Enhanced Partitioned Stereo Residual Echo Estimation

Stefan Goetze and Karl-Dirk Kammeyer
University of Bremen,
Dept. of Communications Engineering,
D-28334 Bremen, Germany
Email: goetze@ant.uni-bremen.de

Markus Kallinger and Alfred Mertins
Carl von Ossietzky University Oldenburg,
Signal Processing Group,
D-26111 Oldenburg, Germany
Email: markus.kallinger@uni-oldenburg.de

***Abstract*— For convenient hands-free telecommunication systems echo cancelers with a low delay are desirable. Since such systems have to cope with long room impulse responses (RIRs) common echo cancelation techniques often suffer from slow convergence or introduce large delays. An algorithm for acoustic echo cancelation (AEC) has to compensate the influence of RIRs of several thousand sampling points. Thus for frequency domain algorithms long DFT-lengths are needed to cover the whole influence of the RIRs. Approaches in the partitioned frequency domain overcome this drawback. Our contribution describes a combined echo cancelation system in the partitioned frequency domain containing a conventional stereo AEC enhanced by a post-filter which requires a reliable residual echo estimate. This contribution focuses on minimizing the estimation bias introduced by partitioned residual echo estimation. For this purpose an optimal smoothing factor for the residual echo differential system will be derived.**

I. Introduction

In hands-free video-conferencing-systems the signal of the far speaker $S_f[m,l]$ is transmitted to the receiving room and then picked up by the microphones due to the acoustic coupling between loudspeakers and microphones. Thus it has to be compensated to prevent it from being transmitted back to the far end speaker. The system model is depicted in Fig. 1. All signals in this contribution are defined in the partitioned discrete frequency domain with a discrete frequency index m and a discrete block index l. The index of the channel is denoted by $i = \{0, 1\}$. The echo signals $\Psi_i[m,l]$ are attenuated by the AEC-filters $\mathbf{C}_i[m,l]$ which calculate an estimate of the echoes $\hat{\Psi}_i[m,l]$ contained in the microphone path. The post-filter $P[m,l]$ is a common enhancement technique to support the AECs which cover only a finite length of the RIRs [1]. Thus some residual echo $\Xi[m,l]$ remains after the AECs [2], [3] and has to be canceled by the post-filter. It should be noted that only one half of the stereo system is depicted in Fig. 1 since the system for the second microphone channel is equivalent and independent of system for the first microphone channel. Thus only one microphone channel is considered here.

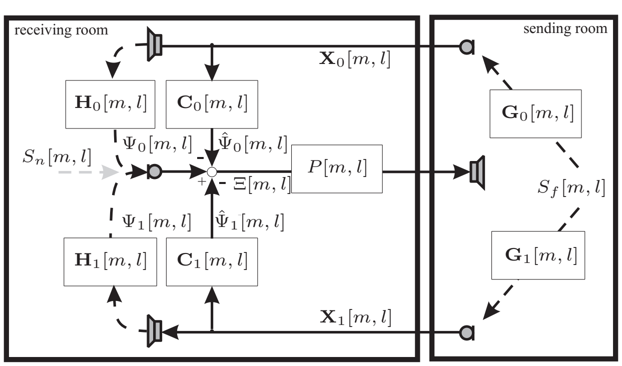

Fig. 1. Acoustic Echo Canceler (AEC) with a post-filter in a stereo setup

The remainder of this paper is organized as follows. Section II briefly introduces the partitioned frequency domain framework used in the following sections. Furthermore the estimation of the residual echo remaining after the acoustic echo canceller (AEC) is described for the stereo case. In Section III we derive an optimal smoothing factor for the estimation of the residual echo, that minimizes the estimation bias. Simulation results are presented in Section IV and Section V concludes this contribution.

II. Residual Echo Estimation in the partitioned frequency domain

Acoustic echo cancellers have to deal with very long impulse responses because of the high reverberation times of common office environments [1]. If long impulse responses have to be taken into account, either the processed blocks have to be of sufficient length to cover the whole impulse response or a partitioned calculation has to be applied. The block length directly influences the delay of the system, which should be kept as small as possible. On the other hand an estimation bias is introduced by an increased number of shorter blocks [4], [2].

With the definition of the short-time spectrum of the signal $x[k]$ and the system $h[k]$ with 50 % overlapping

$$X[m,l] = \sum_{k=0}^{L_{\mathrm{DFT}}/2-1} x\left[l \cdot \frac{L_{\mathrm{DFT}}}{2} + k\right] \cdot e^{-j\,2\pi \frac{m}{L_{\mathrm{DFT}}} k} \qquad (1)$$

and

$$H[m] = \sum_{k=0}^{L_{\mathrm{DFT}}/2-1} h[k] \cdot e^{-j\,2\pi \frac{m}{L_{\mathrm{DFT}}} k} \qquad (2)$$

a partitioned fast convolution using the overlap-save method [5] can be defined as

$$\tilde{\Psi}[m,l] = H[m] \cdot [1,\ (-1)^m] \cdot \begin{bmatrix} X[m,l] \\ X[m,l-1] \end{bmatrix} \qquad (3)$$

for every frequency bin $m = 0, 1, ..., L_{\mathrm{DFT}} - 1$ with the DFT-length L_{DFT}. The factor $(-1)^m = e^{-j\pi m}$ in (3) represents a time-shift of $L_{\mathrm{DFT}}/2$ samples for the previous block $X[m,l-1]$. The result of (3) leads to a convolution in the time-domain, which contains a linear part and a cyclic part. The cyclic part can be removed by an operator $\mathrm{F}_c\{\cdot\}$, which transforms the resulting spectrum back into the time domain, removes the cyclic part by setting the last $L_{\mathrm{DFT}}/2$ samples to zero and transforms the time series back into the DFT-domain [2], [3]:

$$\begin{aligned} \mathbf{\Psi}[l] &= \mathbf{F}\mathbf{W}_{10}\mathbf{F}^{-1}\tilde{\mathbf{\Psi}}[l] \\ &= \mathrm{F}_c\left\{\tilde{\mathbf{\Psi}}[l]\right\} \end{aligned} \qquad (4)$$

