A SPATIALLY ROBUST LEAST SQUARES CROSSTALK CANCELLER

*Markus Kallinger**

Fraunhofer Institute for Integrated Circuits
Erlangen, Germany
markus.kallinger@iis.fraunhofer.de

*Alfred Mertins**

Institute for Signal Processing
University of Lübeck, Germany
mertins@isip.uni-luebeck.de

ABSTRACT

Crosstalk cancellation is a well-known technique to generate virtual 3D sound via loudspeakers. Usually, headphones are used to playback audio material, which has been filtered with HRTFs. Audio playback in a reverberant room using loudspeakers and crosstalk cancellation is our intended application. This raises the need for a robust design, since listeners might slightly move their heads during a listening session. The novel design is based on a known least squares crosstalk canceller design technique, but with added robustness. The robustness is achieved with the help of assumed stochastic perturbation systems, which lie in parallel to the actual propagation impulse responses, during the design process.

Index Terms— Crosstalk, Deconvolution, Equalizers, Robustness, Spatial filters

1. INTRODUCTION

Crosstalk cancellation for acoustic systems is a traditional task in signal processing. Early approaches assume symmetric propagation paths and usually aim at the equalization of head related transfer functions (HRTFs) and the cancellation of crosstalk paths [1]. Subsequent designs were carried out in the DFT-domain. Signal propagation from two loudspeakers to two microphones is described by 2×2-matrices – each entry is a transfer function. Crosstalk cancellation could be separated into two stages: first, the crosstalk paths are perfectly cancelled by designing a 2×2-matrix, which contains the adjugate, i. e. the inverse multiplied with the determinant, of the 2×2-propagation path matrix. However, the remaining task of equalizing the determinant is very demanding, since it contains the difference of two pairs of convoluted impulse responses [2, 3].

Nelson et al. [4] proposed a least squares crosstalk canceller to achieve both equalization and crosstalk cancellation in one step. This technique has been extended by Ward [5], who made a common design for more than one head position simultaneously, which resulted in good spatial robustness. In this paper we try to achieve a similar goal on the basis of statistical knowledge of acoustic transfer functions [6].

The paper is organized as follows: Section 2 contains a generalized derivation of the least squares crosstalk canceller according to Nelson et al. and Ward, respectively. Section 3 illustrates the novel, spatially robust design. Simulation results are given in Section 4, and Section 5 concludes the paper.

Notation Vectors (lowercase) and matrices (uppercase) are printed in boldface. The superscripts ^{T} and * denote transposition and complex conjugation, respectively. The asterisk $*$ denotes convolution. The operator $\mathrm{diag}\{\cdot\}$ turns a vector into a diagonal matrix and $\|\cdot\|$ returns the ℓ_2-norm. Given a vector $\mathbf{c}$ of length L_c, the operator convmtx($\mathbf{c}$, L_h) generates a convolution matrix of size $(L_h+L_c-1)\times L_h$.

2. GENERALIZED LEAST-SQUARES CROSSTALK CANCELLER

We investigate a 2×2-crosstalk canceller as shown in Fig. 1. The responses $h_{il}[n]$ contain the crosstalk canceller coefficients, $c_{il}[n]$ are responses from the loudspeakers to the ears and $p_{il}[n]$ are stochastic perturbations systems, which are added to $c_{il}[n]$ to simulate estimation errors. $d_i[n]$ are desired responses.

