MULTICHANNEL TRANSFORM DOMAIN ADAPTIVE FILTERING: A TWO STAGE APPROACH AND ILLUSTRATION FOR ACOUSTIC ECHO CANCELATION

Sascha Spors and Herbert Buchner

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

ABSTRACT

System identification is the underlying mechanism of adaptive filtering algorithms in the context of acoustic echo cancelation systems. Multichannel system identification is subject to specific problems emerging from spatio-temporal couplings in the input signals of the adaptive filter. This paper presents a generic multichannel transform-domain adaptive filtering approach. It is based upon a two-stage decoupling of the spatio-temporal couplings. The temporal decoupling is yielded by a frequency domain transformation and the spatial decoupling by a further unitary transform. The presented approach reveals strong insights into the fundamental problems of multichannel identification.

Index Terms— multichannel identification, adaptive filtering, acoustic echo cancelation, condition number

1. INTRODUCTION

The problem of system identification plays an prominent role in the development of, e. g., acoustic telecommunication systems. There, a frequent application is acoustic echo cancelation (AEC). A full-duplex communication scenario where AEC is applied is depicted 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. Multichannel systems play an increasing role in teleconferencing applications where the communication is enriched by spatial audio. In this paper we consider systems with multiple speakers and one microphone at the near-end. The near-end room can therefore characterized as multiple-input/single-output (MISO) system. Multichannel system identification is a challenging problem since it is typically ill-conditioned [1]. For instance, the far-end microphone signals in Fig. 1 are produced by one source only and linear filtering by the room acoustics. Hence, the microphone signals are strongly coupled in such a scenario. In advanced adaptation schemes, at least two fundamental approaches exist to cope with these couplings: (1) decoupling of the MISO convolution in the near-end room and (2) decoupling of the input (microphone 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). While a number of multichannel FDAF algorithms have been published, e. g. [2], similar algorithms for TDAF, originally introduced for the single channel case [3], seem to be rare.

This paper presents a generic multichannel TDAF approach. It is based upon a two-stage decoupling of the covariance matrix, where the temporal decoupling is yielded by a frequency domain transformation and the spatial decoupling by a spatial TDAF approach.

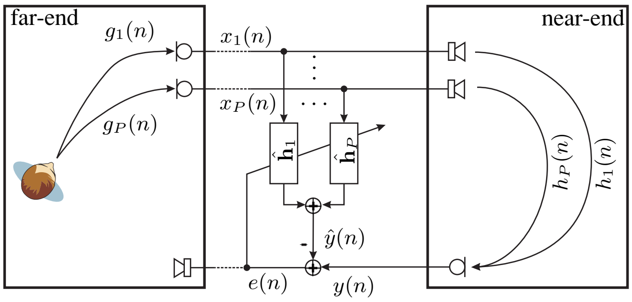


Fig. 1. Block diagram of multichannel acoustic echo cancelation.

Moreover, the illustration of the results shows that this approach allows strong insights into the fundamental problems of multichannel identification. The presented results can be applied to generic MISO identification problems.

2. MULTICHANNEL ACOUSTIC ECHO CANCELATION

In the block diagram of Fig. 1, the echo produced by the acoustic couplings between the P loudspeakers and the microphone at 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 filters that model the acoustic paths $h_p(n)$ from the loudspeakers to the microphones. These filters are derived adaptively from the loudspeaker signals $x_p(n)$ and the error signal $e(n) = y(n) - \hat{y}(n)$. It will be assumed in the following that no double talk is present.