1-4244-0785-0/06/$20.00

with

$$\mathbf{\Psi}[l] = [\Psi[0,l],\ \Psi[1,l],\ ...,\ \Psi[L_{\mathrm{DFT}}-1,l]]^T \tag{5}$$

$$\tilde{\mathbf{\Psi}}[l] = \left[\tilde{\Psi}[0,l],\ \tilde{\Psi}[1,l],\ ...,\ \tilde{\Psi}[L_{\mathrm{DFT}}-1,l]\right]^T \tag{6}$$

$$\mathbf{W}_{10} = \begin{bmatrix} \mathbf{I}_{L_{\mathrm{DFT}/2}\times L_{\mathrm{DFT}/2}} & \mathbf{0}_{L_{\mathrm{DFT}/2}\times L_{\mathrm{DFT}/2}} \\ \mathbf{0}_{L_{\mathrm{DFT}/2}\times L_{\mathrm{DFT}/2}} & \mathbf{0}_{L_{\mathrm{DFT}/2}\times L_{\mathrm{DFT}/2}} \end{bmatrix} \tag{7}$$

$\mathbf{F}$ and $\mathbf{F}^{-1}$ are the DFT- and IDFT-matrices of size $L_{\mathrm{DFT}} \times L_{\mathrm{DFT}}$, respectively. $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ the zero matrix.

If the observed length of the room impulse response (RIR) $\mathbf{H}[m,l]$ is greater than the DFT-length we define the partitioned frequency domain system function as

$$\mathbf{H}[m] = \left[H_0[m],\ H_1[m],\ ...,\ H_{L'_H-1}[m]\right]^T \tag{8}$$

with

$$H_i[m] = \sum_{k=0}^{L_{\mathrm{DFT}}/2-1} h[k+iL_{\mathrm{DFT}}/2]\cdot e^{-j\,2\pi\frac{m}{L_{\mathrm{DFT}}}k}. \tag{9}$$

$L_H = L_{\mathrm{DFT}}\cdot L'_H$ is the length of the RIR which is taken into account. Although the RIR has an infinite length in general it can be assumed to be sufficiently decayed after L_H samples. The vector $\mathbf{X}_{L'_H+1}[m,l]$ is defined accordingly by concatenating L'_H+1 partitions of the loudspeaker signal

$$\mathbf{X}_{L'_H+1}[m,l] = \left[X[m,l],\ X[m,l-1],\ ...,\ X[m,l-L'_H]\right]^T \tag{10}$$

and thus the constrained convolution leads to

$$\Psi[\cdot,l] = \mathrm{F}_c\left\{\mathbf{H}^T[m]\mathbf{T}^T_{L'_H}\mathbf{X}_{L'_H+1}[m,l]\right\} \tag{11}$$

with

$$\mathbf{T}_{L'_H}[m] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ (-1)^m & 1 & \ddots & \vdots \\ 0 & (-1)^m & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & (-1)^m \end{bmatrix}_{L'_H+1\times L'_H}. \tag{12}$$

It should be mentioned, that the dependance of the discrete frequency index m is omitted in this paper for the reason of readability for the matrix $\mathbf{T}_{L'_H}[m]$. The matrix (12) is an extension of the middle part of equation (3), which selects two adjacent spectra at any one block time. The replacement of the discrete frequency index m by $\cdot$ in (11) is done due to the fact, that the result of the $\mathrm{F}_c\{\}$-operator is no longer independent of m.

The formalisms described above can be used for unbiased spectral estimation [6], [3].

$$\tilde{\Phi}_{XX}[\cdot,l] = \frac{2}{L_{\mathrm{DFT}}}\mathrm{F}_c\left\{[1,(-1)^m]\begin{bmatrix} X^*[m,l] \\ X^*[m,l-1]\end{bmatrix} X[m,l]\right\} \tag{13}$$

$$= \mathrm{F}_c\left\{\mathbf{T}_1^T\begin{bmatrix} X^*[m,l] \\ X^*[m,l-1]\end{bmatrix} X[m,l]\right\} \tag{14}$$

To reduce the variance of the estimation two further steps of processing can be applied. The first possibility is to take more blocks into account for the spectral estimation:

$$\bar{\Phi}_{XX}[\cdot,l] = \frac{2}{N\cdot L_{\mathrm{DFT}}}\mathrm{F}_c\left\{\begin{bmatrix} X[m,l] \\ \vdots \\ X[m,l-N+1]\end{bmatrix}^T \mathbf{T}_N^T \begin{bmatrix} X^*[m,l] \\ \vdots \\ X^*[m,l-N]\end{bmatrix}\right\}$$

$$= \mathrm{F}_c\left\{\mathbf{X}_N[m,l]\mathbf{T}_N^T\mathbf{X}^*_{N+1}[m,l]\right\} \tag{15}$$

