A DECOUPLED FILTERED-X LMS ALGORITHM FOR LISTENING-ROOM COMPENSATION

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

*University of Bremen
Dept. of Communications Eng.
28334 Bremen, Germany

†University of Oldenburg
Signal Processing Group
26111 Oldenburg, Germany

‡University of Lübeck
Institute for Signal Processing
23538 Lübeck, Germany

ABSTRACT

In hands-free scenarios the desired speech signal picked up by the microphone is corrupted by various disturbances such as additive noise, acoustic echoes, and room reverberation. Especially the cancelation of room reverberation still remains a challenging task. For time-variant acoustic environments adaptive filters with appropriate learning algorithms based on the well-known least-mean-squares (LMS) algorithm can be used. Examples known from the field of active noise control (ANC) are the filtered-X LMS (FxLMS) or the modified filtered-X LMS (mFxLMS). In this contribution a decoupled version of the mFxLMS with a faster convergence speed will be introduced. Furthermore, an overclocking of the filter update can be applied which allows for even faster convergence at the cost of additional computational load. The new algorithm is evaluated under realistic environments including ambient noise and estimation errors of the room impulse response (RIR).

Index Terms— Listening-room compensation (LRC), acoustic equalization, acoustic echo cancelation (AEC), system identification

1. INTRODUCTION

Equalization of room impulse responses (RIRs) [1] still remains a challenging task. Since a RIR is a mixed-phase system, in general, only its minimum-phase component can be inverted by a causal stable IIR filter [2]. Thus, finite-length FIR filters can be applied that minimize the mean squared error between the signal at the reference microphone and a given desired signal which usually is a delayed or filtered version of the unreverberated sound signal [3]. The straightforward minimization of the system distance of the equalized system $h[k] * c_{\mathrm{EQ}}[k]$ and the desired target system $d[k]$ leads to the well-known least-squares equalizer which implies an inversion of the channel convolution matrix that usually has a dimension of several thousand coefficients [4]. Even for minor RIR changes the equalizer (EQ) filter has to be recalculated [5]. Since RIRs are time-varying, e.g. due to changes in the acoustic environment or changes caused by moving speakers, the EQ has to be updated frequently. Thus, an adaptive algorithm for the equalizer is necessary. Prominent learning algorithms for the EQ filter update are the filtered-X least-mean-squares (FxLMS) algorithm [6, p. 280ff] or the modified filtered-X LMS (mFxLMS) [7]. Fast variants based on the recursive least-squares (RLS) algorithm exist [8] which cause high computational load and might suffer from stability problems.

In this contribution we introduce a decoupled version of the mFxLMS. This new algorithm 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 overclocking the filter update because it is also independent of the number of samples of the original input signal. Thus, a tradeoff between convergence speed and complexity can be utilized to adapt it to the available processing power.

Work supported by German Research Foundation DFG (Ka841-17). The authors would like to thank P. Gessner for fruitful discussions.

The remainder of this paper is organized as follows: Section 2 briefly reviews the FxLMS and the mFxLMS and introduces the decoupled FxLMS. Section 3 discusses disturbances which may occur in a real-world scenario, such as additive noise or imperfect estimates of the RIR. Section 4 presents some simulation results and Section 5 concludes the paper.

Notation: Vectors and matrices are printed in boldface while scalars are printed in italic. k is the discrete time index. The superscripts ^{T} and * denote the transposition and the complex conjugation, respectively. The operator $*$ denotes the convolution of two sequences and the operator convmtx$\{\mathbf{h}, L_c\}$ generates a convolution matrix of size $(L_c + L_h - 1) \times L_c$.

2. LISTENING-ROOM COMPENSATION

In a common setup for listening-room compensation the equalization filter precedes the acoustic channel. The goal of the equalizer is to remove reverberation which is caused by the convolution of the loudspeaker signal with the RIR at the position of the reference microphone where the user of the system is assumed to be located. Spatial robustness with respect to the distance between reference microphone and the user of the LRC system can be increased by multi-microphone systems [4] and will not be considered in this contribution.

2.1. Filtered-X LMS and modified FxLMS

Since the RIR between loudspeaker and microphone is time-variant, in general, the RIR identification as well as the RIR equalization has to be done adaptively. A benchmark for adaptive equalization known from active noise control (ANC) systems is the FxLMS [6] algorithm which is depicted in Fig. 1.

