BLOCK-BASED MULTICHANNEL TRANSFORM-DOMAIN ADAPTIVE FILTERING

Sascha Spors, Herbert Buchner, and Karim Helwani

Deutsche Telekom Laboratories, Technische Universität Berlin,
Ernst-Reuter-Platz 7, 10587 Berlin, Germany.
Email: Sascha.Spors@telekom.de

ABSTRACT

Multichannel adaptive filtering is subject to specific problems emerging from spatio-temporal couplings in the input signals of the adaptive filter. Transform-domain adaptive filtering (TDAF) decouples the input signal of the adaptive filter by a suitably chosen transformation. In a previous paper, the authors have introduced a two-stage approach to multichannel TDAF. However, the approach presented there is based on a sample-by-sample update of the filter coefficients. In this paper we present a more practical block-based formulation of multichannel TDAF that is constructed from a combination of frequency-domain adaptive filtering for temporal decoupling and an unitary transform for spatial decoupling.

1. INTRODUCTION

Telecommunication systems with more than one acoustic transmission channel are being developed and increasingly used. These systems aim at providing additional spatial auditory cues to the listener in contrast to the single channel systems frequently used in the last decades. The spatial cues increase the naturalness of the communication and can facilitate, for instance, the recognition of speakers in a dialogue by their spatial position.

Acoustic echo cancelation (AEC) is required for full-duplex communication in a hands-free communication scenario. The application of AEC to such a scenario is illustrated in Fig. 1. The goal of AEC is to cancel the acoustic echo for the far-end, introduced by the couplings between the loudspeaker(s) and microphone(s) at the near-end. In the block diagram of Fig. 1, the echo produced by the acoustic couplings between the P loudspeakers and the microphone in the near-end room is canceled for the far-end by subtracting the estimate $\hat{y}(n)$ of the microphone signal from the actual microphone signal $y(n)$. The signal $\hat{y}(n)$ is derived by filtering the loudspeaker signals $x_p(n)$ with finite-impulse response (FIR) filters that model the acoustic paths $h_p(n)$ from the loudspeakers to the microphones. The estimation of the acoustic paths $h_p(n)$ represents a multichannel identification problem. It is well known that this identification problem is typically ill-conditioned for the multichannel case if the far-end signals exhibit spatio-temporal correlations [1].

In advanced adaptation schemes, at least two fundamental approaches exist to cope with the far-end correlations: (1) decoupling of the convolution in the near-end room and (2) decoupling of the input (loudspeaker signal) covariance matrix. The first approach is applied in frequency-domain adaptive filtering (FDAF), while the second one is applied in transform-domain adaptive filtering (TDAF).

The authors have introduced a two-stage approach to multichannel TDAF in [2]. However, the approach presented there was based on a sample-by-sample update of the filter coefficients. Block-based adaptation algorithms are typically computationally less complex and therefore favorable. In this paper we present a block-based formulation of multichannel TDAF that is constructed from a combination of FDAF for temporal decoupling and TDAF for spatial decoupling.

We proceed as follows: The next section will introduce the fundamental problem of multichannel system identification. This will be followed by a brief review of TDAF and FDAF before we derive the block-based TDAF algorithm. Some results computed with the proposed algorithm will be shown before concluding the paper.

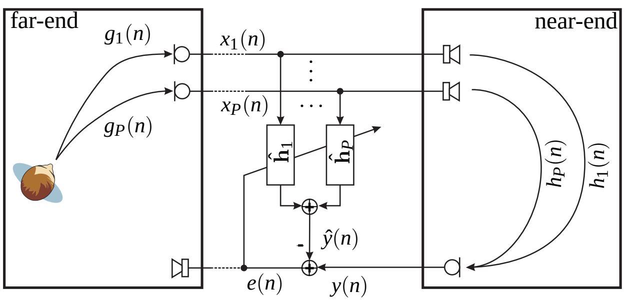


Figure 1: Block diagram of multichannel acoustic echo cancelation.

2. MULTICHANNEL SYSTEM IDENTIFICATION

The estimation of the acoustic paths $h_p(n)$ for $p = 1,2,\ldots,P$ represents a multichannel identification problem. The error $e(n)$ is given as

$$e(n) = y(n) - \sum_{p=1}^{P} \hat{\mathbf{h}}_p^T \mathbf{x}_p(n) , \tag{1}$$

where

$$\hat{\mathbf{h}}_p = [\hat{h}_{p,0}, \hat{h}_{p,1}, \ldots, \hat{h}_{p,L-1}]^T , \tag{2}$$

$$\mathbf{x}_p(n) = [x_p(n), x_p(n-1), \ldots, x_p(n-L+1)]^T , \tag{3}$$

with $\hat{h}_{p,l}$ denoting the l-th coefficient of the p-th channel, L the filter length and n the time instant. Under the assumption of minimizing the mean-square error (MSE) the filter coefficients can be found by solving the multichannel normal equation [1]