A large number of considered blocks N reduces the variance but slows down the adaptation. Furthermore speech signals are only short-time stationary (up to about 20ms). Thus a tradeoff between variance reduction and nonstationary conditions has to be found. For further reducing the variance a first-order recursive smoothing with a smoothing factor $0 \le \beta \le 1$ can be applied.

$$\hat{\Phi}_{XX}[m,l] = \beta\hat{\Phi}_{XX}[m,l-1] + (1-\beta)\bar{\Phi}_{XX}[m,l] \tag{16}$$

A. Estimation of the system misalignment for the stereo case

In practical environments the RIRs are of infinite length and thus longer than the AEC-filters. The AEC-filters $\mathbf{C}_i[m,l]$ only compensate the first part of the impulse responses $\mathbf{H}_i[m,l]$ and a residual echo $\Xi[m,l]$ remains in the microphone path.

A reliable estimate of the power spectral density (PSD) of the residual echo $\hat{\Phi}_{\Xi\Xi}[m,l]$ is indispensable for the design of the post-filter. Under-estimation of $\hat{\Phi}_{\Xi\Xi}[m,l]$ leads to an insufficient residual echo attenuation and an over-estimation leads to signal cancellation and signal distortion. The residual echo's PSD can not be measured directly and thus has to be estimated. In this contribution an estimation of the residual echo PSD by estimating the system misalignment is proposed which was analyzed in [4] for a single-channel residual echo estimation. In Section II-A an extension to the stereo case as an example for a multi-channel system is presented and an optimal smoothing factor for the reduction of the bias introduced by the partitioned estimation is derived for the stereo case.
The stereo system misalignment is given by

$$\mathbf{D}_{\mathrm{st}}[m,l] = \left[\mathbf{D}_0^T[m,l],\ \mathbf{D}_1^T[m,l]\right]^T \tag{17}$$

$$= \left[\mathbf{H}^T_{0,\mathrm{c}}[m,l]-\mathbf{C}_0^T[m,l],\ \mathbf{H}^T_{0,\mathrm{t}}[m,l],\ \mathbf{H}^T_{1,\mathrm{c}}[m,l]-\mathbf{C}_1^T[m,l],\ \mathbf{H}^T_{1,\mathrm{t}}[m,l]\right]^T \tag{18}$$

with

$$\mathbf{H}_0[m,l] = \left[\mathbf{H}^T_{0,\mathrm{c}}[m,l],\mathbf{H}^T_{0,\mathrm{t}}[m,l]\right]^T \tag{19}$$

$$\mathbf{H}_{0,\mathrm{c}}[m,l] = \left[H_{0,0}[m,l],\cdots,H_{0,L'_C-1}[m,l]\right]^T \tag{20}$$

$$\mathbf{H}_{0,\mathrm{t}}[m,l] = \left[H_{0,L'_C}[m,l],\cdots,H_{0,L'_H-1}[m,l]\right]^T \tag{21}$$

$$\mathbf{C}_0[m,l] = \left[C_{0,0}[m,l],\ \cdots,\ C_{0,L'_C-1}[m,l]\right]^T \tag{22}$$

Equation (19) follows the definition of [7] by dividing the RIR into the part $\mathbf{H}_{0,\mathrm{c}}[m,l]$ which can be compensated by the AEC and a tail $\mathbf{H}_{0,\mathrm{t}}[m,l]$ which can not be compensated by the AEC due to the limited length of the AEC. Since in practical cases $L_H > L_C$ only the first part of $\mathbf{H}[m,l]$ can be compensated by $\mathbf{C}[m,l]$ and the residual echo is

$$\Xi[\cdot,l] = \mathrm{F}_c\left\{\mathbf{D}^T_{\mathrm{st}}[m,l]\mathbf{T}^T_{\mathrm{st},L'_d}\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right\} \tag{23}$$

with

$$\mathbf{T}_{\mathrm{st},L'_d}[m] = \begin{bmatrix} \mathbf{T}_{L'_d}[m] & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{L'_d}[m]\end{bmatrix} \tag{24}$$

$$\mathbf{X}_{\mathrm{st},L'_D+1}[m,l] = \left[X_0[m,l],...,X_0[m,l-L'_D],\ X_1[m,l],...,X_1[m,l-L'_D]\right]^T. \tag{25}$$

$L_D = L_{\mathrm{DFT}}\cdot L'_D$ is the length of the system misalignment which is taken into account for each channel. Without loss of generality we set $L'_D = L'_H$ as the length where the impulse response $h[k]$ is sufficiently decayed.

An estimate for a residual echo differential system (18) is obtained by the frequency domain coefficient vector

$$\mathbf{B}_{\mathrm{st}}[m,l] = \left[\mathbf{B}_0^T[m,l],\ \mathbf{B}_1^T[m,l]\right]^T \tag{26}$$

as depicted in Fig. 2.