The estimation of the acoustic paths $h_p(n)$ represents a multichannel identification problem. Under certain reasonable assumptions, adaptive algorithms aim to solve the following normal equation [1]

$$\mathbf{R}_{xx}\,\hat{\mathbf{h}} = \mathbf{r}_{xy}\ . \tag{1}$$

The $PL \times 1$ vector $\hat{\mathbf{h}}$ of estimated filter coefficients is given as

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

$$\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T, \hat{\mathbf{h}}_2^T, \ldots, \hat{\mathbf{h}}_P^T]^T\ , \tag{2b}$$

where $\hat{h}_{p,l}$ denotes the l-th coefficient of the p-th channel and L the filter length. 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

$$\mathbf{R}_{xx}(n) = \hat{\mathcal{E}}\{\mathbf{x}(n)\,\mathbf{x}^T(n)\} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1P} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{R}_{P1} & \dots & \dots & \mathbf{R}_{PP} \end{bmatrix} , \quad (3)$$

where $\hat{\mathcal{E}}\{\cdot\}$ denotes a suitable approximation of the expectation operator and n the time instant. The $PL \times 1$ vector $\mathbf{x}$ of input signals is given as

$$\mathbf{x}_p(n) = [x_p(n), x_p(n-1), \dots, x_p(n-L+1)]^T , \quad (4a)$$
$$\mathbf{x}(n) = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_P^T]^T . \quad (4b)$$

The $PL \times 1$ covariance vector $\mathbf{r}_{xy}(n) = \hat{\mathcal{E}}\{\mathbf{x}(n)\,y(n)\}$. The $L \times L$ sub-matrices $\mathbf{R}_{pq} = \hat{\mathcal{E}}\{\mathbf{x}_p(n)\mathbf{x}_q^T(n)\}$ of $\mathbf{R}_{xx}$ for $p, q = 1, 2, \dots, P$ are assumed to be Toeplitz matrices.
In the recursive least squares (RLS) algorithm the covariance matrix $\mathbf{R}_{xx}(n)$ is estimated as weighted sum [4]

$$\mathbf{R}_{xx}(n) = \sum_{\xi=0}^{n} \lambda^{n-\xi}\mathbf{x}(\xi)\,\mathbf{x}^T(\xi) , \quad (5)$$

where λ denotes the forgetting factor $0 < \lambda < 1$. The coefficient update for the RLS algorithm is given as

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) - \mathbf{R}_{xx}^{-1}(n)\mathbf{x}(n)e(n) . \quad (6)$$

The solution of the normal equation (1) is subject to the conditioning of the covariance matrix $\mathbf{R}_{xx}$. For typical application scenarios temporal and spatial couplings are present in the signal vector $\mathbf{x}(n)$. For the single-channel case, various techniques have been developed in the past to cope with time-domain couplings [4]. One of these is TDAF [3].

3. A TWO-STAGE APPROACH TO MULTICHANNEL TDAF

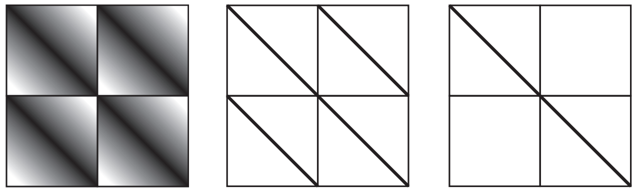
One goal of this paper is to develop a generic multichannel TDAF approach. This is yielded by deriving a fully diagonalized representation of the covariance matrix $\mathbf{R}_{xx}$ in two consecutive steps: (1) block-diagonalization by exploiting the Toeplitz/circulant structure of $\mathbf{R}_{pq}$ and (2) full-diagonalization by applying a unitary transform.

3.1. Temporal decoupling

As first step towards the desired decoupling, we will follow the established procedure of diagonalizing the sub-matrices $\mathbf{R}_{pq}$ of $\mathbf{R}_{xx}$ using a discrete Fourier transformation (DFT) [4]. The sub-matrices $\mathbf{R}_{pq}$ exhibit a Toeplitz structure. Toeplitz matrices are asymptotically equivalent to circulant matrices if their elements are absolutely summable [5]. Hence, for large block lengths ($L \to \infty$) the matrices $\mathbf{R}_{pq}$ become equivalent to circulant matrices. Circulant matrices can be diagonalized by a DFT. This can be formulated in matrix notation as

$$\mathbf{R}_{xx} = \mathbf{F}\,\underline{\mathbf{S}}_{xx}\,\mathbf{F}^H , \quad (7)$$

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



(a) $\mathbf{R}_{xx}$ (b) $\underline{\mathbf{S}}_{xx}$ (c) $\underline{\underline{\mathbf{T}}}_{xx}$

Fig. 2. Illustration of covariance matrix and its representations.