$$\mathbf{R}_{xx}\, \hat{\mathbf{h}} = \mathbf{r}_{xy} , \tag{4}$$

where the $PL \times 1$ vector $\hat{\mathbf{h}}$ of estimated filter coefficients is given as $\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T, \hat{\mathbf{h}}_2^T, \ldots, \hat{\mathbf{h}}_P^T]^T$. The matrix $\mathbf{R}_{xx}$ denotes the covariance matrix of the input signals $\mathbf{x}(n)$ and $\mathbf{r}_{xy}$ the covariance vector between the input $\mathbf{x}(n)$ and the microphone signal $y(n)$. The $PL \times LP$ covariance matrix $\mathbf{R}_{xx}$ is defined as

$$\mathbf{R}_{xx}(n) = \hat{\mathscr{E}}\{\mathbf{x}(n)\mathbf{x}^T(n)\} , \tag{5}$$

where $\hat{\mathscr{E}}\{\cdot\}$ denotes a suitable approximation of the expectation operator and the $PL \times 1$ vector $\mathbf{x}$ of input signals is given as $\mathbf{x}(n) = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_P^T]^T$. The $PL \times 1$ covariance vector $\mathbf{r}_{xy}(n) = \hat{\mathscr{E}}\{\mathbf{x}(n)y(n)\}$. The covariance matrix $\mathbf{R}_{xx}$ is composed from $L \times L$ sub-matrices that are given as $\mathbf{R}_{pq} = \hat{\mathscr{E}}\{\mathbf{x}_p(n)\mathbf{x}_q^T(n)\}$ for $p,q = 1,2,\ldots,P$. Typically, these are assumed to be Toeplitz matrices.

The generalization to multiple microphones and consequently multiple-input multiple-output (MIMO) systems in the near-end room is straightforward. It can be shown that the resulting normal equation for the MIMO case can be decomposed into a series

of independent multiple-input single-output (MISO) normal equations [1] for each microphone channel. Hence, the consideration of a MISO system in the near-end room is sufficient in the context of this work.
The solution of the normal equation (4) is subject to numerical problems when the covariance matrix $\mathbf{R}_{xx}$ is ill-conditioned. It can be shown that this is the case when spatio-temporal correlations exist between the loudspeaker signals $x_p(n)$.

3. A TWO-STAGE APPROACH TO MULTICHANNEL TDAF

Transform-domain adaptive filtering (TDAF) is a technique that performs the filter adaptation in a transform domain. In the ideal case, the far-end signals will be decorrelated by a suitably chosen transformation. The ideal transformation can be deduced from the covariance matrix $\mathbf{R}_{xx}(n)$ and is data-dependent in general. TDAF has originally been introduced for the single channel case [3]. In a previous paper [2] we have proposed multichannel TDAF, based on a two-step decoupling of the covariance matrix. The approach is briefly reviewed in the following.

3.1 Spatio-temporal decoupling

The spatio-temporal decoupling consists of two steps: (1) temporal decoupling using a discrete Fourier transform (DFT) based transformation and (2) a spatial decoupling using a unitary transform.
Assuming stationary signals $x_p(n)$ and the correlation method to estimate the covariance matrix, the sub-matrices $\mathbf{R}_{pq}$ exhibit Toeplitz structure [4]. These assumptions hold well for typical signals. For large block lengths ($L \to \infty$) the matrices $\mathbf{R}_{pq}$ become equivalent to circulant matrices [5]. Circulant matrices can be diagonalized by the DFT

$$\mathbf{R}_{xx} = \mathbf{F}\,\underline{\mathbf{S}}_{xx}\,\mathbf{F}^H \,, \tag{6}$$

where $\mathbf{F}$ denotes a $PL \times LP$ block-diagonal matrix whose diagonal blocks are composed from $L \times L$ DFT matrices $\mathbf{F}_L$. Frequency domain quantities are underlined. The elements of the (normalized) DFT matrices $\mathbf{F}_L$ are given as $f_{nm} = 1/\sqrt{L} \cdot e^{-j2\pi nm/L}$ for $n, m = 0, 1, \ldots, L-1$. The block-matrix $\underline{\mathbf{S}}_{xx}$ is composed from the $L \times L$ diagonal matrices

$$\underline{\mathbf{S}}_{pq} = \operatorname{diag}\{\underline{s}_{pq}^{(0)}, \underline{s}_{pq}^{(1)}, \ldots, \underline{s}_{pq}^{(L-1)}\} \,, \tag{7}$$

where the elements $\underline{s}_{pq}^{(\nu)}$ for $\nu = 0, 1, \ldots, L-1$ are given by the DFT of the first column of $\mathbf{R}_{pq}$. The frequency bin is denoted as ν.
In order to achieve further spatial decoupling, the matrix $\underline{\mathbf{S}}_{xx}$ has to be reordered such that all spatial couplings for one frequency bin are combined into submatrices $\underline{\mathbf{S}}^{(\nu)}$. Formally, this can be reached by a suitably chosen permutation matrix $\mathbf{A}_L$. The submatrices $\underline{\mathbf{S}}^{(\nu)}$ can then be diagonalized by application of the spectral theorem. Combining all described steps, the covariance matrix $\mathbf{R}_{xx}$ can be expressed as

$$\mathbf{R}_{xx} = \mathbf{F}\mathbf{A}_L \underline{\mathbf{U}}_L\, \underline{\underline{\mathbf{T}}}_{xx}\, \underline{\mathbf{U}}_L^H \mathbf{A}_L^T \mathbf{F}^H \tag{8}$$

