

Multi-Channel Listening-Room Compensation using a Decoupled Filtered-X LMS Algorithm

*Stefan Goetze, †Markus Kallinger, ‡Alfred Mertins, and *Karl-Dirk Kammeyer

Abstract—Dereverberation of speech signals in a hands-free scenario by inverse filtering has been a research topic for several years now. However, it is still a challenging problem because of the nature of common room impulse responses (RIRs), which are time-variant mixed phase systems having a large number of zeros close to, on, and even outside the unit circle in the z -domain. In this contribution an adaptive multi-channel equalization algorithm based on a decoupled version of the modified filtered-X LMS (mFxLMS) will be derived in the partitioned frequency domain. This new algorithm allows for fast convergence, computationally efficient implementation, and a low system delay under realistic conditions such as ambient noise and imperfect RIR estimates.

I. INTRODUCTION

For equalization of time-varying room impulse responses (RIRs) by adaptive FIR filters [1], robust, fast converging update algorithms are desirable. Since the acoustic environment is time-variant in a hands-free system, the identification and equalization of room transfer functions has to be performed adaptively. Prominent adaptive filter designs for acoustic equalization known from the field of active noise control (ANC) are based on the least-mean-squares (LMS) algorithm [2], such as the filtered-X least-mean-squares (FxLMS) algorithm [3, p. 280ff] or the modified filtered-X LMS (mFxLMS) [4]. Fast variants based on the recursive least-squares (RLS) algorithm exist [2], but they cause high computational load and may suffer from stability problems.

In this contribution a decoupled version of the multi-channel mFxLMS is proposed which is updated in the partitioned frequency domain [5], [6]. By this, the proposed algorithm is computationally efficient by exploiting the properties of the fast Fourier transform (FFT). Only a small delay is introduced by a block-by-block update. This new algorithm which was described in [7] in time-domain for the single-channel case has the capability to converge faster than FxLMS and mFxLMS (even in speech pauses) because it is excited independently from the input signal. Its convergence speed can be further increased by an iterative processing because it is also independent of the sampling rate of the input signal. Thus, a tradeoff between convergence speed and complexity can be utilized to adapt it to the available processing power.

*Stefan Goetze and Karl-Dirk Kammeyer are with University of Bremen, Dept. of Communications Engineering, 28334 Bremen, Germany, †Markus Kallinger is with Fraunhofer Institute for Integrated Circuits (IIS), 91058 Erlangen, he contributed to this work while he was with University of Oldenburg, Signal Processing Group, 26111 Oldenburg, Germany, ‡Alfred Mertins is with University of Lübeck, Institute for Signal Processing, 23538 Lübeck, Germany. Work was supported in parts by German Research Foundation DFG (Ka841-17).

Notation: Vectors and matrices are printed in boldface while scalars are printed in italic. k , n , and ℓ are the discrete time, frequency, and block index, respectively. All frequency domain variables are printed in sans-serif letters (e.g. $\mathbf{x}[n, \ell]$). By this, time and frequency domain are distinguishable even if the dependence of the variable k or n is omitted as in $\mathbf{x}[\ell]$. The superscripts T , $*$, and H denote the transposition, the complex conjugation, and the Hermitian transposition, respectively. The operator $*$ denotes the convolution of two sequences, $\mathbb{E}\{\cdot\}$ is the expectation, and the operator $\text{convmtx}\{\mathbf{h}^T, L_c\}$ generates a convolution matrix of size $L_c \times (L_c + L_h - 1)$. The operator $\text{diag}\{\cdot\}$ builds up a matrix of size $L \times L$ from a vector of size $L \times 1$ that has the vector's elements on its main diagonal and zeros elsewhere, and the operator $\text{bdiag}\{\cdot\}$ generates a matrix of size $L'L \times L'L$ having matrices of size $L \times L$ on its block diagonal and zeros else.

II. LISTENING ROOM COMPENSATION

A. Least-Squares Equalization of Room Impulse Responses

The common setup for multi-channel listening-room compensation (LRC) is shown in Fig. 1. Here, the equalizer filters $\mathbf{c}_{\text{EQ}}[k]$ precede the acoustic channels $\mathbf{H}[k]$. The goal of the equalizer (EQ) is to remove reverberation which is caused by the convolution of the P loudspeaker signals with the RIRs at the position of the Q reference microphones where the user of the system is assumed to be located. Thus the EQ minimizes the error signals $\mathbf{e}_{\text{EQ}}[k]$ and by this the Euclidean distance between the concatenated systems of EQ filters and RIRs and the desired target systems $\mathbf{d}[k]$.