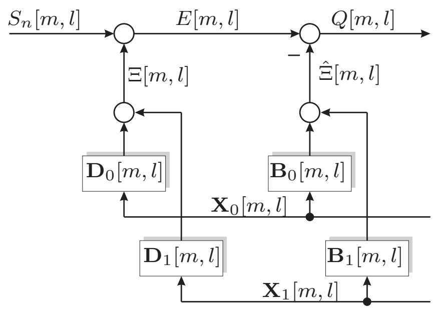

Fig. 2. Differential system for stereo residual echo estimation

The minimization of a error function $\mathrm{E}\left\{|Q[m,l]|^2\right\}$ which can be defined as

$$Q[\cdot,l] = \Xi[\cdot,l] - \mathrm{F}_c\left\{\mathbf{B}_{\mathrm{st}}^T[m,l]\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right\} \tag{27}$$

leads to [8]

$$\begin{aligned}\frac{\partial \mathrm{E}\left\{|Q[\cdot,l]|^2\right\}}{\partial \mathbf{B}_{\mathrm{st}}^*} &= -\mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right)^*\Xi[m,l]\right\}\\ &\quad+\mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right)^*\right.\\ &\quad\left.\mathrm{F}_c\left\{\mathbf{B}_{\mathrm{st}}^T[m,l]\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right\}\right\}.\end{aligned} \tag{28}$$

$\mathrm{E}\{\cdot\}$ is the expectation operator and $(\cdot)^*$ is the conjugate complex. The minimum of (28) gives us the optimal coefficient vector $\mathbf{B}_{\mathrm{st,opt}}[m,l]$ in the minimum mean squared error (MMSE)-sense.

$$\mathbf{B}_{\mathrm{st,opt}}[\cdot,l] = \mathrm{F}_c\left\{\mathbf{R}_{XX}^{-1}[m,l]\mathbf{\Phi}_{X\Xi}[m,l]\right\}. \tag{29}$$

with the correlation matrix $\mathbf{R}_{XX}[m,l]$ and the cross correlation vector $\mathbf{\Phi}_{X\Xi}[m,l]$:

$$\mathbf{R}_{XX}[m,l] = \mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right)^*\mathbf{X}_{\mathrm{st},L'_D}^T[m,l]\right\} \tag{30}$$

$$\mathbf{\Phi}_{X\Xi}[m,l] = \mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},L'_D}^T\mathbf{X}_{\mathrm{st},L'_D+1}[m,l]\right)^*\Xi[m,l]\right\} \tag{31}$$

If we assume the partitions of the loudspeaker channels to be temporarily uncorrelated

$$\frac{\mathrm{E}\{X_i^*[m,l]X_j[m,l+\lambda]\}}{\mathrm{E}\{X_i^*[m,l]X_i[m,l]\}} \ll 1 \tag{32}$$

for $\{i,j\} \in \{0,1\}$, and $\lambda \neq 0$ we can rewrite (29) as:

$$\mathbf{B}_{\mathrm{st}}[\cdot,l] \approx \mathrm{F}_c\left\{\left(\left[\begin{array}{c}\mathbf{I}_{L'_D}\mathbf{I}_{L'_D}\\ \mathbf{I}_{L'_D}\mathbf{I}_{L'_D}\end{array}\right]\odot\mathbf{R}_{XX}[m,l]\right)^{-1}\mathbf{\Phi}_{X\Xi}[m,l]\right\} \tag{33}$$

$\odot$ is the element-wise Hadamard multiplication. With (33) we can calculate the residual echo differential system $\mathbf{B}_{\mathrm{st}}[m,l]$ element by element for each channel. The matrix inversion problem is reduced to that of a 2×2 matrix inversion.

$$\mathbf{B}_{i,\mathrm{st}}[\cdot,l] \approx \mathrm{F}_c\left\{\mathbf{R}_{i,XX}^{-1}[m,l]\cdot\mathbf{\Phi}_{i,X\Xi}[m,l]\right\} \tag{34}$$

with

$$\mathbf{B}_{i,\mathrm{st}}[m,l] = [B_{0,i}[m,l], B_{1,i}[m,l]]^T \tag{35}$$

$$\mathbf{R}_{i,XX}[m,l] = \left[\begin{array}{cc}\Phi_{X_0X_0}[m,l-i] & \Phi_{X_0X_1}[m,l-i]\\ \Phi_{X_1X_0}[m,l-i] & \Phi_{X_1X_1}[m,l-i]\end{array}\right] \tag{36}$$

$$\mathbf{\Phi}_{i,X\Xi}[m,l] = \mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l-i]\right)^*\Xi[m,l]\right\} \tag{37}$$

It should be mentioned that the adaptation for $\mathbf{B}_{i,\mathrm{st}}[m,l]$ has to be stopped for the case of an active near speaker $S_n[m,l] \neq 0$. The estimation for the residual echo PSD $\hat{\Phi}_{\Xi\Xi}[m,l]$ is performed with the frozen $\mathbf{B}_{i,\mathrm{st}}[m,l]$. The double-talk detection used in this contribution is described in [9].

B. Post-Filter Design

With an estimate for the system misalignment $\mathbf{D}_{\mathrm{st}}[m,l]$, we obtain $\Xi[m,l]$ from (23). A reliable estimate of the residual echo PSD is essential for the design of the post-filter in order to avoid remaining echoes as well as desired speech distortions. The Wiener post-filter is given by

$$\begin{aligned}P[m,l] &= \frac{\hat{\Phi}_{S_nS_n}[m,l]}{\hat{\Phi}_{S_nS_n}[m,l]+\hat{\Phi}_{\Xi\Xi}[m,l]}\\ &= \frac{\hat{\Phi}_{EE}[m,l]-\hat{\Phi}_{\Xi\Xi}[m,l]}{\hat{\Phi}_{EE}[m,l]}.\end{aligned} \tag{38}$$

The PSD estimation in (38) can be obtained from (16).