We search for a least squares solution of a difference system $\mathbf{e}_{\mathrm{st}}$ in terms of the equalizer coefficients, which are all collected into a stacked vector $\mathbf{h}_{\mathrm{st}}$:

$$\begin{aligned}\mathbf{e}_{\mathrm{st}} &= \tilde{\mathbf{C}}_{\mathrm{st}}\mathbf{h}_{\mathrm{st}} - \mathbf{d}_{\mathrm{st}} \\ &= (\mathbf{C}_{\mathrm{st}} + \mathbf{P}_{\mathrm{st}})\,\mathbf{h}_{\mathrm{st}} - \mathbf{d}_{\mathrm{st}} \qquad (1)\end{aligned}$$

using

$$\mathbf{h}_{\mathrm{st}} = \left[\mathbf{h}_{11}^T, \mathbf{h}_{12}^T, \mathbf{h}_{21}^T, \mathbf{h}_{22}^T\right]^T, \qquad (2)$$

$$\mathbf{h}_{il} = \left[h_{il}[0], \ldots, h_{il}[L_h-1]\right]^T, \qquad (3)$$

$$\mathbf{C}_{\mathrm{st}} = \begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{bmatrix}, \qquad (4)$$

$$\mathbf{C} = \begin{bmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{bmatrix}, \qquad (5)$$

*Both authors have performed the present work at the Signal Processing Group, University of Oldenburg, Germany.

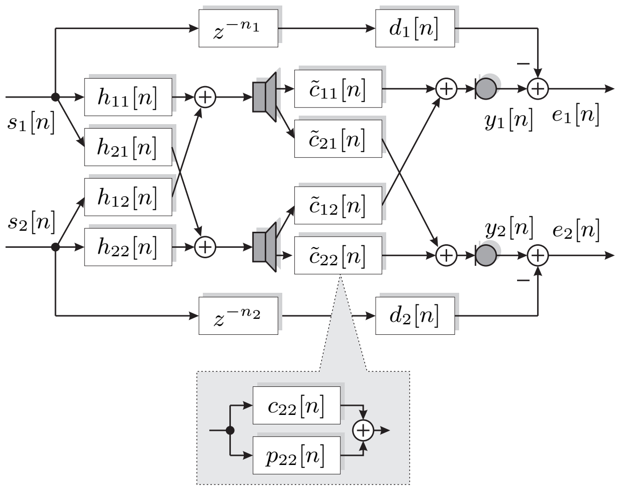

Fig. 1. 2×2-setup of a crosstalk canceller. For illustration purposes the generation of the error signals is also shown.

$$\mathbf{C}_{il} = \text{convmtx}\Big(\left[c_{il}[0], \ldots, c_{il}[L_c - 1]\right]^T, L_h\Big), \qquad (6)$$

$$\mathbf{P}_{\text{st}} = \begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix}, \qquad (7)$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}, \qquad (8)$$

$$\mathbf{P}_{il} = \text{convmtx}\Big(\left[p_{il}[0], \ldots, p_{il}[L_c - 1]\right]^T, L_h\Big), \qquad (9)$$

$$\mathbf{d}_{\text{st}} = \left[\mathbf{d}_1^T, \mathbf{0}^T, \mathbf{0}^T, \mathbf{d}_2^T\right]^T \qquad (10)$$

and

$$\mathbf{d}_i = \qquad (11)$$

$$\Bigg[\underbrace{0, \ldots, 0}_{n_i}, d_i[0], \ldots, d_i[L_{\text{BP}} - 1], \underbrace{0, \ldots, 0}_{L_h + L_c - 1 - L_d - n_i}\Bigg]^T.$$

The aim of the design is to minimize the mean squared error $\mathrm{E}\{\|\mathbf{e}_{\text{st}}\|^2\}$. The expectation operator is used to account for the stochastic perturbations described by $\mathbf{P}$. This way of expressing an estimation uncertainty had been used in [7] during the design of channel-shortening filters.