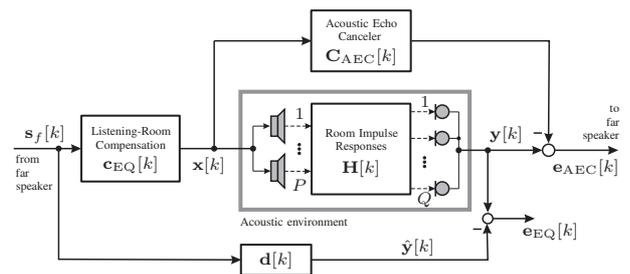


Fig. 1. Multi-channel listening-room compensation system with an AEC for system identification.

For acoustic equalization an estimate of the RIR is needed which has to be tracked adaptively. This is done in the schematic in Fig. 1 by the acoustic echo canceller (AEC) $\mathbf{C}_{\text{AEC}}[k]$. The system identification performed by the AEC will be imperfect due to the unmodeled tails of the RIRs and

estimation errors especially in periods of initial convergence or after RIR changes [8]. Furthermore, if more than one loudspeaker is used the AEC may suffer from the so-called stereo problem of acoustic echo cancellation [9]. Thus, the AEC may not converge to the desired *system identification* solution since the problem is overdetermined and the solution is non unique.

B. Multi-Channel LS-EQ

Since the direct inversion is not possible for most real-world RIRs due to their non-minimum phase property [10] least-squares equalizers (LS-EQs) minimize the Euclidean distance between the concatenated overall system of EQ filters and RIRs and the desired target systems

$$\hat{\mathbf{c}}_{\text{EQ}} = \underset{\mathbf{c}_{\text{EQ}}}{\text{argmin}} \|\mathbf{c}_{\text{EQ}} \mathbf{H}_{\text{CM}} - \mathbf{d}\|^2. \quad (1)$$

Solving (1) leads to the well known least squares equalizer

$$\mathbf{c}_{\text{EQ}} = \mathbf{H}_{\text{CM}}^+ \mathbf{d} \quad (2)$$

with the following vector and matrix definitions:

$$\mathbf{c}_{\text{EQ}} = [\mathbf{c}_{\text{EQ},1}^T, \mathbf{c}_{\text{EQ},2}^T, \dots, \mathbf{c}_{\text{EQ},P}^T] \quad (3)$$

$$\mathbf{c}_{\text{EQ},p} = [c_{\text{EQ},p,0}, c_{\text{EQ},p,1}, \dots, c_{\text{EQ},p,L_c,\text{EQ}-1}]^T \quad (4)$$

$$\mathbf{H}_{\text{CM}} = \begin{bmatrix} \mathbf{H}_{\text{CM},11} & \cdots & \mathbf{H}_{\text{CM},1Q} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{\text{CM},P1} & \cdots & \mathbf{H}_{\text{CM},PQ} \end{bmatrix} \quad (5)$$

$$\mathbf{H}_{\text{CM},pq} = \text{convmtx}\{\mathbf{h}_{pq}^T, L_{c,\text{EQ}}\} \quad (6)$$

$$\mathbf{h}_{pq} = [h_{pq,0}, h_{pq,1}, \dots, h_{pq,L_h}]^T \quad (7)$$

$$\mathbf{d} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_Q^T] \quad (8)$$

$$\mathbf{d}_q = [0, \dots, 0, d_0, d_1, \dots, d_{L_d-1}, \underbrace{0, \dots, 0}_{L_h+L_c,\text{EQ}-1-L_d-k_{0,q}}]^T \quad (9)$$

Here \mathbf{c}_{EQ} and \mathbf{H}_{CM} are the coefficient vector of the equalizer and the channel convolution matrix of size $PL_{c,\text{EQ}} \times (L_h + L_{c,\text{EQ}} - 1)Q$ built from the RIR coefficients, respectively. The subindex CM indicates the definition as a time-domain convolution matrix unlike the definition of the MIMO channel matrix \mathbf{H} used later in this paper in (17). \mathbf{H}_{CM}^+ is the Moore-Penrose pseudoinverse of \mathbf{H}_{CM} and \mathbf{d} is the vector containing the desired systems, which can be chosen as delayed unit impulses, delayed high passes or delayed band passes. We build up \mathbf{d} from 40th order finite impulse response (FIR) highpasses with band limits at 200Hz at a sampling frequency of $f_s = 8\text{kHz}$. The delay introduced by the equalizer is denoted by $k_{0,q}$. Differing delays for different channels q can be advantageous if the theoretical delay differences between loudspeakers and microphones are known from the geometry [1]. For proper choices for $k_{0,q}$ see [11].