$n, m = 0, 1, \dots, L-1$. The block-matrix $\underline{\mathbf{S}}_{xx}$ is composed from the $L \times L$ diagonal matrices

$$\underline{\mathbf{S}}_{pq} = \mathrm{diag}\{\underline{s}_{pq}^{(0)}, \underline{s}_{pq}^{(1)}, \dots, \underline{s}_{pq}^{(L-1)}\} , \quad (8)$$

where the elements $\underline{s}_{pq}^{(\nu)}$ for $\nu = 0, 1, \dots, L-1$ are given by the DFT of the first column of $\mathbf{R}_{pq}$. The frequency bin is denoted as ν. Note, that Eq. (7) can be understood as a unitary transform of the covariance matrix $\mathbf{R}_{xx}$.
The achieved frequency domain representation of the covariance matrix $\mathbf{R}_{xx}$ results only in a diagonalization of the blocks $\mathbf{R}_{pq}$. This is illustrated in Fig. 2. Figure 2(a) shows the typical structure of the covariance matrix $\mathbf{R}_{xx}$ for $P = 2$. The transformation by the DFT matrices results in the block-diagonalized matrix $\underline{\mathbf{S}}_{xx}$, as illustrated by Fig. 2(b).
For a full diagonalization, the spatial couplings between the P channels will be exploited by applying a further unitary transform to the covariance matrix.

3.2. Spatial decoupling

The diagonal matrices $\underline{\mathbf{S}}_{pq}$ are composed from the elements $\underline{s}_{pq}^{(\nu)}$ for one particular combination of input signals. In order to achieve the desired decoupling, the matrix $\underline{\mathbf{S}}_{xx}$ is rearranged such that all combinations $\underline{s}_{pq}^{(\nu)}$ for one particular frequency bin ν are combined into one $P \times P$ matrix $\underline{\mathbf{S}}^{(\nu)}$. These blocks are then arranged on the main diagonal of the matrix $\tilde{\underline{\mathbf{S}}}_{xx}$. This operation can be described by a suitably chosen permutation matrix $\mathbf{A}$ [6]

$$\underline{\mathbf{S}}_{xx} = \mathbf{A}\,\tilde{\underline{\mathbf{S}}}_{xx}\,\mathbf{A}^T , \quad (9)$$

where $\tilde{\underline{\mathbf{S}}}_{xx}$ is a $LP \times PL$ block-diagonal matrix composed from the blocks $\underline{\mathbf{S}}^{(\nu)}$. The elements of $\underline{\mathbf{S}}^{(\nu)}$ exhibit the specific symmetry property $\underline{s}_{pq}^{(\nu)} = (\underline{s}_{qp}^{(\nu)})^*$ due to their interpretation as power spectral densities. Consequently, the matrices $\underline{\mathbf{S}}^{(\nu)}$ are normal and the spectral theorem can be applied. Hence, $\underline{\mathbf{S}}^{(\nu)}$ can be diagonalized by a unitary transform [6]

$$\underline{\mathbf{S}}^{(\nu)} = \underline{\mathbf{U}}^{(\nu)}\,\underline{\underline{\mathbf{T}}}^{(\nu)}\,(\underline{\mathbf{U}}^{(\nu)})^H , \quad (10)$$

where $\underline{\mathbf{U}}^{(\nu)}$ denotes a $P \times P$ unitary matrix composed from the singular vectors of $\underline{\mathbf{S}}^{(\nu)}$. Spatially transformed quantities are underlined twice. The matrix $\underline{\underline{\mathbf{T}}}^{(\nu)}$ denotes a diagonal matrix constructed from the singular values $\underline{\underline{t}}_{\eta}^{(\nu)}$ of $\underline{\mathbf{S}}^{(\nu)}$

$$\underline{\underline{\mathbf{T}}}^{(\nu)} = \mathrm{diag}\{\underline{\underline{t}}_1^{(\nu)}, \underline{\underline{t}}_2^{(\nu)}, \cdots, \underline{\underline{t}}_P^{(\nu)}\} . \quad (11)$$