Generalization can be obtained by weighting of each desired path:

$$\mathbf{e}_{\text{st}} = \mathbf{W}\left(\left(\mathbf{C}_{\text{st}} + \mathbf{P}_{\text{st}}\right)\mathbf{h}_{\text{st}} - \mathbf{d}_{\text{st}}\right) \qquad (12)$$

with

$$\mathbf{W} = \text{diag}\left\{\left[\mathbf{w}_{11}^T, \mathbf{w}_{21}^T, \mathbf{w}_{12}^T, \mathbf{w}_{22}^T\right]\right\}. \qquad (13)$$

Vectors $\mathbf{w}_{il}$ represent individual weights of cancellers and equalized paths, respectively. In this paper, each vector $\mathbf{w}_{il}$ only contains constant elements w_{il}. Starting the derivation of the generalized least squares crosstalk canceller we want to minimize

$$\mathrm{E}\left\{\|\mathbf{e}_{\text{st}}\|^2\right\} = \mathrm{E}\Big\{\left(\left(\mathbf{C}_{\text{st}} + \mathbf{P}_{\text{st}}\right)\mathbf{h}_{\text{st}} - \mathbf{d}_{\text{st}}\right)^T \mathbf{W}^T\mathbf{W}\left(\left(\mathbf{C}_{\text{st}} + \mathbf{P}_{\text{st}}\right)\mathbf{h}_{\text{st}} - \mathbf{d}_{\text{st}}\right)\Big\}. \qquad (14)$$

For further conversions we assume that all stochastic perturbations are zero-mean:

$$\mathrm{E}\{p_{il}[n]\} = 0. \qquad (15)$$

The derivative of $\mathrm{E}\left\{\|\mathbf{e}_{\text{st}}\|^2\right\}$ with respect to $\mathbf{h}_{\text{st}}$ amounts to:

$$\frac{\partial \mathrm{E}\left\{\|\mathbf{e}_{\text{st}}\|^2\right\}}{\partial \mathbf{h}_{\text{st}}} = 2\mathbf{C}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{C}_{\text{st}}\mathbf{h}_{\text{st}} + 2\mathrm{E}\left\{\mathbf{P}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\text{st}}\right\}\mathbf{h}_{\text{st}} - 2\mathbf{C}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{d}_{\text{st}}. \qquad (16)$$

Setting this expression to zero and solving for $\mathbf{h}_{\text{st}}$ yields:

$$\mathbf{h}_{\text{st}} = \left(\mathbf{C}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{C}_{\text{st}} + \mathrm{E}\left\{\mathbf{P}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\text{st}}\right\}\right)^{-1} \mathbf{C}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{d}_{\text{st}}. \qquad (17)$$

Referring to equations (4) and (7), the first and the second half of vector $\mathbf{h}_{\text{st}}$ can be calculated separately. The required matrix inversion does not have to be carried out at once – the inversion of the upper left and lower right quadrants suffices. Taking a closer look at $\mathrm{E}\left\{\mathbf{P}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\text{st}}\right\}$, we see that, in our case, each $\mathbf{w}_{il}$, which makes up the diagonal matrix $\mathbf{W}$, is constant. Therefore, the weights can be varied for emphasized optimization of ipsilateral or contralateral paths.

3. INCORPORATING SPATIAL ROBUSTNESS

In the following, we want to choose $\mathrm{E}\left\{\mathbf{P}_{\text{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\text{st}}\right\}$ in a way to increase the spatial robustness of the investigated LS-crosstalk canceller. We refer to a statistical point of view, which was treated by Radlović et al. [6]. If the following conditions are met, the transfer function between a loudspeaker and a microphone is a stochastic one:

- The dimensions of the room must be large compared the wavelengths of interest. This is true especially for speech signals.
- Statistical assumptions can be met for frequencies above the Schroeder large-room frequency

$$f_{\text{SL}} = 2000\sqrt{\frac{\tau_{60}}{V}}\ \text{Hz}. \qquad (18)$$

Our room possesses a Schroeder large-room frequency of 215 Hz (for dimensions, please refer to Section 4).

- All loudspeakers and microphones should have a distance of at least half a wavelength to adjacent walls.