The unreverberated speech signal $s[k]$ is fed through the room equalization filter $c_{\mathrm{EQ}}[k]$ which precedes the acoustic channel $h[k]$. The aim of the equalizer is to minimize the distance between the equalized system $h[k] * c_{\mathrm{EQ}}[k]$ and a desired target system $d[k]$. The input signal of the LMS update path has to be filtered with the acoustic channel $h[k]$ [6] which is not available in real world systems. Thus, an estimate $\hat{h}[k]$ of the RIR (known as the plant model in ANC systems) is needed. Since changes of the filter coefficients $c_{\mathrm{EQ}}[k]$ do not have an immediate impact on the error signal $e_{\mathrm{EQ}}[k]$

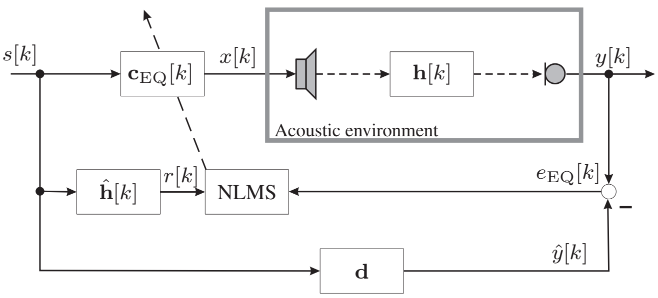

Fig. 1. Block diagram of filtered-X LMS (FxLMS).

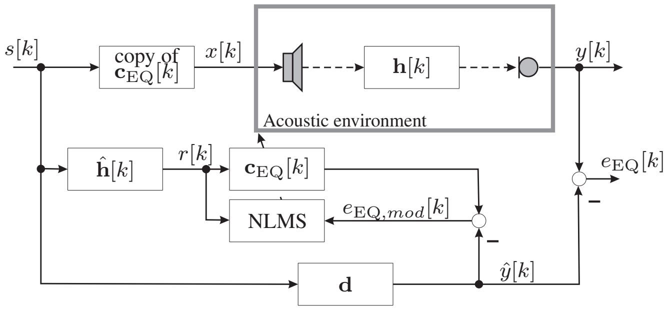

Fig. 2. Block diagram of modified Filtered-X LMS (mFxLMS).

due to the delay of the RIR a small stepsize μ for the filter update is required to ensure stability. Thus, especially if a large filter length is required the FxLMS algorithm suffers from slow convergence.

The FxLMS algorithms can be written in matrix/vector notation as follows:

Algorithm 1 Filtered-X LMS (FxLMS)

1: $\mathbf{r}[k] = \hat{\mathbf{H}}^T[k]\mathbf{s}_{\mathrm{I}}[k]$
2: $e_{\mathrm{EQ}}[k] = \mathbf{s}_{\mathrm{II}}^T[k]\mathbf{H}[k]\mathbf{c}_{\mathrm{EQ}}[k] - \mathbf{s}_{\mathrm{II}}^T[k]\mathbf{d}$
3: $\mathbf{c}_{\mathrm{EQ}}[k+1] = \mathbf{c}_{\mathrm{EQ}}[k] + \mu\mathbf{r}[k]e_{EQ}[k]$

Here $\mathbf{c}_{\mathrm{EQ}}[k]$ is the coefficient vector of the equalizer, $\mathbf{H}[k]$ and $\hat{\mathbf{H}}[k]$ are the convolution matrices of the RIR and its estimate, respectively, $\mathbf{d}$ is the desired system vector, and $\mathbf{s}[k]$ is the input signal vector given here for two different lengths:

$$\mathbf{c}_{\mathrm{EQ}}[k] = \left[\, c_{\mathrm{EQ},0}[k],\, c_{\mathrm{EQ},1}[k],\, \dots,\, c_{\mathrm{EQ},L_{c,\mathrm{EQ}}-1}[k] \,\right]^T \tag{1}$$

$$\mathbf{H}[k] = \text{convmtx}\left\{[h_0[k], h_1[k], ..., h_{L_h}[k]]^T, L_{c,\mathrm{EQ}}\right\} \tag{2}$$

$$\hat{\mathbf{H}}[k] = \text{convmtx}\left\{\left[\hat{h}_0[k], \hat{h}_1[k], ..., \hat{h}_{L_{\hat{h}}}[k]\right]^T, L_{c,\mathrm{EQ}}\right\} \tag{3}$$