Some of the eigenvalues might be zero (or in practice very small) if the matrix $\underline{\mathbf{S}}^{(\nu)}$ exhibits a rank deficit.
Constructing a block-diagonal matrix $\underline{\mathbf{U}}$ from the unitary matrices

$\underline{\mathbf{U}}^{(\nu)}$ for all ν and similarily a block-diagonal matrix $\underline{\underline{\mathbf{T}}}_{xx}$ with the diagonal matrices $\underline{\underline{\mathbf{T}}}^{(\nu)}$ yields

$$\tilde{\underline{\mathbf{S}}}_{xx} = \underline{\mathbf{U}}\,\underline{\underline{\mathbf{T}}}_{xx}\,\underline{\mathbf{U}}^H \; , \tag{12}$$

where $\underline{\underline{\mathbf{T}}}_{xx}$ is a diagonal matrix constructed from all the eigenvalues $\underline{\underline{t}}_{\eta}^{(\nu)}$. Substituting (9) and (12) into (7) yields the desired representation of the covariance matrix

$$\mathbf{R}_{xx} = \mathbf{F}\mathbf{A}\underline{\mathbf{U}}\,\underline{\underline{\mathbf{T}}}_{xx}\,\underline{\mathbf{U}}^H\mathbf{A}^T\mathbf{F}^H \tag{13}$$

in terms of the diagonal matrix $\underline{\underline{\mathbf{T}}}_{xx}$. Hence, the desired decoupling of the covariance matrix has been achieved by a set of suitably chosen unitary transforms. These transforms consist of two steps: (1) temporal decoupling using a DFT based transformation and (2) a spatial decoupling using a unitary transform. Figure 2 illustrates the effect of these two transformations on the covariance matrix.

3.3. Multichannel TDAF

Introducing Eq. (13) into the normal equation (1) and exploiting the unitarity of the transform matrices yields the transformed normal equation

$$\underline{\underline{\mathbf{T}}}_{xx}\,\underbrace{\underline{\mathbf{U}}^H\mathbf{A}^T\mathbf{F}^H\hat{\mathbf{h}}}_{\hat{\underline{\underline{\mathbf{h}}}}} = \underbrace{\underline{\mathbf{U}}^H\mathbf{A}^T\mathbf{F}^H\mathbf{r}_{xy}}_{\underline{\underline{\mathbf{t}}}_{xy}} \; , \tag{14}$$

where $\hat{\underline{\underline{\mathbf{h}}}}$ 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 (1) has been decomposed by the transformations into a series of scalar equations.
The application of the transformations will be illustrated briefly at the example of the RLS algorithm. The coefficient update of the multichannel TDAF approach can be derived straightforwardly by introducing Eq. (13) into Eq. (6) as

$$\hat{\underline{\underline{\mathbf{h}}}}(n) = \hat{\underline{\underline{\mathbf{h}}}}(n-1) - \underline{\underline{\mathbf{T}}}_{xx}^{-1}(n)\,\underbrace{\underline{\mathbf{U}}^H\mathbf{A}^T\mathbf{F}^H\mathbf{x}(n)}_{\underline{\underline{\mathbf{x}}}(n)}\,e(n) \; , \tag{15}$$

where $\underline{\underline{\mathbf{x}}}(n)$ denotes the transformed vector of input signals $\mathbf{x}(n)$. The covariance matrix $\underline{\underline{\mathbf{T}}}_{xx}(n)$ can be estimated from the transformed input signals by introducing Eq. (13) into (5) and rearranging the sum as

$$\underline{\underline{\mathbf{T}}}_{xx}(n) = \lambda\,\underline{\underline{\mathbf{T}}}_{xx}(n-1) + \underline{\underline{\mathbf{x}}}(n)\,\underline{\underline{\mathbf{x}}}^T(n) \; . \tag{16}$$