In general multi-channel LRC is superior to single-channel LRC due to the following reasons: If spatial diversity can be exploited by the use of multiple loudspeakers perfect inversion may be possible [12] exploiting the so-called multiple input/output inverse theorem (MINT) if the RIRs do

not have common zeros in the z -domain. Multi-microphone systems increase spatial robustness compared to single-channel LRC systems [8]. Fig. 2 shows the performance of multi-channel LS equalization for different numbers of loudspeakers and microphones in terms of the signal-to-reverberation ratio enhancement (SRRE) [8]. In Fig. 2 (a)-(c) the EQ is designed for a single loudspeaker system ($J = 1$) and for $K \in \{1, 12, 28\}$ reference microphones lying on a rectangle in the center of the specific subpicture. The room dimensions are (5.6m x 4.375m x 3.5m) and the loudspeaker position is at (1.7m, 2.0m, 1.0m). It can be seen that the use of multiple microphones increases spatial robustness while the maximum achievable SRR enhancement decreases slightly from 14.8 dB to 12 dB. This is due to the fact that a multi-microphone LRC system leads to a mean equalization for the given spatial positions of the reference microphones. It should be emphasized that spatial robustness of a LRC device is an essential requirement for the use in hands-free telecommunication systems [8]. If a second loudspeaker is added to the system (see Fig. 2d-f) the overall performance is increased as it can be seen from the achievable maximum SRRE values. However, using multiple loudspeakers leads to a loss of spatial robustness which becomes obvious especially by comparing Figs. 2 (a) and (d).

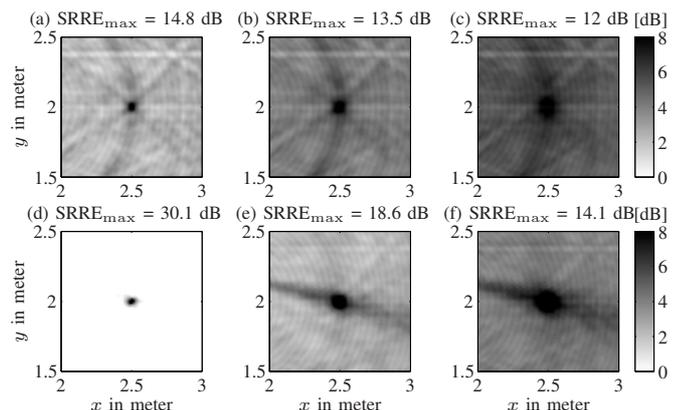


Fig. 2. Spatial robustness of listening-room compensation in terms of SRRE as a function of the location for different number of microphones and loudspeakers. a)-c) one loudspeaker system with 1, 12, and 28 reference microphones, d)-f) two loudspeaker system with 1, 12, and 28 microphones.

C. System Identification for Listening Room Compensation

All algorithms in this contribution rely on knowledge of the RIRs which have to be equalized. In real-world systems this knowledge is not available and the RIRs have to be identified by adaptive algorithms. Acoustic echo cancelers perform system identification at least for the single loudspeaker case [8]. Adaptive tracking of the RIR estimates is necessary due to the time-varying nature of common RIRs. Thus, estimation errors are inevitable, e.g. in periods of initial convergence or after RIR changes. As known from extensive research in the field of stereo acoustic echo cancellation a multi-loudspeaker system identification is not solvable uniquely [9] due to the high

correlation of the loudspeaker signals $x_p[k]$ which originate from the same source most of the time.

As illustrated in Fig. 3 for a two-loudspeaker scenario exemplarily the RIRs $\mathbf{h}_p[k]$ can be split up into one part $\hat{\mathbf{h}}_p[k]$ which is correctly identified by the AEC and an estimation error $\tilde{\mathbf{h}}_p[k]$:

$$\mathbf{h}_p[k] = \begin{bmatrix} \hat{\mathbf{h}}_p[k] \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{h}}_p[k] = \begin{bmatrix} \mathbf{c}_{\text{AEC},p}[k] \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{h}}_p[k] \quad (10)$$

$$\mathbf{h}_p[k] = [h_{p,0}[k], \dots, h_{p,L_h-1}[k]]^T \quad (11)$$

$$\hat{\mathbf{h}}_p[k] = [c_{\text{AEC},p,0}[k], \dots, c_{\text{AEC},p,L_{c,\text{AEC}}-1}[k]]^T \quad (12)$$

$$\tilde{\mathbf{h}}_p[k] = [\tilde{h}_{p,0}[k], \dots, \tilde{h}_{p,L_h-1}[k]]^T \quad (13)$$