in terms of the diagonal matrix $\underline{\underline{\mathbf{T}}}_{xx}$ which is composed from the spatio-temporal eigenvalues of $\mathbf{R}_{xx}$. These eigenvalues can be linked to the spatio-temporal correlation coefficients of the input signals $x_p(n)$ [2].
The $LP \times PL$ matrix $\underline{\mathbf{U}}_L$ denotes a block-diagonal matrix composed from the $P \times P$ submatrices $\underline{\mathbf{U}}^{(\nu)}$ constructed from the singular vectors of $\underline{\mathbf{S}}^{(\nu)}$. Note, that the desired decoupling of the covariance matrix has been achieved by a set of suitably chosen unitary transforms. This favorable property is beneficial for mathematical rearrangements in the algorithm.

3.2 Multichannel TDAF

Introducing Eq. (8) into the normal equation (4) and exploiting the unitarity of the transform matrices yields the transformed normal equation

$$\underline{\underline{\mathbf{T}}}_{xx}\, \underbrace{\underline{\mathbf{U}}_L^H \mathbf{A}_L^T \mathbf{F}^H \hat{\mathbf{h}}}_{\underline{\underline{\hat{\mathbf{h}}}}} = \underbrace{\underline{\mathbf{U}}_L^H \mathbf{A}_L^T \mathbf{F}^H \mathbf{r}_{xy}}_{\underline{\underline{\mathbf{t}}}_{xy}} \,, \tag{9}$$

where $\underline{\underline{\hat{\mathbf{h}}}}$ and $\underline{\underline{\mathbf{t}}}_{xy}$ denote the transformed vector of filter coefficients $\hat{\mathbf{h}}$ and the transformed covariance vector $\mathbf{r}_{xy}$, respectively. Since $\underline{\underline{\mathbf{T}}}_{xx}$ is diagonal, the normal equation (4) has been decomposed by the transformations into a series of scalar equations.
The solution of the normal equation (9) involves the inversion of the diagonal matrix $\underline{\underline{\mathbf{T}}}_{xx}$ containing the spatio-temporal eigenvalues of $\mathbf{R}_{xx}$. If one or more of these are zero or close to zero this will be subject to numerical problems. It was shown in [2] that these eigenvalues are linked to the spatio-temporal correlations in the far-end signals and that strong correlations lead to eigenvalues that are (close to) zero. One benefit of TDAF is that a regularization can be performed spatially and temporally frequency-bin selective.
The derived transformations have been applied straightforwardly to the recursive-least squares (RLS) algorithm in [2]. The formulation is based on a sample-by-sample update of the filter coefficients. However, for a practical implementation block-based algorithms are favorable. The presented two-step approach to multichannel TDAF allows the utilization of known frequency domain techniques like frequency domain adaptive filtering (FDAF) for the temporal decoupling. After a brief review of generalized FDAF in the next section, a combination of TDAF and FDAF will be developed in Section 5.

4. FREQUENCY-DOMAIN ADAPTIVE FILTERING

This section presents a brief review of generalized FDAF [6, 7]. FDAF is essentially based on a block formulation of the identification problem. This block formulation is derived by combining L consecutive samples into blocks, formulating the error signal (1) in terms of blocks and minimizing the error. For this purpose, the convolution operation in (1) is reformulated in terms of a matrix operation, where the input signals are combined into a matrix with Toeplitz structure. A Toeplitz matrix can be transformed into a circulant matrix by doubling its size. This concept is a fundamental building block of FDAF where the circulant matrix is then diagonalized by the DFT. This results in an overlap save formulation of the convolution by incorporating window functions.
The concept of generalized multichannel FDAF is closely linked to TDAF in the sense that it also aims at temporal decoupling. It is well known that the Fourier transformation diagonalizes linear time-shift invariant systems. FDAF employs the DFT for temporal decoupling of the near-end system. It can be shown [6] that this leads also to an approximate temporal decoupling of the covariance matrix $\mathbf{R}_{xx}$. This is due to fact that the DFT only approximately decouples the covariance matrix for the finite blocksize in practical implementations [5].

4.1 Algorithm

The time-domain block error signal $\mathbf{e}(m)$ for a block length of L samples is defined as

$$\mathbf{e}(m) = [e(mL), e(mL+1), \ldots, e(mL+L-1)]^T \,, \tag{10}$$

where m denotes the block index. The microphone signal $\mathbf{y}(m)$ is defined in a similar fashion as $\mathbf{e}(m)$. In order to derive an algorithm that requires only DFTs of size $2L$, the error and microphone signals are zero padded before transformation into the frequency domain

$$\underline{\mathbf{e}}'(m) = \mathbf{F}_{2L}\left[\mathbf{0}_{1\times L}, \mathbf{e}^T(m)\right]^T \,, \tag{11}$$

and similarly for the microphone signal. The loudspeaker signals in the frequency domain are given as

$$\underline{\mathbf{X}}_p(m) = \operatorname{diag}\{\mathbf{F}_{2L}[x_p(mL-L), \ldots, x_p(mL+L-1)]^T\} \,, \tag{12a}$$
$$\underline{\mathbf{X}}(m) = [\underline{\mathbf{X}}_1(m), \ldots, \underline{\mathbf{X}}_P(m)] \,. \tag{12b}$$