Given these assumptions Radlović et al. defined a frequency dependent distance measure of

$$Q_{il}(\Omega) = \mathrm{E}\left\{\left|\tilde{C}_{il}(\Omega)H_{il}(\Omega) - 1\right|^2\right\}. \qquad (19)$$

With $\tilde{C}_{il}(\Omega) = C_{il}(\Omega) + P_{il}(\Omega)$ and an assumed perfect equalizer $H_{il}(\Omega) = 1/C_{il}(\Omega)$ we want to calculate $\mathrm{E}\{|P_{il}(\Omega)|^2\}$ in order to insert its time domain equivalent into equation (17). Ω denotes the angular frequency. The distance measure amounts to

$$Q_{il}(\Omega) \cong 2 - 2\frac{\sin(\Omega f_s R/c)}{\Omega f_s R/c}, \qquad (20)$$

in the far field in reverberant environments [6]. f_s represents the sampling frequency in Hz, R is the deviation of the microphone from its assumed position in m, and c is the sound-propagation velocity (340 m/s).

$$\begin{aligned} Q_{il}(\Omega) &= \mathrm{E}\left\{\left|\frac{C_{il}(\Omega) + P_{il}(\Omega)}{C_{il}(\Omega)} - 1\right|^2\right\} \\ &= \frac{|C_{il}(\Omega)|^2 + \mathrm{E}\{|P_{il}(\Omega)|^2\}}{|C_{il}(\Omega)|^2} + 1 \\ &\quad - 2\Re\left\{\frac{C_{il}(\Omega) + \mathrm{E}\{P_{il}(\Omega)\}}{C_{il}(\Omega)}\right\} \\ &= \frac{|C_{il}(\Omega)|^2 + \mathrm{E}\{|P_{il}(\Omega)|^2\}}{|C_{il}(\Omega)|^2} - 1 \\ \mathrm{E}\{|P_{il}(\Omega)|^2\} &= |C_{il}(\Omega)|^2\left(2 - 2\frac{\sin(\Omega f_s R/c)}{\Omega f_s R/c}\right). \qquad (21) \end{aligned}$$

In order to get an equivalent of $\mathrm{E}\{|P_{il}(\Omega)|^2\}$ in the time domain, we first sample $Q_{il}(\Omega)$ at K discrete frequencies Ω_k. A succeeding inverse discrete Fourier transform (IDFT) delivers the discrete series $q_{il}[n]$, which in turn is needed to receive

$$\begin{aligned} r_{pp,il}[n] &= r_{cc,il}[n] * q_{il}[n] \qquad (22) \\ &= (c_{il}[n] * c_{il}[-n]) * q_{il}[n] \\ &\longleftrightarrow \mathrm{E}\{|P_{il}[\Omega_k]|^2\}. \end{aligned}$$

For equation (17) we need to calculate $\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}$. As a simplification we have considered the four partitions of $\mathbf{W}$ to be identity matrices being weighted by a scalar. Therefore, the two diagonal partitions of $\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}$ are almost identical except for their different weights (see equation (7)). In the following, let us take a closer look at the upper left partition, $[\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}]_{11}$. It possesses four quadrants, too. The upper left one is a symmetric Toeplitz matrix of size $L_h \times L_h$, made up by the vector

$$\begin{bmatrix} w_{11}r_{pp,11}[0] + w_{21}r_{pp,21}[0] \\ \vdots \\ w_{11}r_{pp,11}[L_h-1] + w_{21}r_{pp,21}[L_h-1] \end{bmatrix}$$

and the lower right one is symmetric Toeplitz, as well. It is defined by vector

$$\begin{bmatrix} w_{11}r_{pp,12}[0] + w_{21}r_{pp,22}[0] \\ \vdots \\ w_{11}r_{pp,12}[L_h-1] + w_{21}r_{pp,22}[L_h-1] \end{bmatrix}.$$

The two off-diagonal quadrants of $[\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}]_{11}$ contain zeros for the following reason: the off-diagonal quadrants only contain cross spectral densities of the stochastic perturbation like $\mathrm{E}\{P_{11}^*[\Omega_k]P_{12}[\Omega_k]\}$. Perturbations of different acoustic paths are assumed to be uncorrelated.