Here, $L_{c,\text{AEC}}$ is the length of the AEC filter which equals $L_{\tilde{h}}$ and is, in general, smaller than the length of the RIR L_h . Thus, the so-called tail of the RIR which cannot be identified by the AEC always contributes to the estimation error $\tilde{\mathbf{h}}_p[k]$ and leads to a decreased performance of the equalizer [13].

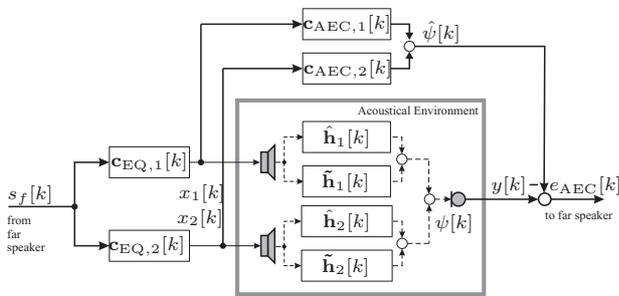


Fig. 3. Combined system of equalizer and acoustic echo canceler. The RIRs can be split into a part modeled by the AEC $\hat{\mathbf{h}}_p[k] = \mathbf{c}_{\text{AEC},p}[k]$ and the AEC system misalignment $\tilde{\mathbf{h}}_p[k]$ (estimation error).

It can be seen easily from the error signal of a stereo AEC filter

$$\|e_{\text{AEC}}\|^2 = \|x_1[k] * (\mathbf{h}_1[k] - \mathbf{c}_{\text{AEC},1}[k]) + x_2[k] * (\mathbf{h}_2[k] - \mathbf{c}_{\text{AEC},2}[k])\|^2 \quad (14)$$

that several solutions exist for minimizing (14). Different approaches exist for decorrelation of the loudspeaker signals, such as adding (masked) uncorrelated noise, nonlinear processing, etc. However, the system identification performance of multi-loudspeaker AEC systems is not sufficient for an equalizer relying on this information as it is illustrated in Fig. 4.

The upper part of Fig. 4 shows the inversion of a system with two loudspeakers and one microphone in time domain (left) and frequency domain (right) for correct system identification $\mathbf{c}_{\text{AEC},1}[k] = \mathbf{h}_1[k]$ and $\mathbf{c}_{\text{AEC},2}[k] = \mathbf{h}_2[k]$. The equalization is nearly perfect due to exploiting spatial diversity [12]. The lower part of Fig. 4 shows the inversion for $\mathbf{c}_{\text{AEC},1}[k] = \mathbf{h}_1[k] + \mathbf{n}[k]$ and $\mathbf{c}_{\text{AEC},2}[k] = \mathbf{h}_2[k] - \mathbf{n}[k]$. Since the same noise signal $\mathbf{n}[k]$ was used, $\|e_{\text{AEC}}\|^2$ in (14) still equals zero. The lower part of Fig. 4 shows that inversion

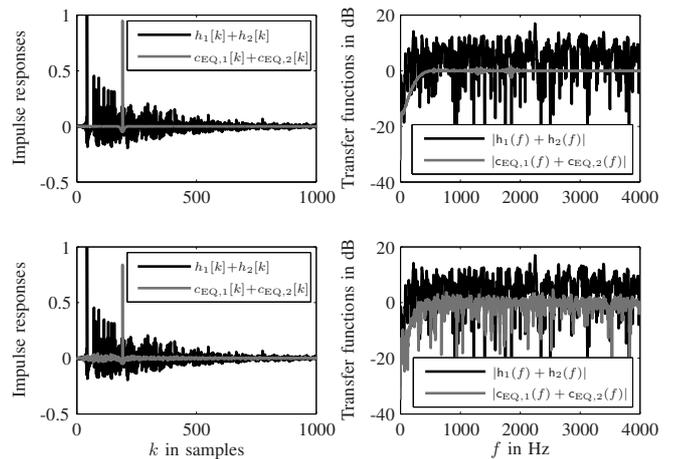


Fig. 4. Multi-channel (MISO) equalization of room impulse responses. Room reverberation time is $\tau_{60} = 250\text{ms}$, RIR length and EQ length are $L_{\text{EQ}} = 1024$ and $L_h = 2048$, respectively.

fails even if $\mathbf{n}[k]$ has very low power. Since lags in tracking of time-variant RIRs and the so-called *tail-effect* of stereo acoustic echo cancellation lead to further estimation errors multi-loudspeaker system inversion is often not sufficiently robust to be used in fast-changing real-world systems.