The generic FDAF algorithm for MISO systems can then be summarized as follows [7]

$$\underline{\mathbf{S}}(m) = \lambda \underline{\mathbf{S}}(m-1) + (1-\lambda)\underline{\mathbf{X}}^H(m)\mathbf{G}_1\underline{\mathbf{X}}(m) , \tag{13a}$$

$$\underline{\mathbf{K}}(m) = (1-\lambda)\underline{\mathbf{S}}^{-1}(m)\underline{\mathbf{X}}^H(m) , \tag{13b}$$

$$\underline{\mathbf{e}}'(m) = \underline{\mathbf{y}}'(m) - \mathbf{G}_2\underline{\mathbf{X}}(m)\hat{\underline{\mathbf{h}}}'(m-1) , \tag{13c}$$

$$\hat{\underline{\mathbf{h}}}'(m) = \hat{\underline{\mathbf{h}}}'(m-1) + \mathbf{G}_3\underline{\mathbf{K}}(m)\underline{\mathbf{e}}'(m) , \tag{13d}$$

where λ denotes the forgetting factor and $\hat{\underline{\mathbf{h}}}'(m)$ the zero padded vector of estimated filter coefficients which is defined as

$$\hat{\underline{\mathbf{h}}}'(m) = \mathbf{G}^{10}_{2LP\times LP}\hat{\underline{\mathbf{h}}}'(m) , \tag{14}$$

where $\mathbf{G}^{10}_{2LP\times LP}$ denotes a window matrix that performs the zero padding. It is defined as follows

$$\mathbf{G}^{10}_{2LP\times LP} = \text{Bdiag}\{\mathbf{G}^{10}_{2L\times L}, \ldots, \mathbf{G}^{10}_{2L\times L}\} , \tag{15a}$$

$$\mathbf{G}^{10}_{2L\times L} = \mathbf{F}_{2L}[\mathbf{I}_{L\times L}, \mathbf{0}_{L\times L}]^T \mathbf{F}_L^{-1} . \tag{15b}$$

In the FDAF algorithm the finite block length is explicitly accounted for by the constraint matrices $\mathbf{G}_1$, $\mathbf{G}_2$ and $\mathbf{G}_3$. These are defined as

$$\mathbf{G}_1 = \mathbf{G}_2 = \mathbf{F}_{2L}\text{Bdiag}\{\mathbf{0}_{L\times L}, \mathbf{I}_{L\times L}\}\mathbf{F}_{2L}^{-1} , \tag{16a}$$

$$\mathbf{G}_3 = \text{Bdiag}\{\mathbf{G}^{10}_{2L\times 2L}, \ldots, \mathbf{G}^{10}_{2L\times 2L}\} , \tag{16b}$$

$$\mathbf{G}^{10}_{2L\times 2L} = \mathbf{F}_{2L}\text{Bdiag}\{\mathbf{I}_{L\times L}, \mathbf{0}_{L\times L}\}\mathbf{F}_{2L}^{-1} . \tag{16c}$$

The frequency domain algorithm given by (13) provides the optimal solution of the normal equation (4). The formulation can be extended straightforwardly to include partitioned impulse responses [6]. Partitioning improves the performance in the context of nonstationary signals and time-varying near-end system. It also allows to improve the delay in a practical implementation. For $K = L$ partitions, this algorithm is equivalent to the time domain RLS algorithm.
Based on the generic FDAF algorithm a number of special cases and approximations can be derived that lead to most known algorithms and efficient algorithms [6, 7]. One frequently applied approximation, that is relevant in the context of this paper, will be discussed in the following.

4.2 Approximations

As discussed for TDAF in Section 3.1, the submatrices $\mathbf{R}_{pq}$ of $\mathbf{R}_{xx}$ are assumed to be Toeplitz. The same holds when doubling the block-size, as performed in FDAF. However, the DFT only diagonalizes Toeplitz matrices in the limiting case for $L \to \infty$. The resulting submatrices of $\underline{\mathbf{S}}(m)$ will contain off-diagonal elements in practical implementations with finite block lengths. As a result, the frequency domain covariance matrix $\underline{\mathbf{S}}(m)$ is not exactly (blockwise) diagonal in general. Hence, computing the inverse in (13b) results in a high computational complexity.
Approximating the constraint matrix $\mathbf{G}_1$ by $\mathbf{G}_1 = \mathbf{I}/2$ results in a blockwise diagonal structure of $\underline{\mathbf{S}}(m)$. It has been shown in [6] that this approximation provides good results for sufficiently large block lengths L.

5. BLOCK-BASED MULTICHANNEL TDAF

In order to derive a block-based algorithm for multichannel TDAF both block-based FDAF and the concept of TDAF are combined in the following. The two stage approach to TDAF presented in Section 3 separates the temporal decoupling from the spatial decoupling. Hence, FDAF can be utilized for temporal decoupling combined with the concept of spatial decoupling from TDAF. For this purpose the eigenvalue decomposition of TDAF is introduced into (13). It will be assumed in the following that $\mathbf{G}_1 = \mathbf{I}/2$. The generalization is straightforward as will be discussed later.