$$\mathbf{d} = [\underbrace{0,\, ...,\, 0}_{k_0},\, d_0,\, d_1,\, ...,\, d_{L_d-1},\, \underbrace{0,\, ...,\, 0}_{L_h+L_{c,\mathrm{EQ}}-1-L_d-k_0}\,]^T \tag{4}$$

$$\mathbf{s}_{\mathrm{I}}[k] = [\, s[k],\, \dots,\, s[k - L_{\hat{h}} - L_{c,\mathrm{EQ}} + 2]\,]^T \tag{5}$$

$$\mathbf{s}_{\mathrm{II}}[k] = [\, s[k],\, \dots,\, s[k - L_h - L_{c,\mathrm{EQ}} + 2]\,]^T \tag{6}$$

The lengths of the RIR, the RIR estimate, the LRC filter and the desired system are denoted by L_h, $L_{\hat{h}}$, $L_{c,\mathrm{EQ}}$, and L_d, respectively. k_0 is the delay of the desired target system that will be introduced by the equalizer.

To overcome the problem of a heavily reduced allowed convergence speed in comparison to the conventional LMS algorithm the modified filtered-X LMS (mFxLMS) algorithm [7] was introduced which is depicted in Fig. 2. In contrast to the FxLMS the calculation of the error signal of the modified FxLMS $e_{\mathrm{EQ},mod}[k]$ is independent of the room impulse response and thus independent of the microphone signal $y[k]$. Instead, the filter update is based on the RIR estimate $\hat{h}[k]$ only. By this, the equalizer $\mathbf{c}_{\mathrm{EQ}}[k]$ succeeds the RIR estimate $\hat{h}[k]$ and for the case of a correct system identification ($\hat{h}[k] = h[k]$) the convergence performance of the mFxLMS is the same as for the conventional LMS algorithm because the update of the filter coefficients has direct impact on the error signal $e_{\mathrm{EQ},mod}[k]$.

The mFxLMS algorithm can be summarized as follows:

Algorithm 2 Modified filtered-X LMS (mFxLMS)

1: $\mathbf{r}[k] = \hat{\mathbf{H}}^T[k]\mathbf{s}_{\mathrm{I}}[k]$
2: $e_{\mathrm{EQ},mod}[k] = \mathbf{r}^T[k]\mathbf{c}_{\mathrm{EQ}}[k] - \mathbf{s}_{\mathrm{I}}^T[k]\mathbf{d}$
3: $\mathbf{c}_{\mathrm{EQ}}[k+1] = \mathbf{c}_{\mathrm{EQ}}[k] + \mu'\mathbf{r}[k]e_{\mathrm{EQ},mod}[k]$

2.2. A decoupled version of modified filtered-X LMS allowing for an update overclocking

The mFxLMS depicted in Fig. 2 and described by Algorithm 2 allows for a larger stepsize than the conventional FxLMS and thus for faster convergence. Since the filter update path is more or less independent of the system which should be equalized (the error signal in line 2 of Algorithm 2 is independent of the real RIR and only the estimate or model is used) we propose to feed the update path with an independent excitation $s_{dec}[k]$ as it is depicted in Fig. 3.

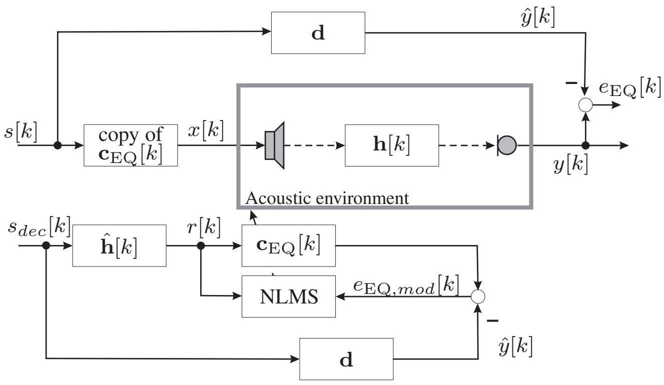

Fig. 3. Block diagram of decoupled Filtered-X LMS.