D. Frequency-Domain Gradient Algorithms for LRC

Direct application of the least-squares equalization (2) involves the inversion of the RIR matrix having a size of several thousand taps (see (2) and (5)). For real-time systems adaptive gradient algorithms like the FxLMS [3, p. 280ff] have to be used which suffer from slow convergence speed. Faster algorithms exist [2], however they might suffer from high computational load or stability problems. In the following a simple, efficient and fast converging algorithm will be derived which is based on the modified filtered-X least-mean-squares (mFxLMS) algorithm. Fig. 5 shows a schematic of the conventional FxLMS (for the single channel case for simplicity reasons) which is the basis for the following mFxLMS and dFxLMS algorithms and, thus, will be described briefly in the block frequency domain in the following.

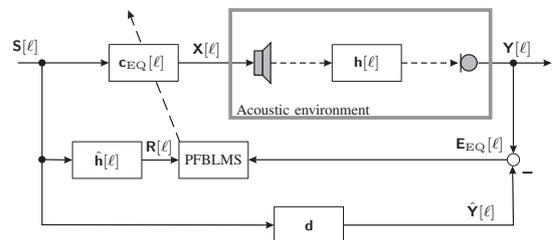


Fig. 5. Block diagram of the single channel filtered-X LMS (FxLMS) in partitioned frequency domain.

A fast converging and computationally efficient approach for adaptive filtering is the so-called multi delay filtering

(MDF) [6], [14] which can be expressed easily by

$$\mathbf{Y}[\ell] = \mathbf{G}\mathbf{X}[\ell]\mathbf{H}[\ell] \quad (15)$$

using the following definitions:

$$\mathbf{H}[\ell] = \text{bdiag}\{\underbrace{\mathbf{F}_{2L \times L}, \mathbf{F}_{2L \times L}, \dots, \mathbf{F}_{2L \times L}}_{PL'_h}\}\mathbf{H}[\ell] \quad (16)$$

$$\mathbf{H}[\ell] = \begin{bmatrix} \mathbf{h}_{11}[\ell] & \cdots & \mathbf{h}_{1Q}[\ell] \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{P1}[\ell] & \cdots & \mathbf{h}_{PQ}[\ell] \end{bmatrix} \quad (17)$$

$$\mathbf{X}[\ell] = [\check{\mathbf{X}}_1[\ell], \dots, \check{\mathbf{X}}_1[\ell - L'_h + 1], \dots, \check{\mathbf{X}}_P[\ell], \dots, \check{\mathbf{X}}_P[\ell - L'_h + 1]] \quad (18)$$

$$\check{\mathbf{X}}_p[\ell] = \text{diag}\{\mathbf{F}_{2L \times L}\mathbf{x}_p[\ell] + \tilde{\mathbf{I}}_{2L \times 2L}\mathbf{F}_{2L \times L}\mathbf{x}_p[\ell - 1]\} \quad (19)$$

$$\mathbf{x}_p[\ell] = [x_p[\ell L], x_p[\ell L + 1], \dots, x_p[(\ell + 1)L - 1]]^T \quad (20)$$

$$\tilde{\mathbf{I}}_{2L \times 2L} = \text{diag}\left\{ \left[e^{-j\pi^0}, e^{-j\pi^1}, \dots, e^{-j\pi(2L-1)} \right] \right\} \\ = \text{diag}\{[1, -1, \dots, 1, -1, \dots]\} \quad (21)$$

$$\mathbf{G} = \mathbf{F}_{2L \times 2L}\mathbf{W}_{2L \times L}^{01}(\mathbf{W}_{2L \times L}^{01})^T\mathbf{F}_{2L \times 2L}^{-1} \quad (22)$$

$$\mathbf{F}_{2L \times L} = \mathbf{W}_{2L \times L}^{10}\mathbf{F}_{2L \times 2L} \quad (23)$$

$$\mathbf{W}_{2L \times L}^{01} = [\mathbf{0}_{L \times L}, \mathbf{I}_{L \times L}]^T \quad (24)$$

$$\mathbf{W}_{2L \times L}^{10} = [\mathbf{I}_{L \times L}, \mathbf{0}_{L \times L}]^T \quad (25)$$