5.1 Algorithm

As for the TDAF approach introduced in Section 3, a reordering of the covariance matrix is desirable that combines the spatial couplings for one frequency bin into sub-matrices. This can be achieved in the framework of FDAF by post multiplying the frequency domain signal matrix $\underline{\mathbf{X}}(m)$ by a $2LP \times 2LP$ permutation matrix

$$\mathbf{A} = [\mathbf{1}_{ij}] , \tag{17}$$

composed from $P \times 2L$ submatrices for $i = 1,...,2L,\ j = 1,...,P$ where $\mathbf{1}_{ij}$ denotes a submatrix which contains a one at position (i, j) and zeros at all other positions. The reordering of the signal matrix $\underline{\mathbf{X}}(m)$ results in a reordering of the covariance matrix. The reordered covariance matrix is a block diagonal matrix composed from $P \times P$ matrices representing the spatial couplings for one of the $2L$ frequency bins. As for the sample-by-sample TDAF concept this allows a bin-wise eigenvalue decomposition. Hence, the concepts outlined in Section 3 can be introduced into the FDAF algorithm.
Introducing the eigenvalue decomposition of the covariance matrix

$$\underline{\mathbf{S}}(m) = \mathbf{A}\underline{\mathbf{U}}(m)\underline{\underline{\mathbf{T}}}(m)\underline{\mathbf{U}}^H(m)\mathbf{A}^T \tag{18}$$

for the block index m and $m-1$ into (13a) together with $\mathbf{G}_1 = \mathbf{I}/2$, and utilizing the unitarity of the spatial transformation $\underline{\mathbf{U}}$ yields

$$\underline{\underline{\mathbf{T}}}(m) = \lambda \underline{\mathbf{U}}^H(m)\underline{\mathbf{U}}(m-1)\underline{\underline{\mathbf{T}}}(m-1)\underline{\mathbf{U}}^H(m-1)\underline{\mathbf{U}}(m) + \frac{1}{2}(1-\lambda)\underline{\underline{\mathbf{X}}}^H(m)\underline{\underline{\mathbf{X}}}(m) , \tag{19}$$

where

$$\underline{\underline{\mathbf{X}}}(m) = \underline{\mathbf{X}}(m)\mathbf{A}\underline{\mathbf{U}}(m) \tag{20}$$

denotes the matrix of transformed far-end signals. The multiplication of the reordered signal matrix $\underline{\mathbf{X}}(m)\mathbf{A}$ by the singular matrix of $\underline{\mathbf{U}}(m)$ can be interpreted as filtering the far-end signals, where the MIMO filter is given by the singular vectors of $\underline{\mathbf{S}}(m)$. Hence, the desired decoupling can be achieved by filtering the far-end signals. The update equation (19) for the transformed covariance matrix $\underline{\underline{\mathbf{T}}}(m)$ contains combinations of singular matrices from the actual and the previous block index. This combination can be interpreted as a transformation of $\underline{\underline{\mathbf{T}}}(m-1)$ from the previous eigenspace to the actual one. This transformation is required for an exact formulation since the spatial eigenvectors are not constant from one block to the next one in general.
Introducing (18) and (20) into Eq. (13b) results in the following relation for the Kalman gain

$$\underline{\underline{\mathbf{K}}}(m) = (1-\lambda)\underline{\underline{\mathbf{T}}}^{-1}(m)\underline{\underline{\mathbf{X}}}^H(m) , \tag{21}$$

where

$$\underline{\underline{\mathbf{K}}}(m) = \underline{\mathbf{U}}^H(m)\mathbf{A}^T\underline{\mathbf{K}}(m) \tag{22}$$

denotes the transformed Kalman gain. Furthermore introducing (20) into the error signal of FDAF (13c) reads

$$\underline{\mathbf{e}}'(m) = \underline{\mathbf{y}}'(m) - \mathbf{G}_2\underline{\underline{\mathbf{X}}}(m)\underline{\mathbf{U}}^H(m)\underline{\mathbf{U}}(m-1)\hat{\underline{\underline{\mathbf{h}}}}'(m-1) , \tag{23}$$

where the transformed filter coefficients are defined as

$$\hat{\underline{\underline{\mathbf{h}}}}'(m-1) = \underline{\mathbf{U}}^H(m-1)\mathbf{A}^T\hat{\underline{\mathbf{h}}}'(m-1) . \tag{24}$$

Finally introducing (24) and (22) into (13d) yields the coefficient update as

$$\hat{\underline{\underline{\mathbf{h}}}}'(m) = \underline{\mathbf{U}}^H(m)\underline{\mathbf{U}}(m-1)\hat{\underline{\underline{\mathbf{h}}}}'(m-1) + \tilde{\mathbf{G}}_3\underline{\underline{\mathbf{K}}}(m)\underline{\mathbf{e}}'(m) , \tag{25}$$

where the constraint $\tilde{\mathbf{G}}_3$ is given as

$$\tilde{\mathbf{G}}_3 = \underline{\mathbf{U}}^H(m)\mathbf{A}^T\mathbf{G}_3\mathbf{A}\underline{\mathbf{U}}(m) . \tag{26}$$