III. Optimal Smoothed System Misalignment for the Stereo Case

In [4] an optimal smoothing for the residual echo differential system $\mathbf{B}[m,l]$ was derived for a single channel system by the authors. The smoothing was optimized with respect to the minimization of the bias introduced by the partitioned calculation of the system misalignment. In this contribution a frequency- and block-dependent adaptive smoothing will be derived for the stereo case.

A. Problem statement

As a first step we rewrite equation (23) as a sum of the vectors' elements.

$$\Xi[\cdot,l] = \sum_{i=0}^{L'_D-1}\mathrm{F}_c\left\{\mathbf{D}_{\mathrm{st},i}^T[m,l]\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l-i]\right\} \tag{39}$$

With (39) we can now rewrite (29) exemplarily for the first partition

$$\begin{aligned}\mathbf{B}_{\mathrm{st},0}[\cdot,l] &\approx \mathbf{R}_{0,XX}^{-1}[m,l]\cdot\mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\cdot\right.\\ &\qquad\left.\sum_{i=0}^{L'_D-1}\mathrm{F}_c\left\{\mathbf{D}_{\mathrm{st},i}^T[m,l]\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l-i]\right\}\right\}\\ &= \mathbf{D}_{\mathrm{st},0}[m,l]+\mathbf{R}_{0,XX}^{-1}[m,l]\cdot\\ &\quad\sum_{i=1}^{L'_D-1}\underbrace{\mathrm{E}\left\{\left(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}_{\mathrm{st},1}^T[m,l-i]\right\}}_{=\mathbf{0}}\mathbf{D}_{\mathrm{st},i}[m,l]\end{aligned} \tag{40}$$

Equation (40) illustrates that the residual echo system will be estimated correctly if the assumption of temporarily uncorrelated partitions of the loudspeaker signal $\mathbf{X}_{\mathrm{st}}[m,l]$ in (32) is fulfilled. For practical realizations the expectation operator $\mathrm{E}\{\cdot\}$ has to be replaced by an spectral estimation $\hat{\mathrm{E}}\{\cdot\}$ which introduces an additive bias:

$$\hat{\mathbf{R}}_{0,XX}^{-1}[m,l]\sum_{i=1}^{L'_D-1}\underbrace{\hat{\mathrm{E}}\{(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l])^*\mathbf{X}_{\mathrm{st},1}^T[m,l-i]\}}_{\neq\mathbf{0}}\mathbf{D}_{\mathrm{st},i}[m,l] \tag{41}$$

The calculation for all other partitions can be found analogously. From (41) we see that the correlations between the different partitions of the loudspeaker signals $\mathbf{X}_{\mathrm{st},L'_D}[m,l]$ contribute a major part of the bias of the residual echo differential system estimate.

B. Derivation of an adaptive smoothing factor

Since the residual echo differential system $\hat{\mathbf{B}}_{\mathrm{st},i}[m,l]$ is less time variant than the exciting speech signal the variance of $\hat{\mathbf{B}}_{\mathrm{st},i}[m,l]$ can be reduced by a first order recursive smoothing

$$\begin{aligned}\hat{\mathbf{B}}_{\mathrm{st},i}[m,l] &= \breve{\boldsymbol{\alpha}}_i[m,l]\hat{\mathbf{B}}_{\mathrm{st},i}[m,l-1]\\ &\quad+(\mathbf{I}_2-\breve{\boldsymbol{\alpha}}_i[m,l])\hat{\mathbf{R}}^{-1}_{i,XX}[m,l]\hat{\boldsymbol{\Phi}}_{i,X\Xi}[m,l]\end{aligned} \tag{42}$$

with a smoothing factor

$$\breve{\boldsymbol{\alpha}}_i[m,l]=\begin{bmatrix}\breve{\alpha}_{0,i}[m,l] & 0\\ 0 & \breve{\alpha}_{1,i}[m,l]\end{bmatrix} \tag{43}$$

In the following we derive a frequency dependent smoothing factor $\boldsymbol{\alpha}_{\mathrm{opt}}[m,l]=|\breve{\boldsymbol{\alpha}}_{\mathrm{opt}}[m,l]|^2$ which minimizes the distance between the squared estimate of the residual echo difference system $\hat{\mathbf{B}}_{\mathrm{st},i}[m,l]$ and the expectation of an squared stochastical system $\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]$. With an optimal $\boldsymbol{\alpha}_{\mathrm{opt}}[m,l]$ the expectation $\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]|^2\right\}$ should converge to the true squared system misalignment $|\mathbf{D}[m,l]|^2$:

$$\boldsymbol{\alpha}_{\mathrm{opt},i}[m,l]=\min_{\boldsymbol{\alpha}_i[m,l]}\left\||\hat{\mathbf{B}}_{\mathrm{st},i}[m,l]|^2-\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]|^2\right\}\right\|^2 \tag{44}$$