Here $\mathbf{F}_{2L \times 2L}$ and $\mathbf{F}_{2L \times 2L}^{-1}$ are the discrete Fourier transform (DFT) and the inverse DFT (IDFT) matrix, respectively, and $\mathbf{W}_{2L \times L}^{01}$ and $\mathbf{W}_{2L \times L}^{10}$ are window matrices. The constraining matrix \mathbf{G} suppresses cyclic convolution products [14], and the shifting matrix $\tilde{\mathbf{I}}_{2L \times 2L}$ represents a time shift of one block in frequency domain. By using $\tilde{\mathbf{I}}_{2L \times 2L}$ in (19), the DFT $\mathbf{F}_{2L \times L}\mathbf{x}_p[\ell - 1]$ of the previous block needs not be recalculated. The modified DFT matrix $\mathbf{F}_{2L \times L}$ of size $2L \times L$ represents a FFT with 50% zero-padding. L'_h denotes the number of partitions used to express the RIR vector \mathbf{h} and can be calculated by dividing the length of the RIR L_h by the block length: $L'_h = L_h/L$.

Similar to (15) we can express the block frequency domain signals $\mathbf{R}[\ell]$ and $\mathbf{E}_{\text{EQ}}[\ell]$ which we need for the formulation of a partitioned frequency block LMS (PFBLMS) gradient update algorithm:

$$\mathbf{E}_{\text{EQ}}[\ell] = \mathbf{G}\mathbf{X}[\ell]\mathbf{H}[\ell] - \hat{\mathbf{Y}}[\ell] \quad (26)$$

$$\hat{\mathbf{Y}}[\ell] = \mathbf{G}\mathbf{S}_f[\ell]\mathbf{D} \quad (27)$$

$$\mathbf{R}[\ell] = \mathbf{G}[\mathbf{S}_f[\ell], \dots, \mathbf{S}_f[\ell]]\hat{\mathbf{H}}[\ell] \quad (28)$$

With (26)-(28) the FxLMS can be expressed in the partitioned frequency domain by minimizing the error criterion

$$J[\ell] = (1 - \lambda) \sum_{i=1}^{\ell} \lambda^{\ell-i} \text{tr}\{\mathbf{E}_{\text{EQ}}^H[i]\mathbf{E}_{\text{EQ}}[i]\} \quad (29)$$

as

$$\Phi_{\text{RR}}^{-1}[\ell] = \lambda\Phi_{\text{RR}}^{-1}[\ell - 1] + (1 - \lambda)\mathbf{R}^H[\ell]\mathbf{G}\mathbf{R}[\ell] \quad (30)$$

$$\mathbf{C}_{\text{EQ}}^{\text{Fx}}[\ell] = \mathbf{C}_{\text{EQ}}^{\text{Fx}}[\ell - 1] + \mu\Phi_{\text{RR}}^{-1}[\ell]\mathbf{R}^H[\ell]\mathbf{E}[\ell]. \quad (31)$$

Here, $0 \leq \lambda \leq 1$ is the *forgetting* or *smoothing* factor. A simple algorithm with increased convergence speed compared to the FxLMS which allows for a larger stepsize than the conventional FxLMS and, thus, for faster convergence is the mFxLMS [4]. As already shown in [7] for the single-channel case slight modifications of Fig. 5 lead to the mFxLMS and, furthermore, to the decoupled filtered-X least-mean-squares (dFxLMS) algorithms [7], which are depicted in Fig. 6. If the switch Sw_1 is in the right position, Fig. 6 depicts the mFxLMS. With (30) and (28) the update of the block frequency domain

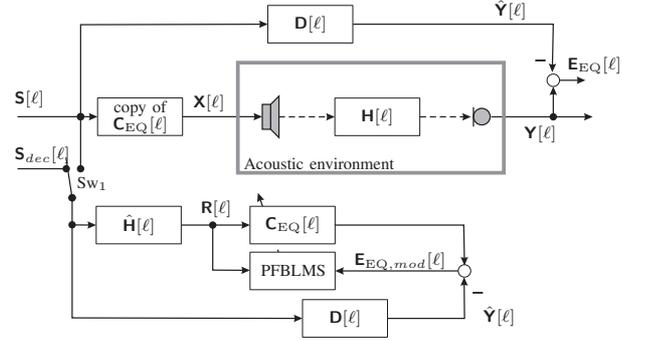


Fig. 6. Block diagram of modified filtered-X (mFxLMS) and decoupled filtered-X LMS (dFxLMS) in the partitioned frequency domain.