Equations (15) and (16) reveal that the filter coefficients $\hat{\underline{\underline{\mathbf{h}}}}(n)$ can be computed from the transformed input $\underline{\underline{\mathbf{x}}}(n)$ and error $e(n)$ signal if the transformation matrices are known. The temporal transformation $\mathbf{F}$ is given by the DFT matrix, and hence fixed. The required spatial transformation $\underline{\mathbf{U}}$ is composed from the eigenvectors of $\underline{\mathbf{S}}^{(\nu)}$, and hence depends on the spatial properties of the far-end acoustics. The eigenvectors can be computed from the covariance matrix $\mathbf{R}_{xx}$ and hence from the input signals $\mathbf{x}(n)$. The computational complexity for the eigenvector calculation can be quite high. However, the block structure of the covariance matrix allows to decompose the problem into a series of lower dimensional problems. Additionally, the sparse structure of the transformed covariance matrix can be exploited in practical implementations. This indicates that the computation of the transformations is tractable in practice. Further, a fixed transformation derived from analyzing the eigenvectors, like in single-channel TDAF algorithms [3], might be a potential solution.

The solution of the normal equation (14) and the coefficient update (15) require to invert the covariance matrix $\underline{\underline{\mathbf{T}}}_{xx}(n)$. If the eigenvalues $\underline{\underline{t}}_{\eta}^{(\nu)}$ are close to zero this will be subject to numerical problems. Potential countermeasures are to perform a regularization of the respective values or to discard the respective bins in the computation of the filter update. For the single-channel case, the latter is known as reduced rank TDAF. In practice, microphone noise and the impulse response tail effect will limit the numerical problems [1].
Note that the inverse of the transformed covariance matrix, occurring in Eq. (15), can be understood as the power normalization step typically applied in single channel LMS-based TDAF algorithms [3].

4. ILLUSTRATION

4.1. Conditioning and misalignment

It was already outlined in Section 2, that the computation of the filter coefficients essentially requires to invert the covariance matrix $\mathbf{R}_{xx}$. The condition number of this matrix is a measure how numerically well-posed this inversion process is. If the condition number is high, the problem is said to be ill-conditioned. Small variations in the covariance matrix lead to large variations in the filter coefficients. The condition number of a matrix depends on the underlying matrix norm. The Frobenius norm of $\mathbf{R}_{xx}^{1/2}$ has proven to be a good measure in the context of adaptive filtering [7]. This section derives the conditioning of the covariance matrix in terms of its transformed representation. It will be assumed that $\mathbf{R}_{xx}$ has full rank for ease of illustration.
The condition number $\kappa_{\mathrm{F}}\{\mathbf{R}_{xx}^{1/2}\}$ of $\mathbf{R}_{xx}^{1/2}$ is given as [6]

$$\kappa_{\mathrm{F}}\{\mathbf{R}_{xx}^{\frac{1}{2}}\} = \left\|\mathbf{R}_{xx}^{-\frac{1}{2}}\right\|_{\mathrm{F}} \left\|\mathbf{R}_{xx}^{\frac{1}{2}}\right\|_{\mathrm{F}} \; , \tag{17}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. The Frobenius norm of $\mathbf{R}_{xx}^{1/2}$ can be expressed in terms of the following matrix traces

$$\left\|\mathbf{R}_{xx}^{\frac{1}{2}}\right\|_{\mathrm{F}} = \mathrm{tr}\{\mathbf{R}_{xx}\}^{\frac{1}{2}} = \mathrm{tr}\{\underline{\underline{\mathbf{T}}}_{xx}\}^{\frac{1}{2}} \; , \tag{18}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a matrix. Note that the invariance of the trace operator with respect to unitary transformations has been exploited to yield the last equality. Expressing the Frobenius norm of $\mathbf{R}_{xx}^{-1/2}$ in a similar fashion, and introducing both into Eq. (17) yields