In the following we will derive the optimal smoothing factor again for the first partition exemplarily. We have

$$\begin{aligned}&\left\|\left|\hat{\mathbf{B}}_{\mathrm{st},0}[m,l]\right|^2-\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]|^2\right\}\right\|^2\\ &=\left\|\boldsymbol{\alpha}_0[m,l]\left|\hat{\mathbf{B}}_{\mathrm{st},0}[m,l-1]\right|^2+(\mathbf{I}_2-\boldsymbol{\alpha}_0[m,l])\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]\right.\right.\\ &\quad+\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\cdot\sum_{i=1}^{L'_D-1}\hat{\mathrm{E}}\left\{\left(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}\\ &\quad\left.\left.\cdot\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]|^2\right\}-\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]|^2\right\}\right\|^2\end{aligned} \tag{45}$$

The elements of $\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]$ are assumed to be zero-mean, mutually uncorrelated and statistically independent of the loudspeaker signals:

$$[\mathrm{E}\{\mathcal{D}_{0,i}[m,l]\},\mathrm{E}\{\mathcal{D}_{1,i}[m,l]\}]^T=\mathbf{0}_{2\times1}\ \text{for}\ 0\le i\le L'_D-1$$

$$\mathrm{E}\left\{\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\boldsymbol{\mathcal{D}}^T_{\mathrm{st},i}[m,l]\right\}=\begin{bmatrix}\mathrm{E}\{|\mathcal{D}_{0,i}[m,l]|^2\} & 0\\ 0 & \mathrm{E}\{|\mathcal{D}_{1,i}[m,l]|^2\}\end{bmatrix}$$

$$\mathrm{E}\left\{\mathbf{X}^*_{\mathrm{st},1}[m,l-i]\boldsymbol{\mathcal{D}}_{\mathrm{st},j}[m,l]^T\right\}=\mathbf{0}_{4\times4}\ \text{for}\ 0\le\{i,j\}\le L'_D-1$$

With these assumptions the expectation operator containing the summation term can be split into two parts.

$$\begin{aligned}&\mathrm{E}\left\{\left|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]+\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\cdot\right.\right.\\ &\quad\left.\left.\sum_{i=1}^{L'_D-1}\hat{\mathrm{E}}\left\{\left(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}\cdot\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\right|^2\right\}\\ &=\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]|^2\right\}+\mathrm{E}\left\{\left|\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\cdot\right.\right.\\ &\quad\left.\left.\sum_{i=1}^{L'_D-1}\hat{\mathrm{E}}\left\{\left(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}\cdot\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\right|^2\right\}\end{aligned}$$

With $|\mathbf{a}|^2=\mathrm{diag}\left\{\mathbf{a}^*\mathbf{a}^T\right\}$ for an arbitrary complex vector $\mathbf{a}$ we further rewrite the summation part:

$$\begin{aligned}&\mathrm{E}\left\{\left|\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\cdot\right.\right.\\ &\quad\left.\left.\sum_{i=1}^{L'_D-1}\hat{\mathrm{E}}\left\{\left(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}\cdot\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\right|^2\right\}\\ &=\mathrm{E}\left\{\sum_{i=1}^{L'_D-1}\mathrm{diag}\left\{(\hat{\mathbf{R}}^*_{0,XX}[m,l])^{-1}\cdot\right.\right.\\ &\quad\hat{\mathrm{E}}\left\{(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l])\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}^*\mathrm{diag}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]|^2\right\}\\ &\quad\left.\left.\hat{\mathrm{E}}\left\{(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l])^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}^T\cdot\hat{\mathbf{R}}^{-T}_{0,XX}[m,l]\right\}\right\}\\ &\approx\sum_{i=1}^{L'_D-1}\mathrm{E}\left\{\left|\hat{\mathrm{E}}\left\{\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}^H_{\mathrm{st},2}[m,l-i]\mathrm{diag}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]|^2\right\}\right.\right.\right.\\ &\quad\left.\left.\left.\mathbf{X}_{\mathrm{st},1}[m,l-i]\right\}\right|\right\}\cdot\mathrm{diag}\left\{\left|\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\right|\right\}\end{aligned} \tag{46}$$

It should be mentioned that the inversion for the main diagonal elements of $\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]$ which is a 2×2 matrix can be easily calculated by (dependency of $[m,l]$ is omitted)

$$\mathrm{diag}\left\{\left|\hat{\mathbf{R}}^{-1}_{0,XX}\right|\right\}=\left|\frac{1}{\hat{\Phi}_{X_0X_0}\hat{\Phi}_{X_1X_1}\left(1-\left|\hat{\Gamma}_{X_0X_1}\right|^2\right)}\begin{bmatrix}\hat{\Phi}_{X_1X_1}\\ \hat{\Phi}_{X_0X_0}\end{bmatrix}\right| \tag{47}$$

with the magnitude squared coherence (MSC)

$$\left|\hat{\Gamma}_{X_0X_1}[m,l]\right|^2=\frac{\left|\hat{\Phi}_{X_0X_1}[m,l]\right|^2}{\hat{\Phi}_{X_0X_0}[m,l]\hat{\Phi}_{X_1X_1}[m,l]}. \tag{48}$$