Accordingly, the diagonal quadrants of $[\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}]_{22}$ are symmetric Toeplitz and made up by

$$\begin{bmatrix} w_{12}r_{pp,11}[0] + w_{22}r_{pp,21}[0] \\ \vdots \\ w_{12}r_{pp,11}[L_h-1] + w_{22}r_{pp,21}[L_h-1] \end{bmatrix}$$

and

$$\begin{bmatrix} w_{12}r_{pp,12}[0] + w_{22}r_{pp,22}[0] \\ \vdots \\ w_{12}r_{pp,12}[L_h-1] + w_{22}r_{pp,22}[L_h-1] \end{bmatrix},$$

respectively. The two off-diagonal partitions of $\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}$ contain only zeros. Finally, we get $\mathrm{E}\{\mathbf{P}_{\mathrm{st}}^T\mathbf{W}^T\mathbf{W}\mathbf{P}_{\mathrm{st}}\}$ with 16 quadratic, uniformly sized partitions, and only those on the diagonal are non-zero.

4. SIMULATION RESULTS

For the following simulations we have measured impulse responses in an office of size 2.63 m × 3.10 m × 4.22 m (height × width × length) with a reverberation time of 400ms. Sources were two Genelec 8020 loudspeakers. Two in-ear microphones were placed into the ears of a 4128C head and torso simulator by Brüel&Kjær, "ear types" 4158C and 4159C. Both speakers had a distance of 1.7 m to the back wall and 1.0 m to each other. The torso was located with a distance of 1.0 m to the line between the loudspeakers, i. e. we had a distance of 1.41 m from each loudspeaker to the torso. Measurements using maximum-length sequences have been carried out at 44.1 kHz sampling frequency; all impulse responses have been resampled to $f_s = 12$ kHz.

We used $L_c = L_p = 1000$ samples for the design of a $L_h = 1000$ crosstalk canceller. Evaluation of the designed crosstalk canceller was carried out with longer impulse responses ($L_c = 1800$). The delay in front of the desired impulse responses $d_1[n]$ and $d_2[n]$ was set to $n_1 = n_2 = 150$. Both desired impulse responses were chosen to be bandpass filters. We took a recursive Butterworth design in order to achieve a causal character for $d_1[n]$ and $d_2[n]$, respectively. At an order of 10, cutoff frequencies were located at 200 Hz and 5.5 kHz. The infinite impulse response was truncated to a length of $L_{\mathrm{BP}} = 400$. All w_{il} were set to 0.5.

Fig. 2 shows measured magnitude transfer functions (black) and processed magnitude transfer functions (gray). They describe the propagation from the sources $s_1[n]$ and $s_2[n]$, respectively, to the "left ear", i. e. to signal $y_1[n]$. We can observe the impact of equalization in the upper left plot (transfer function of ipsilateral path). The lower left plot shows the equalization performance of a spatially robust design with an assumed microphone deviation of $R = 2$ cm (ipsilateral path). The equalizer causes a slight lowpass character. The crosstalk cancellation performance of both designs is rather equal (contralateral path, right plots).