$$\kappa_{\mathrm{F}}^2\{\mathbf{R}_{xx}^{\frac{1}{2}}\} = \sigma_x^2\,\sum_{\eta=1}^{P}\sum_{\nu=0}^{L-1}(\underline{\underline{t}}_{\eta}^{(\nu)})^{-1} \; , \tag{19}$$

with $\sigma_x^2 = \sum_{\eta=1}^{P}\sum_{\nu=0}^{L-1}\underline{\underline{t}}_{\eta}^{(\nu)}$ being the variance of the input signals.
Equation (19) illustrates that the conditioning of $\mathbf{R}_{xx}^{1/2}$ is given by summing up the eigenvalues $\underline{\underline{t}}_{\eta}^{(\nu)}$ and its inverses $(\underline{\underline{t}}_{\eta}^{(\nu)})^{-1}$ for all frequencies ν and dimensions η. It is evident from (19) that the condition number gets high if one or more of the eigenvalues $\underline{\underline{t}}_{\eta}^{(\nu)}$ are close to zero.
The condition number can be related to the normalized misalignment μ_{min} by introducing Eq. (19) into [2, Eq.(80)]

$$\mu_{\mathrm{min}} = 10\log_{10}\left(\frac{\sigma_n^2}{\sigma_x^2}\,\frac{(1-\lambda)^2}{\|\mathbf{h}\|_2^2}\,\kappa_{\mathrm{F}}^2\{\mathbf{R}_{xx}^{\frac{1}{2}}\}\right) \tag{20}$$

where σ_n^2 denotes the variance of the additive noise.
Due to the decoupling of the proposed multichannel TDAF approach

small and well-posed problems occur (no summations in Eq. (19)). However, selective regularization is needed, as already mentioned above.

4.2. Interpretation of eigenvalues

In order to gain more insight into the coefficients $\underline{\underline{t}}_\eta^{(\nu)}$ these are interpreted in the following. The covariance matrix $\mathbf{R}_{pq}$ expressing the spatio-temporal couplings between the p-th and q-th input channel can be expressed by rewriting (7) as sum

$$\mathbf{R}_{pq} = \sum_{\nu=0}^{L-1} \underline{s}_{pq}^{(\nu)} \, \mathbf{f}^{(\nu)} (\mathbf{f}^{(\nu)})^H \, , \tag{21}$$

where $\mathbf{f}^{(\nu)}$ denotes a $L \times 1$ vector containing the ν-th column of the DFT matrix. Consequently, the elements of $\mathbf{f}^{(\nu)}$ are given as $e^{-j2\pi l\nu/L}$ with $l = 0, 1, \ldots, L-1$. The first column of the covariance matrix $\mathbf{R}_{pq}$ contains the temporal covariance coefficients $r_{pq}(l)$ between the p-th and q-th input channel. Simplifying (21) to the first column yields

$$r_{pq}(l) = \sum_{\nu=0}^{L-1} \underline{s}_{pq}^{(\nu)} \, e^{-j2\pi l\nu/L} \, . \tag{22}$$

Hence, the spectral coefficients $\underline{s}_{pq}^{(\nu)}$ represent the temporal couplings between the channels. The spatial couplings between the input signals can be derived by rewriting (10) as sum

$$\underline{\mathbf{S}}^{(\nu)} = \sum_{\eta=1}^{P} \underline{\underline{t}}_\eta^{(\nu)} \underline{\mathbf{u}}_\eta^{(\nu)} (\underline{\mathbf{u}}_\eta^{(\nu)})^H \, , \tag{23}$$

where $\underline{\mathbf{u}}_\eta^{(\nu)}$ denotes the η-th column of $\underline{\mathbf{U}}^{(\nu)}$. Hence, the coefficients $\underline{\underline{t}}_\eta^{(\nu)}$ represent the spatio-temporal couplings between the channels. Note, that the two step approach to decouple $\mathbf{R}_{xx}$, presented in this paper, allows to interpret the eigenvalues $\underline{\underline{t}}_\eta^{(\nu)}$ in terms of the spatio-temporal correlation between the input signals. For highly spatio-temporally correlated input signals most of the eigenvalues are close to zero and consequently the covariance matrix $\mathbf{R}_{xx}$ is ill-conditioned.