It is now possible to separate (45) into a part depending on $\boldsymbol{\alpha}_0[m,l]$ and one part that is not depending on $\boldsymbol{\alpha}_0[m,l]$:

$$\begin{aligned}&\left\|\left|\hat{\mathbf{B}}_{\mathrm{st},0}[m,l]\right|^2-\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]|^2\right\}\right\|^2\\ &\qquad=\|\boldsymbol{\alpha}_0[m,l]\boldsymbol{\mathcal{A}}[m,l]+\boldsymbol{\mathcal{B}}[m,l]\|^2\end{aligned} \tag{49}$$

with

$$\begin{aligned}\boldsymbol{\mathcal{B}}[m,l]=\mathrm{E}&\left\{\left|\sum_{i=1}^{L'_d-1}\hat{\mathbf{R}}^{-1}_{0,XX}[m,l]\right.\right.\\ &\left.\left.\hat{\mathrm{E}}\left\{\left(\mathbf{T}^T_{\mathrm{st},1}\mathbf{X}_{\mathrm{st},2}[m,l]\right)^*\mathbf{X}^T_{\mathrm{st},1}[m,l-i]\right\}\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\right|^2\right\}\end{aligned} \tag{50}$$

and

$$\begin{aligned}\boldsymbol{\mathcal{A}}[m,l]&=\left|\hat{\mathbf{B}}_{\mathrm{st},0}[m,l-1]\right|^2-\mathrm{E}\left\{|\boldsymbol{\mathcal{D}}_{\mathrm{st},0}[m,l]|^2\right\}-\boldsymbol{\mathcal{B}}[m,l]\\ &=\boldsymbol{\mathcal{G}}[m,l]-\boldsymbol{\mathcal{B}}[m,l].\end{aligned} \tag{51}$$

With the definition of $\boldsymbol{\alpha}_i[m,l]$ as a diagonal matrix after (43) for the avoidance of interference between the loudspeaker channels the solution which minimizes the distance between the ℓ_2-norm of the difference between the estimated residual echo differential system and the stochastical residual echo differential system after (49) is given with

$$\boldsymbol{\alpha}_0[m,l]=\begin{bmatrix}-\dfrac{\mathcal{B}_0[m,l]}{\mathcal{A}_0[m,l]} & 0\\ 0 & -\dfrac{\mathcal{B}_1[m,l]}{\mathcal{A}_1[m,l]}\end{bmatrix}. \tag{52}$$

After putting (50),(51),(46), and (47) into (52) we finally get

$$\operatorname{diag}\{\boldsymbol{\alpha}_{\mathrm{opt},0}[m,l]\} = \begin{bmatrix} \left(1 - \frac{\mathcal{G}_0[m,l]}{\mathcal{B}_0[m,l]}\right)^{-1} \\ \left(1 - \frac{\mathcal{G}_1[m,l]}{\mathcal{B}_1[m,l]}\right)^{-1} \end{bmatrix} = \mathbf{1}_2 \oslash \tag{53}$$

$$\left(1 + \frac{\mathrm{E}\left\{\left|\hat{\Phi}_{X_0X_0}[m,l]\hat{\Phi}_{X_1X_1}[m,l]\left(1-\left|\hat{\Gamma}_{X_0X_1}[m,l]\right|^2\right)\right|\right\}}{\sum_{i=1}^{L'_d-1}\mathrm{E}\left\{\left|\hat{\mathrm{E}}\left\{(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l-i])^H\operatorname{diag}\left\{\left|\boldsymbol{\mathcal{D}}_{\mathrm{st},i}[m,l]\right|^2\right\}\mathbf{X}_{\mathrm{st},1}[m,l-i]\right\}\right|\right\}} \cdot \begin{bmatrix} \frac{\left|\mathrm{E}\left\{\left|\mathcal{D}_{0,0}[m,l]\right|^2\right\}-\left|\hat{B}_{0,0}[m,l-1]\right|^2\right|}{\mathrm{E}\left\{\left|\hat{\Phi}_{X_1X_1}[m,l]\right|\right\}} \\ \frac{\left|\mathrm{E}\left\{\left|\mathcal{D}_{1,0}[m,l]\right|^2\right\}-\left|\hat{B}_{1,0}[m,l-1]\right|^2\right|}{\mathrm{E}\left\{\left|\hat{\Phi}_{X_0X_0}[m,l]\right|\right\}} \end{bmatrix}\right)$$

with $\mathbf{1}_N \oslash \mathbf{x} = [1/x_1, 1/x_2, ..., 1/x_N]^T$ as the element-wise division defined equivalently to the well known Hadamard Multiplication. It should be mentioned that $\boldsymbol{\alpha}_{\mathrm{opt},0}[m,l]$ depends on the MSC. More precisely a high MSC slows down the adaptation which is similar to the findings concerning step-size control of a stereo AEC in [10]. Furthermore a high PSD in the considered partition compared to the PSDs of the other partitions speeds up the adaptation while high PSDs in adjacent partitions slow down the adaptation. Thus the adaptation is slowed down for the case of a high bias for the estimate (see equation (41)).