The derived block-based TDAF algorithm will be summarized in the following. Equation (19) is in principle only required for the derivation, since the transformed covariance matrix $\underline{\underline{\mathbf{T}}}(m)$ is directly given by the eigenvalue decomposition. The following equations constitute the TDAF algorithm

$$\underline{\mathbf{S}}(m) = \lambda \underline{\mathbf{S}}(m-1) + \frac{1}{2}(1-\lambda)\underline{\mathbf{X}}^H(m)\underline{\mathbf{X}}(m) , \quad (27a)$$

$$\underline{\underline{\mathbf{T}}}(m) = \underline{\mathbf{U}}^H(m)\mathbf{A}^T\underline{\mathbf{S}}(m)\mathbf{A}\underline{\mathbf{U}}(m) , \quad (27b)$$

$$\underline{\underline{\mathbf{K}}}(m) = (1-\lambda)\underline{\underline{\mathbf{T}}}^{-1}(m)\underline{\underline{\mathbf{X}}}^H(m) , \quad (27c)$$

$$\underline{\mathbf{e}}'(m) = \underline{\mathbf{y}}(m) - \mathbf{G}_2\underline{\underline{\mathbf{X}}}(m)\mathbf{G}_U\hat{\underline{\underline{\mathbf{h}}}}'(m-1) , \quad (27d)$$

$$\hat{\underline{\underline{\mathbf{h}}}}'(m) = \mathbf{G}_U\hat{\underline{\underline{\mathbf{h}}}}'(m-1) + \tilde{\mathbf{G}}_3\underline{\underline{\mathbf{K}}}(m)\underline{\mathbf{e}}'(m) , \quad (27e)$$

where

$$\mathbf{G}_U = \underline{\mathbf{U}}^H(m)\underline{\mathbf{U}}(m-1) . \quad (28)$$

The algorithm defined by Eq. (27) constitutes a combination of the concepts of FDAF and TDAF. The decoupling of the covariance matrix is performed in a two-step approach. The DFT is used for temporal decoupling and an eigenvalue decomposition for spatial decoupling. The temporal decoupling is performed in a very efficient manner by applying FDAF. The required DFTs can be realized efficiently by the fast Fourier transform (FFT). For an exact spatial decoupling, an eigenvalue decomposition has to be performed. However, the derived formulation allows also to use a generic MIMO filtering of the far-end signals with the potential of finding efficient approximations of the exact solution.
The following section will briefly discuss variants of the baseline algorithm.

5.2 Variants

The block-based TDAF algorithm (27) requires to compute the covariance matrix $\underline{\mathbf{S}}(m)$ in order to derive the transformation $\underline{\mathbf{U}}(m)$ of the far-end signals. An alternative is to derive the transformation directly from $\underline{\mathbf{X}}^H(m)\underline{\mathbf{X}}(m)$ and to formulate a recursive update of the decoupled covariance matrix $\underline{\underline{\mathbf{T}}}(m)$. The eigenvalue decomposition of $\underline{\mathbf{X}}^H(m)\underline{\mathbf{X}}(m)$ is given as

$$\underline{\mathbf{X}}^H(m)\underline{\mathbf{X}}(m) = \mathbf{A}\underline{\mathbf{U}}(m)\tilde{\underline{\underline{\mathbf{T}}}}(m)\underline{\mathbf{U}}^H(m)\mathbf{A}^T . \quad (29)$$

Hence, we can define the transformed far-end signals similar to the derivation of TDAF, as given in the previous section by (20). These transformed signals can then be introduced into the derivation of FDAF as given in [7]. The resulting transformed covariance matrix is given as

$$\underline{\underline{\mathbf{T}}}(m) = \lambda\mathbf{G}_U\underline{\underline{\mathbf{T}}}(m-1)\mathbf{G}_U^H + (1-\lambda)\tilde{\underline{\underline{\mathbf{T}}}}(m) , \quad (30)$$

which can be combined straightforwardly with (27c)-(27e). However, the matrix $\mathbf{G}_U$ is different in this case

$$\mathbf{G}_U = \underline{\mathbf{U}}^H(m)\mathbf{A}^T G^{10}_{2LP\times LP}\mathbf{B}\underline{\mathbf{U}}_L(m)\times \\ \times\underline{\mathbf{U}}_L(m-1)\mathbf{B}^T(G^{10}_{2LP\times LP})^H\mathbf{A}\underline{\mathbf{U}}(m-1) , \quad (31)$$

where the permutation matrix $\mathbf{B}$ is defined in [7]. The constraint matrix $\mathbf{G}_U$ for both variants of the algorithm considers the transition of the eigenspaces of $\underline{\mathbf{S}}(m)$ over time. The matrix $\mathbf{G}_U$ takes the spatial changes in the far-end signals into account. This is due to the two-step approach to spatio-temporal decoupling of the far-end signals applied in the presented TDAF approach. For spatially (quasi) stationary signals (i.e., $\underline{\mathbf{U}}(m) \approx \underline{\mathbf{U}}(m-1)$) this matrix can be approximated quite well by $\mathbf{G}_U = \mathbf{I}$ for both variants. Note, that under this approximation both variants of the block-based TDAF algorithm are equivalent. The approximation of $\mathbf{G}_U$ is also reasonable for situations where the forgetting factor λ is chosen close to one.