By this, even faster convergence can be achieved since the excitation signal of the update path $s_{dec}[k]$ can be chosen arbitrarily. Best convergence would be archived for so-called *perfect sequences* as an input signal for the NLMS $r[k]$. Since generation of such a signal by crafting the excitation signal $s_{dec}[k]$ at the input of the RIR estimate is difficult we propose to use a white Gaussian excitation for $s_{dec}[k]$. An additional advantage of the proposed algorithm is the fact that with a decoupled input signal for the update path an overclocking of the filter update is possible and by this a tradeoff between convergence speed and computational complexity. The proposed decoupled filtered-X least-mean-squares (dFxLMS) algorithm is summarized in Algorithm 3. Here, $O \geq 1, O \in \mathbb{N}$ is the overclocking factor. The term overclocking does not mean oversampling since this would affect both, the sampling rate of the signals and the systems. The sampling rate of the systems, such as that of $\hat{\mathbf{h}}[k]$, is unchanged while more than one input sample $s_{dec}[k]$ is processed before copying the filter weights $\mathbf{c}_{\mathrm{EQ}}[k]$ to the upper branch.

Algorithm 3 Decoupled version of modified filtered-X LMS

```
for i = 0 : O − 1 do
    r[k + i] = Ĥ^T[k] s_dec[k + i]
    e_EQ,mod[k + i] = r^T[k + i] c_EQ[k + i] − s_dec^T[k + i] d
    c_EQ[k + i + 1] = c_EQ[k + i] + μ' r[k + i] e_EQ,mod[k + i]
end for
Copy updated EQ coefficients c_EQ[k + i + 1] to upper branch
```

3. ERROR INFLUENCES

Two kinds of error influences may have impact on the equalization filter. The first is additive local disturbance at the microphone (such as ambient noise $n[k]$ or an *interfering* local speaker $s_n[k]$) and the second is the estimation mismatch between the RIR $h[k]$ that shall be equalized and its estimate $\hat{h}[k]$.

Since the mFxLMS and the dFxLMS described in the previous section work independently from the microphone signal $y[k]$ even strong local disturbances present at the microphone have no influence on the filter adaptation. This property is advantageous compared to the conventional FxLMS algorithm, especially for a hands-free scenario with a high background noise level and competing speakers. It should be mentioned that strong disturbances at the microphone of course have negative influence on the RIR identification which is needed by all the algorithms (see next paragraph).

System Identification

All algorithms described so far rely on an estimate $\hat{\mathbf{h}}[k]$ of the acoustic channel (the room impulse response $\mathbf{h}[k]$). As described in [4] adaptive tracking of the time-variant RIR is necessary and, thus, it is inevitable that estimation errors occur, e.g. in periods of initial convergence or after RIR changes. Acoustic echo cancelers (AECs) estimate and subtract the acoustic echo $\psi[k]$ from the microphone signal $y[k]$ by performing system identification. Thus, the AEC filter coefficients $\mathbf{c}_{\mathrm{AEC}}[k]$ can be used as an estimate for the room impulse response ($\hat{\mathbf{h}}[k] = \mathbf{c}_{\mathrm{AEC}}[k]$). As illustrated in Fig. 4 the RIR $\mathbf{h}[k]$ can be split up into one part $\hat{\mathbf{h}}[k]$ which is correctly identified by the acoustic echo canceller (AEC) and an estimation error $\tilde{\mathbf{h}}[k]$:

$$\mathbf{h}[k] = \begin{bmatrix} \hat{\mathbf{h}}[k] \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{h}}[k] = \begin{bmatrix} \mathbf{c}_{\mathrm{AEC}}[k] \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{h}}[k] \tag{7}$$

with

$$\mathbf{h}[k] = [h_0[k]\,,\, h_1[k]\,,\, \ldots\,,\, h_{L_h-1}[k]]^T \tag{8}$$

$$\hat{\mathbf{h}}[k] = \left[c_{\mathrm{AEC},0}[k],\, c_{\mathrm{AEC},1}[k],\, \ldots,\, c_{\mathrm{AEC},L_{c,\mathrm{AEC}}-1}[k]\right]^T \tag{9}$$

$$\tilde{\mathbf{h}}[k] = \left[\tilde{h}_0[k]\,,\, \tilde{h}_1[k]\,,\, \ldots\,,\, \tilde{h}_{L_{\tilde{h}}-1}[k]\right]^T \tag{10}$$

Here, $L_{c,\mathrm{AEC}}$ is the length of the AEC filter which equals $L_{\tilde{h}}$ and is, in general, less 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 $\tilde{\mathbf{h}}[k]$ and leads to a decreased performance of the equalizer [9].