As in [4] we now introduce an approximation which is independent of all unknown systems. The misalignment $\mathrm{E}\{|\mathcal{D}_{0,0}[m,l]|^2\}$ is not available and to avoid recursive structures, $|\hat{B}_{0,0}[m,l-1]|^2$ is omitted, too. Thus both systems are replaced by a factor $\eta = (1/C - 1)\,2(L'_D - 1)$ which depends on the number of blocks of $\hat{\mathbf{D}}[m,l]$ and a fixed constant C:

$$\operatorname{diag}\left\{\boldsymbol{\alpha'}_0[m,l]\right\} = \mathbf{1}_2 \oslash \tag{54}$$

$$\left(1 + \frac{\eta\left|\hat{\Phi}_{X_0X_0}[m,l]\hat{\Phi}_{X_1X_1}[m,l]\left(1-\left|\hat{\Gamma}_{X_0X_1}[m,l]\right|^2\right)\right|}{\sum_{i=1}^{L'_d-1}\left|\hat{\mathrm{E}}\left\{(\mathbf{T}_{\mathrm{st},1}^T\mathbf{X}_{\mathrm{st},2}[m,l-i])^H\mathbf{X}_{\mathrm{st},1}[m,l-i]\right\}\right|} \begin{bmatrix} \frac{1}{\left|\hat{\Phi}_{X_1X_1}[m,l]\right|} \\ \frac{1}{\left|\hat{\Phi}_{X_0X_0}[m,l]\right|} \end{bmatrix}\right)$$

IV. Simulation Results

Fig. 3 shows the relative system misalignment

$$\Delta D[l] = \frac{\sum_{m=0}^{L_{\mathrm{DFT}}}\left\|\hat{\mathbf{B}}[m,l] - \mathbf{D}[m,l]\right\|^2}{\sum_{m=0}^{L_{\mathrm{DFT}}}\left\|\mathbf{D}[m,l]\right\|^2} \tag{55}$$

for an speech signal for $\mathbf{X}_{\mathrm{st}}[m,l]$. The room reverberation time was $\tau_{60} = 0.4$sec. The DFT-length was $L_{\mathrm{DFT}} = 512$ and the considered length of the RIRs were $L_D = 1024$. After 5 seconds the RIRs in the sending room were switched. It can be seen that the optimal smoothing significantly reduces the relative system misalignment according to (55) and that the approximation $\boldsymbol{\alpha'}[m,l]$ in (54) gives a good approximation which is suitable for practical implementations.

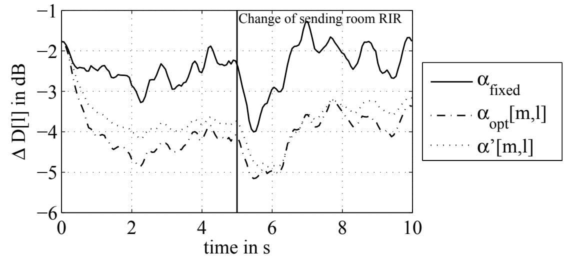

Fig. 3. Relative system misalignment $\Delta D[l]$ for speech as loudspeaker signal

Fig. 4 compares the proposed algorithm with the approximation for the optimal smoothing to the conventional Partitioned Frequency Block LMS (PFBLMS) algorithm. It can be seen, that the proposed method leads to a faster but more coarse adaptation than the PFBLMS.

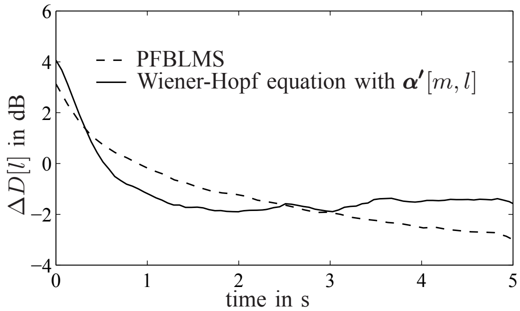

Fig. 4. Comparison between the proposed algorithm with adaptive smoothing and direct Wiener Hopf-Solution

V. Conclusions

In this paper we pursued the identification of the residual echo differential system by solving the Wiener-Hopf equation in the DFT domain in a stereo setup. This allows for faster but more coarse adaptation than gradient algorithms. The bias caused by a partitioned calculation was analyzed and an adaptive smoothing was proposed which minimizes this bias. Simulations show a significant reduction of the estimation bias and thus a more accurate estimate of the residual echo which is necessary for the design of a post-filter.

References

[1] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic Echo Control – An Application of Very-High-Order Adaptive Filters," *IEEE Signal Processing Magazine*, pp. 42–69, July 1999.

[2] M. Kallinger, J. Bitzer, and K. D. Kammeyer, "Post-Filtering for Stereo Acoustic Echo Cancellation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sep 2003.

[3] G. Enzner, R. Martin, and P. Vary, "Unbiased Residual Echo Power Estimation for Hands–Free Telephony," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP-2002)*, Orlando, Florida, USA, May 2002.

[4] S. Goetze, M. Kallinger, and K.-D. Kammeyer, "Residual Echo Power Spectral Density Estimation Based on an Optimal Smoothed Misalignment For Acoustic Echo Cancelation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005) , Eindhoven, The Netherlands*, 12.-15. Sept. 2005, pp. 209–212.

[5] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, pp. 14–34, January 1992.

[6] C. M. Rader, "An Improved Algorithm for High Speed Autocorrelation with Application to Spectral Estimation," *IEEE Trans. on Audio and Electroacoustics*, vol. 18, Dec 1979.

[7] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 2, pp. 156–165, Mar 1998.

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

[9] M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Enhanced Double-Talk Detection Based on Pseudo-Coherence in Stereo," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005) , Eindhoven, The Netherlands*, 12.-15. Sept. 2005, pp. 177–180.

[10] M. Gauger, "An Improved Method for Stereo Acoustic Echo Cancelling," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2003, pp. 596–599.