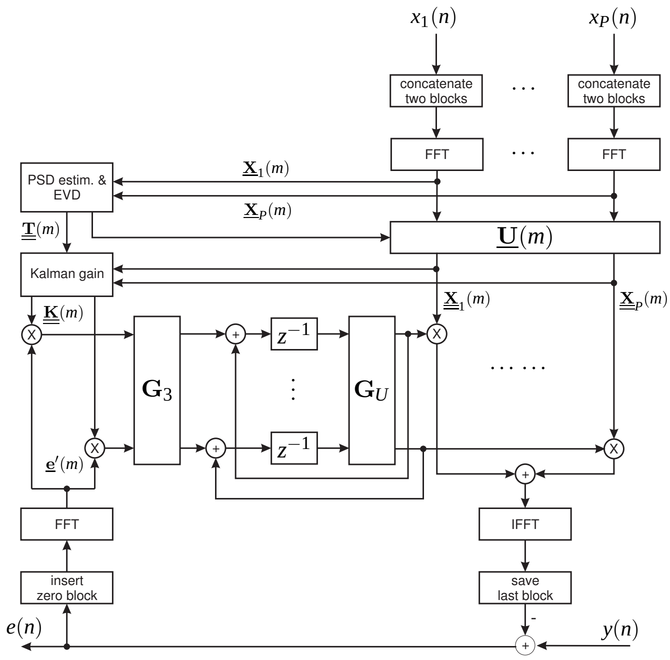


Figure 2: Block diagram of TDAF algorithm (adapted from [7]).

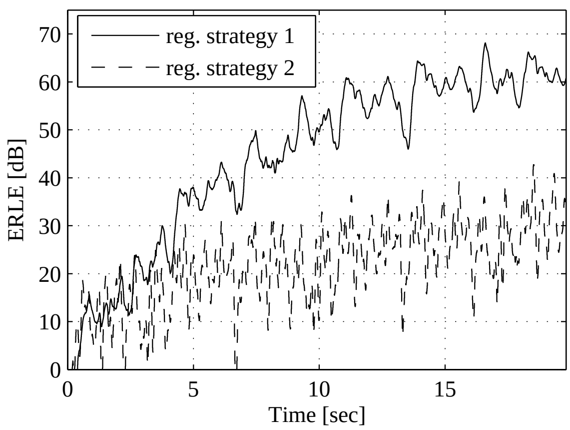
In theory, the number of eigenvalues for one frequency bin which are not zero is given by the number of independent active sources in the far-end room. The other eigenvalues are zero or in practice close to zero. Hence, these eigenvalues and the associated eigenchannels can be neglected for the adaptation. This technique is also known as reduced rank TDAF. The application of a so-called thin eigenvalue decomposition computing only some of the eigenvalues provides the potential to lower the computational complexity.
The derivation of the block-based TDAF algorithm is so far based on the assumption that the constraint $\mathbf{G}_1$ is approximated by $\mathbf{G}_1 = \mathbf{I}/2$. This assumption allows the bin-wise computation of the eigenvalue decomposition. However, the TDAF framework is also applicable when this constraint is not approximated. The singular vectors and values of the reordered covariance matrix have then to be computed by considering the entire $2LP \times 2LP$ matrix, which may be computationally very demanding in practice.

5.3 Implementation

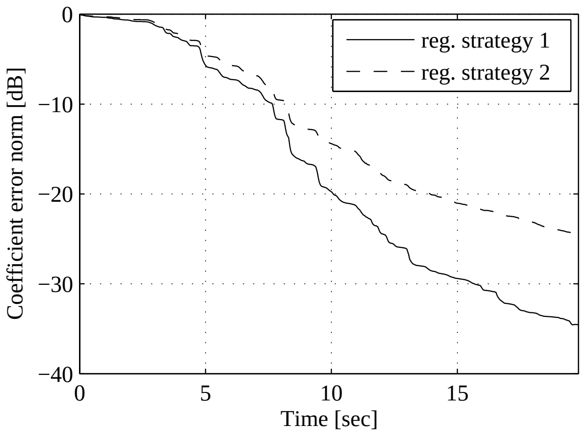
Some of the matrices in the formulation (27) of the TDAF algorithm exhibit sparse or diagonal structures giving the potential for optimization in a practical implementation. Figure 2 illustrates a block-diagram of the algorithm exploiting these structures. Due to the frequency-domain formulation provided by FDAF all operations can be performed efficiently in a bin-wise (scalar) fashion. However, the matrices $\underline{\mathbf{U}}(m)$ and the constraints $\mathbf{G}_3$ and $\mathbf{G}_U$ constitute MIMO systems. If the constraints $\mathbf{G}_3$ and $\mathbf{G}_U$ are approximated by identity matrices, then the multichannel identification problem is reduced to P decoupled SISO identification problems within the transform domain.