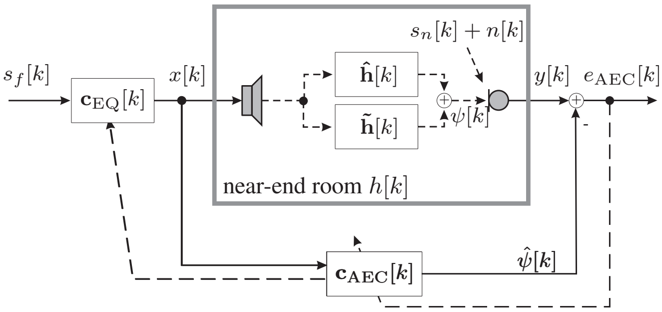

Fig. 4. Combined system of equalizer and acoustic echo canceler. The RIR can be split into a part modeled by the AEC $\mathbf{c}_{\mathrm{AEC}}[k]$ and the AEC system misalignment $\tilde{\mathbf{h}}[k]$.

4. SIMULATION RESULTS

RIRs of length $L_h = 4096$ with a reverberation time of $\tau_{60} = 500ms$ were generated [10] at a sampling frequency of $f_s = 8$kHz. The length of the acoustic echo canceler and the room equalization filter were set to $L_{c,\mathrm{AEC}} = 2048$ and $L_{c,\mathrm{EQ}} = 1024$, respectively. The desired target system $d[k]$ was chosen as a 40th order FIR high-pass with 3dB frequency at 200Hz. The delay introduced by the equalizer was $k_0 = 511$ samples. The algorithms were implemented in a partitioned frequency domain [11, 12] to reduce the computational load and the delay introduced by the system.

Evaluation of the algorithms is done by means of the normalized system distance

$$D_{\mathrm{dB}}[k] = 10\log_{10} \frac{||\mathbf{H}[k]\mathbf{c}_{\mathrm{EQ}}[k] - \mathbf{d}||^2}{||\mathbf{d}||^2} \tag{11}$$

and the segmental signal to reverberation ratio (SSRR) [13]

$$\mathrm{SSRR}[\ell] = 10\log_{10} \frac{\sum_{k=0}^{L_{Bl}-1} \hat{y}[\ell L_{Bl} + k]^2}{\sum_{k=0}^{L_{Bl}-1} (\hat{y}[\ell L_{Bl} + k] - y[\ell L_{Bl} + k])^2} \tag{12}$$

Here $L_{Bl} = 128$ is the block length, and ℓ is the block index.

Fig. 5 compares the convergence behavior of the modified filtered-X least-mean-squares (mFxLMS) and the dFxLMS with an overclocking of 2 and 4 times the block length for a white input signal. Please note that for a white noise input signal and no overclocking the dFxLMS equals the mFxLMS. For this simulation the RIR estimate was set to the correct RIR but with a reduced length of $L_{\hat{h}} = L_{c,\mathrm{AEC}} = 2048$. It can be seen that the convergence speed can be increased drastically exploiting the overclocking capabilities of the decoupled structure introduced in Section 2.2.

Fig. 6 shows the performance gain which is possible due to the decoupling of input signal and excitation signal for the filter update path of the dFxLMS. The dashed line shows the convergence behavior of the algorithm for $s_{dec}[k] = s[k]$ if $s[k]$ is speech input (male speaker). No overclocking is performed, thus the convergence equals that of the mFxLMS. The solid line shows the performance for a white noise as excitation for the update path $s_{dec}[k]$. The performance of the equalizer can be drastically increased compared to a coupled structure (such as the mFxLMS) if a speech signal is the input which is the common case for all practically relevant systems. If a white noise signal is used for the update path signal $s_{dec}[k]$ instead of the speech input the algorithm converges much faster.