mFxLMS is given by

$$\mathbf{E}_{\text{EQ},\text{mod}}[\ell] = \mathbf{G}\mathbf{R}[\ell]\mathbf{C}_{\text{EQ}}^{\text{mFx}}[\ell] - \hat{\mathbf{Y}}[\ell] \quad (32)$$

$$\mathbf{C}_{\text{EQ}}^{\text{mFx}}[\ell] = \mathbf{C}_{\text{EQ}}^{\text{mFx}}[\ell - 1] + \mu\Phi_{\text{RR}}^{-1}[\ell]\mathbf{R}^H[\ell]\mathbf{E}_{\text{EQ},\text{mod}}[\ell] \quad (33)$$

Since the filter update path of the mFxLMS is more or less independent of the system which shall be equalized we propose to feed the update path with an independent excitation $\mathbf{S}_{\text{dec}}[\ell]$ as it is depicted in Fig. 6 (Sw_1 in left position). By this, even faster convergence can be achieved, e.g. if a white excitation $\mathbf{S}_{\text{dec}}[\ell]$ is used. An additional advantage of the decoupled input signal for the update path is the possibility of an overclocking for the filter update. Thus, the convergence speed can be further increased by the decoupled version (abbreviated here by dFxLMS) at cost of additional computational effort.

The filter update of the dFxLMS can be expressed in block frequency domain by:

for $i = 1 : O$

$$\mathbf{R}[\ell] = \mathbf{G}[\mathbf{S}_{\text{dec}}[\ell], \dots, \mathbf{S}_{\text{dec}}[\ell]]\hat{\mathbf{H}}[\ell] \quad (34)$$

$$\hat{\mathbf{Y}}[\ell] = \mathbf{G}\mathbf{S}_{\text{dec}}[\ell]\mathbf{D} \quad (35)$$

$$\mathbf{E}_{\text{EQ},\text{mod}}[\ell] = \mathbf{G}\mathbf{R}[\ell]\mathbf{C}_{\text{EQ}}^{\text{dFx}}[\ell] - \hat{\mathbf{Y}}[\ell] \quad (36)$$

$$\mathbf{C}_{\text{EQ}}^{\text{dFx}}[\ell] = \mathbf{C}_{\text{EQ}}^{\text{dFx}}[\ell - 1] + \mu\Phi_{\text{RR}}^{-1}[\ell]\mathbf{R}^H[\ell]\mathbf{E}_{\text{EQ},\text{mod}}[\ell] \quad (37)$$

end

It should be emphasized that the dFxLMS (as well as the mFxLMS) is independent of additional noise at the microphone since the microphone signal $\mathbf{Y}[\ell]$ has no influence on its update.

III. SIMULATION RESULTS

The lengths of the RIRs, the LRC filters and the AEC filters were chosen to $L_h = 4096$, $L_{c,EQ} = 1024$, and $L_{c,AEC} = 1024$ at a sampling frequency of $f_s = 8000\text{Hz}$, respectively. The room reverberation time was $\tau_{60} = 500\text{ms}$. Although in practical environments a RIR is of infinite length it can be truncated after $L_h = 4096$ samples since it is sufficiently decayed. The following simulations are given for $P = 1$ loudspeakers to lead to a spatially more robust design (compare Section II-C) and $Q = 3$ microphones. The delay introduced by the equalizer was $k_0 = 512$ samples for all channels.

Fig. 7 compares the proposed dFxLMS algorithm with the FxLMS and the mFxLMS by means of the system distance

$$D_{\text{dB}}[k] = 10 \log_{10} \frac{\|\mathbf{H}[k] \mathbf{c}_{EQ}[k] - \mathbf{d}\|^2}{\|\mathbf{d}\|^2}. \quad (38)$$

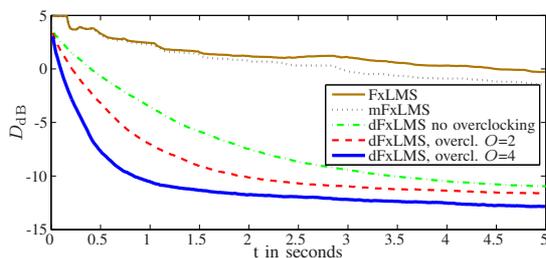


Fig. 7. Comparison of FxLMS, mFxLMS and dFxLMS (speech input).