6. RESULTS

A typical multichannel AEC application scenario will be considered in the following to illustrate the properties of the developed TDAF algorithm.
The simulated geometrical setup consists of a near-end room with size $6 \times 6 \times 3$ meters containing two loudspeakers ($P = 2$) and one microphone. The near-end room was acoustically modeled by the image source method with an acoustic reflection factor at the walls of $\rho = 0.9$. The loudspeakers and the microphone are located at a



(a) Echo return loss enhancement (ERLE).



(b) Normalized misalignment.

Figure 3: Simulation results for the proposed multichannel TDAF algorithm for two different regularization strategies.

height of 1.5 meters. The position of the loudspeakers is $[2.8, 5]$ m and $[3.2, 5]$ m, and of the microphone $[5, 2]$ m.

The signal of a male speaker was fed equally to both loudspeakers (phantom source stationary in center). The loudspeaker signals were pre-processed by a nonlinearity [1] in order to cope for the non-uniqueness problem. Noise with a level of approximately -50 dB with respect to the echo was added to the microphone signals, in order to simulate microphone and other noise sources at the near-end.

The algorithm was implemented in MATLAB, as depicted by Fig. 2. The filter length was chosen as $L = 4096$ at a sampling rate of $f_s = 44.1$ kHz. In order to illustrate the effect of selective regularization in the eigenspace, two regularization strategies have been implemented: (1) both spatial and temporal frequency-bin selective regularization and (2) only temporal frequency-bin selective regularization. The latter shows a similar performance as a straightforward implementation of FDAF. The dynamic regularization scheme introduced in [6] has been used for both strategies.

Figure 3(a) shows the echo return loss enhancement (ERLE) for the simulated scenario. It can be seen that the algorithm converges fast and provides a good amount of echo attenuation (in dB) that is bounded by the near-end noise. The spatio-temporal frequency-bin selective regularization shows better results than the temporal frequency-bin regularization. Figure 3(b) shows the normalized misalignment. Again the spatio-temporal frequency-bin selective regularization performs better. Note, that the proposed TDAF algorithm, unlike multichannel FDAF, provides inherently the possibility for this beneficial regularization strategy.

7. CONCLUSION

This paper presents a block-based reformulation of the sample-by-sample multichannel TDAF approach introduced in [2]. Its two-stage approach to spatio-temporal decoupling has been exploited in order to perform the temporal decoupling efficiently by the FDAF algorithm in combination with an eigenvalue decomposition to cope for the spatial couplings. In contrast to a sample-by-sample update the presented block-based approach benefits from the computational savings of the FDAF algorithm. The results show that the resulting algorithm performs well in a typical multichannel AEC scenario. One benefit of the proposed adaptation scheme, working in the eigenspace of the far-end signal covariance matrix, is the possibility of selective regularization in that eigenspace. The benefit of this regularization was demonstrated in Section 6. The block-based TDAF algorithm is formally equivalent to wave-domain adaptive filtering (WDAF) developed by the authors in [7]. This link is quite interesting since WDAF is based on decoupling of the near-end system by a singular value decomposition (SVD). The only formal difference between the TDAF and the WDAF algorithm, besides the MISO/MIMO formulation, is the matrix $\mathbf{G}_U$ which accounts for the change of the eigenspace over time. Since for the derivation of WDAF (like for FDAF) it is assumed that the near-end room acoustics is time-invariant this matrix does not show up there explicitly. The formulations of WDAF and multichannel TDAF, as presented by the authors, are based on transforming (filtering) the far-end signals in order to overcome fundamental problems of the multichannel identification problem. The transformations are linked to the eigenspace of the near-end system or the covariance matrix of the far-end signals. The formulations of the algorithms are also applicable for generic MIMO filters. This opens up the potential to find efficient approximations of these transformations. The basic concept of TDAF also has a strong relation to blind source separation (BSS). The BSS algorithms based on second-order statistics try to find a demixing system that diagonalizes the covariance matrix of the demixed signals. The transformation $\underline{\mathbf{U}}$ can also be interpreted as demixing system in this context, since the goal is to provide independent signals to the adaptive filters.

In the future, further work is planned on the detailed analysis of the properties of the presented multichannel TDAF algorithm.

REFERENCES

[1] J. Benesty, Y. Huang, and J. Chen, "Wiener and adaptive filters," in *Speech and Audio Processing in Adverse Environments*, E. Haensler and G. Schmidt, Eds., pp. 103–120. Springer, 2008.

[2] S. Spors and H. Buchner, "Multichannel transform domain adaptive filtering: A two stage approach and illustration for acoustic echo cancellation," in *11th International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, September 2008.

[3] S.S. Narayan, A.M. Peterson, and M.J. Narasimha, "Transform domain LMS algorithm," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, June 1983, vol. ASSP-31.

[4] J.D Markel and A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, 1976.

[5] R.M. Gray, "On the asymptotic eigenvalue distribution of toeplitz matrices," *IEEE Transactions on Information Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[6] H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive algorithms with application to acoustic echo cancellation," in *Adaptive signal processing: Application to real-world problems*, J. Benesty and Y. Huang, Eds. Springer, 2003.

[7] H. Buchner and S. Spors, "A general derivation of wave-domain adaptive filtering and application to acoustic echo cancellation," in *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, USA, October 2008.