In Fig. 7 simulations including a RIR estimation error are performed. For this purpose the RIR estimate is generated by adding

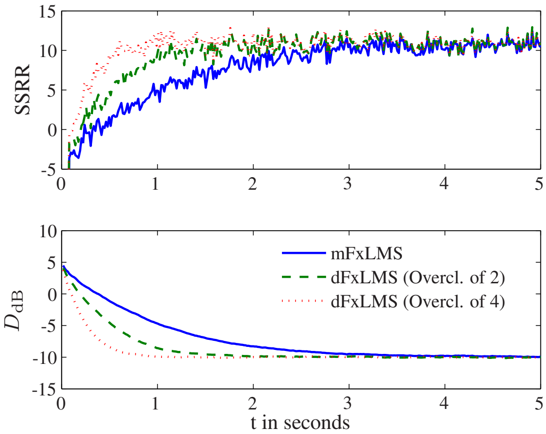

Fig. 5. Comparison of mFxLMS and dFxLMS with overclocking for white input signals. The blue solid line shows convergence of the mFxLMS which equals the convergence of a dFxLMS without overclocking.

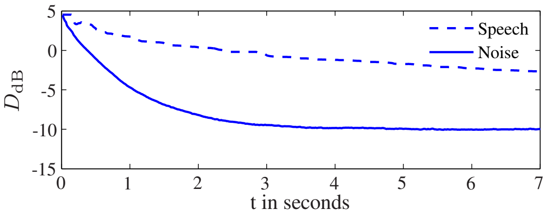

Fig. 6. Performance of the dFxLMS for a speech input and a white input signal $s_{dec}[k]$ used for the update path.

white Gaussian noise to the correct RIR estimate at different SNRs. Fig. 7 shows the convergence curves for the three algorithms discussed above for SNR ratios between the true RIR $\mathbf{h}[k]$ and the estimation error $\tilde{\mathbf{h}}[k]$ of 20dB (solid lines), 10dB (dashed lines) and 0dB (dotted lines). Here the term SNR denotes the ratio between RIR power $||\mathbf{h}[k]||^2$ and error power $||\tilde{\mathbf{h}}[k]||^2$. It is clearly visible that the dFxLMS algorithm (lower three curves) outperforms both the FxLMS and the mFxLMS, even for an SNR of 0dB. The curves for the FxLMS and the mFxLMS algorithms for a speech input signal (marked by the upper circle in Figure 7) are not clearly distinguishable because they lie closely together for all SNRs. This is due to the fact that the performance loss due to the correlated input signal is dominant. The dFxLMS update path still works with a white excitation signal. Thus, the lower solid line in Fig. 7 is about the same as the solid line in Fig. 6. Please note that the simulation results in Figure 7 do not indicate a higher robustness to possible RIR estimation errors. However, the dFxLMS is capable of tracking changes of the RIR model much faster.

5. CONCLUSION

In this contribution a decoupled version of the modified filtered-X LMS algorithm for listening-room compensation was introduced and analyzed. Due to the decoupled structure of the new algorithm a far better convergence behavior can be achieved compared to the conventional FxLMS and the mFxLMS. This is due to the fact that the filter update path can be driven by an appropriate excitation signal. An overclocking to further increase the performance is possible. The proposed filter structure is easy to implement and computationally efficient.

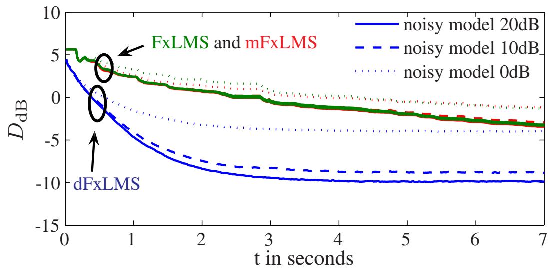

Fig. 7. Influences of a noisy RIR model / erroneous RIR estimate on the equalizer performance.

6. REFERENCES

[1] J. N. Mourjopoulos, "Digital Equalization of Room Acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[2] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *Journal of the Acoustical Society of America (JASA)*, vol. 66, pp. 165–169, Jul. 1979.

[3] 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.

[4] 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, pp. 224–227, May 2008.

[5] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an Acoustic Reverberant Environment: Robustness Results," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

[6] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, 1985.

[7] 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, pp. 1053–1056, 1992.

[8] 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.

[9] 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, March 2007.

[10] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[11] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Enhanced Partitioned Stereo Residual Echo Estimation," in *Asilomar 2006 - Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 1326–1330, Oct. 2006.

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

[13] P. Naylor and N. Gaubitch, "Speech Dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sep. 2005.