It is obvious that FxLMS and mFxLMS algorithms perform poor since their update is based on the highly correlated speech input signal $\mathbf{S}_f[\ell]$. The mFxLMS algorithm (dotted line) performs slightly better than the conventional FxLMS algorithm. A large performance gain is achieved by the dFxLMS algorithm (dash-dotted line) even without overlocking. Please note that mFxLMS and dFxLMS algorithms are the same for speech input (switch Sw_1 in right position) and no overlocking. Thus, the distance between the dotted line (mFxLMS) and the dash-dotted line (dFxLMS, $O = 1$) is due to the white excitation used for $\mathbf{S}_{dec}[\ell]$. Further performance gain is achieved if an overlocking factor $O \geq 1$ is used as it can be seen from the lower two curves.

Simulation results for imperfect RIR estimates are shown in Fig. 8. For this purpose the RIR estimate is generated by adding white Gaussian noise to the correct RIR with different SNRs. Here the term SNR denotes the ratio between RIR power $\|\mathbf{h}[k]\|^2$ and error power $\|\tilde{\mathbf{h}}[k]\|^2$.

Obviously an imperfect estimate of the RIR leads to a decreased performance of the equalizer. However, the proposed dFxLMS algorithm still clearly outperforms FxLMS and mFxLMS algorithms. Although the dFxLMS algorithm is not more robust in terms of RIR estimation errors since it is just a fast converging version of the mFxLMS algorithm it becomes obvious from Fig. 8 that the mFxLMS algorithm as well as the FxLMS algorithm are not suitable for a real-world hands-free scenario due to their slow convergence, whereas the dFxLMS algorithm can be applied in fast-changing environments.

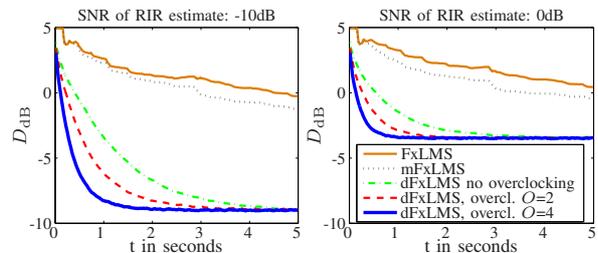


Fig. 8. Comparison of FxLMS, mFxLMS and dFxLMS for imperfect RIR estimates.

IV. CONCLUSION

In this contribution a robust algorithm for listening-room compensation is proposed which is computationally efficient and allows for fast convergence. Due to the decoupled structure of the dFxLMS algorithm it allows for feeding the update path with arbitrary input having advantageous correlation properties. By this, much faster convergence is achieved. Furthermore, the decoupled structure allows for an overlocking of the update path and, by this, the convergence speed can be further improved at the cost of additional computations.

REFERENCES

- [1] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.
- [2] M. Bouchard and S. Quednau, "Multichannel Recursive-Least-Squares Algorithms and Fast-Transversal-Filter Algorithms for Active Noise Control and Sound Reproduction Systems," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 5, pp. 606–618, Sep. 2000.
- [3] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, 1985.
- [4] E. Bjarnason, "Active Noise Cancellation using a Modified Form of the Filtered-X LMS Algorithm," in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Brussels, Belgium, 1992.
- [5] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, pp. 14–34, January 1992.
- [6] J.-S. Soo and K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Trans. on Acoustics Speech and Signal Processing*, vol. 38, no. 2, pp. 373–376, Feb 1990.
- [7] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "A Decoupled Filtered-X LMS Algorithm for Listening-Room Compensation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sep. 2008.
- [8] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "System Identification for Multi-Channel Listening-Room Compensation using an Acoustic Echo Canceller," in *Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, Trento, Italy, May 2008.
- [9] 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.
- [10] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *Journal of the Acoustical Society of America (JASA)*, vol. 66, Jul. 1979.
- [11] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Estimation of the Optimum System Delay for Speech Dereverberation by Inverse Filtering," to be published at *International Conference on Acoustics (NAG/DAGA 2009)*, Rotterdam, The Netherlands, Mar. 2009.
- [12] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb. 1988.
- [13] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Least Squares Equalizer Design under Consideration of Tail Effects," in *German Annual Conference on Acoustics (DAGA)*, Stuttgart, Germany, pp. 599–600, Mar 2007.
- [14] E. Moulines, O. Ait Amrane, and Y. Grenier, "The Generalized Multi Delay Filter: Structure and Convergence Analysis," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 14–28, Jan 1995.