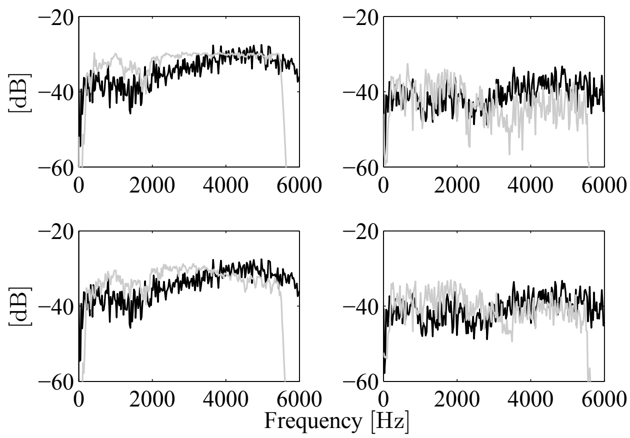

Fig. 2. Magnitude transfer functions (black) and processed magnitude transfer functions (gray). The upper left plot shows the ipsilateral path to the left ear and the upper right one the contralateral path to the left ear. For the upper plots we used a non-robust design and a robust design for the lower plots (again, ipsilateral path to the left ear on the lower left and contralateral path to the left ear on the lower right plot). The room impulse responses were not changed.

Fig. 3 shows results for the previously designed crosstalk cancellation filters being applied to modified measured impulse responses. Here, the advantage of the robust design becomes evident. While hardly any crosstalk cancellation can be achieved anymore, the robust design (lower plots) does not produce any additional spectral peaks. The non-robust design produces peaks for both the ipsi- and the contralateral paths. Note that crosstalk cancellation performance in such an echoic room cannot be compared to known investigations, where only HRTFs are addressed.

5. CONCLUSIONS

In this paper we have proposed a novel way to design spatially robust crosstalk cancellation filters through an extension of a known least squares design. An additional advantage is the

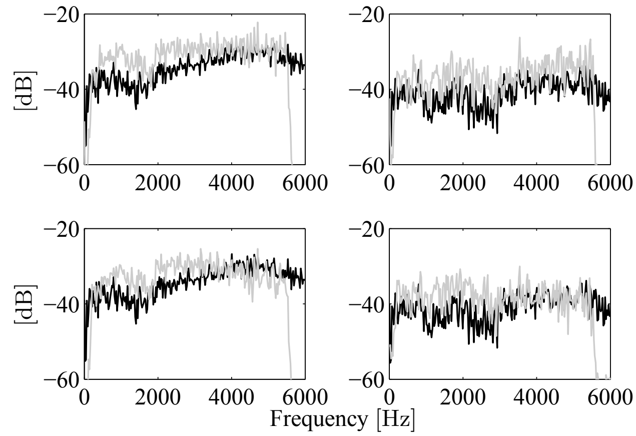

Fig. 3. Same configuration as Fig. 2 except for the torso, which was moved towards the loudspeakers by 2 cm after the crosstalk-canceller design.

possibility to set a parameter for the expected deviation from the "sweet spot". Simulation results on the basis of measured impulse responses from a reverberant office room show the robustness of the extended least squares design.

ACKNOWLEDGEMENT

The authors would like to thank Ms. A. Müller for measuring loads of impulse responses.

6. REFERENCES

[1] P. Damaske, "Head-Related Two-Channel Stereophony with Loudspeaker Reproduction," *Journal of the Acoustical Society of America (JASA)*, vol. 50, 1971.

[2] C. Bourget and T. Aboulnasr, "Inverse Filtering of Room Impulse Response for Binaural Recording Playback Through Loudspeakers," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Adelaide, Australia, Apr. 1994, vol. 3, pp. 301–304.

[3] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduña Bustamante, "Fast Deconvolution of Multichannel Systems using Regularization," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, Mar. 1998.

[4] P. A. Nelson, H. Hamada, and S. J. Elliott, "Adaptive Inverse Filters for Stereophonic Sound Reproduction," *IEEE Trans. on Signal Processing*, vol. 40, no. 7, pp. 1621–1632, Jan. 1992.

[5] D. B. Ward, "Joint Least Squares Optimization for Robust Acoustic Crosstalk Cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 2, pp. 211–215, Feb. 2000.

[6] B. D. Radlović, 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.

[7] M. Kallinger and A. Mertins, "Impulse Response Shortening for Acoustic Listening Room Compensation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.