4.3. Link to generalized coherence

The magnitude-squared coherence $|\gamma(\nu)|^2$ is a frequently applied measure for the dependencies between two signals. For instance in [7] is was used to characterize the misalignment of stereophonic AEC. The concept of the coherence has been generalized to the case of more than two signals in [8]. In the following the link between the generalized coherence and the condition number is illustrated.
The generalized coherence, in terms of the quantities introduced in this paper, is given as

$$|\gamma(\nu)|^2 = 1 - \frac{\det\{\underline{\mathbf{S}}^{(\nu)}\}}{\prod_{\eta=1}^{P} \underline{s}_{\eta\eta}^{(\nu)}} = 1 - \prod_{\eta=1}^{P} \frac{\underline{\underline{t}}_\eta^{(\nu)}}{\underline{s}_{\eta\eta}^{(\nu)}} \, , \tag{24}$$

where the second equality was derived from the properties of the determinant. The generalized coherence $|\gamma(\nu)|^2$ is zero if all input signals are mutually independent from each other and will approach one if linear dependencies between the signals exist. It is assumed in the following, that the inputs $\mathbf{x}$ are excited with unit variance white (Gaussian) noise. Hence, that $\underline{s}_{\eta\eta}^{(\nu)} = 1$. Rearranging (24), taking the natural logarithm on both sides and using the approximation $\ln x \approx x - 1$ results in

$$\sum_{\eta=1}^{P} (\underline{\underline{t}}_\eta^{(\nu)})^{-1} \approx P - \ln(1 - |\gamma|^2) \, . \tag{25}$$

Introducing (25) into (19) together with $\sigma_x^2 = LP$ for the assumed excitation yields the desired relation as

$$\kappa_{\mathrm{F}}^2\{\mathbf{R}_{xx}^{\frac{1}{2}}\} \approx L^2P^2 - L^2P\ln(1 - |\gamma|^2) \, . \tag{26}$$

Equation (26) states that κ_{F}^2 assumes its minimum value L^2P^2 if $|\gamma(\nu)|^2 = 0$, hence when the input signals are independent from each other. Its also evident from (26) that κ_{F}^2 increases towards infinity when the generalized coherence increases towards one. Both extrema are in conjunction with Eq. (19) for such conditions. This indicates that the assumptions made to derive (26) are reasonable.

5. CONCLUSION

We presented a multichannel TDAF approach that is based upon a two step decoupling of the input covariance matrix using unitary transformations. These transformations depend on the spatio-temporal couplings of the far-end microphone signals. Since the room is assumed to be a linear slowly time-variant system, a Fourier transformation serves well for temporal decoupling. For the spatial decoupling a SVD has been applied. For an efficient implementation the spatial transformation could be applied in the multichannel FDAF framework. As a side effect, the proposed two step approach provides interesting insights into the ill-conditioning of the problem for the multichannel case and potential countermeasures.

6. REFERENCES

[1] J. Benesty, D.R. Morgan, and M.M. Sondhi, “A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation,” *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, March 1998.

[2] H. Buchner, J. Benesty, and W. Kellermann, “Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication,” *Signal Processing*, vol. 85, pp. 549–570, 2005.

[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] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 2002.

[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] G.H. Golub and C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1989.

[7] A.W.H. Khong, J. Benesty, and P.A. Naylor, “Stereophonic acoustic echo cancellation: Analysis of the misalignment in the frequency domain,” *IEEE Signal Processing Letters*, vol. 13, no. 1, pp. 33–36, January 2006.

[8] H. Gish and D. Cochran, “Generalized coherence,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New York, USA, April 1988.