)** distortion of the desired signal measured by the segmental SNR $SSNR(k)$ between the FBF output and the GSC output for data blocks of length 512 (Sampling rate $12\,\mathrm{kHz}$, $N = 256$, $R = 64$, $\lambda = 0.97$, $k_0 = 0.5$, MC-FDAF: $\mu_c = [0.7, 0.2] \cdot (1 - \lambda)$ [BM,IC], MC-BRFDAF: $\mu_c = [1.3, 1.3] \cdot (1 - \lambda)$ [BM,IC]).

0 and a frequency-independent constant value μ_c for disabling and enabling the adaptation.

D. Experimental results

The GSC realized by MC-BRFDAF and by MC-FDAF is applied to a microphone array with $12\,\mathrm{cm}$ aperture and $M = 4$ equally spaced sensors in an office room with $T_{60} = 250\,\mathrm{ms}$ reverberation time. The desired signal of Fig.2a arrives from broadside direction from a distance of $60\,\mathrm{cm}$. The interference of Fig.2b is located at $120\,\mathrm{cm}$ in endfire direction. The average SIR at the sensors is $3\,\mathrm{dB}$. The parameters which are summarized below Fig. 3 are optimized for maximum convergence speed and maximum noise-suppression after convergence. They are the same for both GSC realizations except for the constant stepsize parameter μ_c. Figure 3a–c show the suppression $TR_{\mathrm{BM}}(k)$ of the desired signal by the blocking matrix, the interference suppression $IR(k)$ of the GSC, and the distortion $SSNR(k)$ of the desired signal by the GSC as a function of time after initialization of the system, respectively. $SSNR(k)$ is the segmental SNR between the output of the fixed beamformer and the output of the GSC for the desired signal only. Ideally, $SSNR(k) = \infty$ since the interference canceller should not distort the desired signal. It may be seen that the blocking matrix (Fig. 3a) and the interference canceller (Fig. 3b) converge faster for MC-BRFDAF than for MC-FDAF, since larger stepsizes can be chosen for MC-BRFDAF due to the improved robustness against double-talk. While $TR_{\mathrm{BM}}(k)$ converges for both GSCs to nearly the same value, $IR(k)$ is about $4\,\mathrm{dB}$ greater for MC-BRFDAF than for MC-FDAF after convergence. This result is confirmed by application of the algorithm to various mixtures of speech signals. The distortion $SSNR(k)$ (Fig. 3c) for MC-BRFDAF is slightly higher for MC-BRFDAF than for MC-FDAF.

IV. CONCLUSIONS

We presented a DFT-domain adaptive filter which is robust against double-talk and which is especially designed for wide-band signals, where adaptation with DFT-bin-wise stepsize controls is advantageous. The efficiency of the approach was motivated by applying the adaptive filter to a generalized sidelobe canceller for audio signals. Experiments showed that the improved robustness leads to an improved convergence behavior and higher noise-reduction even for small-scale microphone arrays.

REFERENCES

[1] M.M. Sondhi. An adaptive echo canceller. *The Bell System Technical Journal*, XLVI(3):497–510, March 1967.

[2] T. Gänsler. A double-talk resistant subband echo canceller. *Signal Processing*, 65(1):89–101, February 1998.

[3] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[4] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty. Double-talk robust fast converging algorithms for network echo cancellation. *IEEE Trans. on Speech and Audio Processing*, 8(6):656–663, November 2000.

[5] J. Benesty and T. Gänsler. A robust fast converging least-squares adaptive algorithm. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 6:3785–3788, May 2001.

[6] J.J. Shynk. Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine*, pages 14–37, January 1992.

[7] J.-S. Soo and K.K. Pang. Multidelay block frequency domain adaptive filter. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 38(2):373–376, February 1990.

[8] J. Benesty and D.R. Morgan. Multi-channel frequency-domain adaptive filtering. In S.L. Gay and J. Benesty, editors, *Acoustic Signal Processing for Telecommunication*, chapter 7, pages 121–133. Kluwer Academic Publishers, Boston, MA, 2000.

[9] H. Buchner, J. Benesty, and W. Kellermann. Multichannel frequency-domain adaptive filtering with application to multichannel acoustic echo cancellation. In J. Benesty and Y. Huang, editors, *Adaptive Signal Processing: Applications to Real-World Problems*. Springer, 2003.

[10] B. H. Nitsch. A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain. *Signal Processing*, 80(9):1733–1745, September 2000.

[11] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann. An outlier robust extended multidelay filter with application to acoustic echo cancellation. *Int. Workshop on Acoustic Echo and Noise Control*, September 2003.

[12] W. Herbordt. *Combination of robust adaptive beamforming with acoustic echo cancellation for acoustic human/machine front-ends*. PhD thesis, University Erlangen-Nuremberg, Germany, 2004.

[13] L.J. Griffiths and C.W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. on Antennas and Propagation*, 30(1):27–34, January 1982.

[14] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. on SP*, 47(10):2677–2684, Oct. 1999.

[15] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Upper Saddle River, NJ, 1993.