EFFICIENT CROSSTALK CANCELER DESIGN WITH IMPULSE RESPONSE SHORTENING FILTERS

Terence Betlehem[*], *Paul D. Teal*[†], *Yusuke Hioka*[‡]

[*]Industrial Research Ltd, Lower Hutt, New Zealand

[†]School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

[‡]NTT Cyber Space Laboratories, NTT Corporation, Tokyo, Japan

ABSTRACT

An impulse response shortening approach is used to perform acoustic crosstalk cancellation. Crosstalk canceler filters are traditionally designed using least squares, with an approach that equalizes all room reverberation. However, depending upon end application, some reverberation may be permissible in the delivered signals. This idea is used to create more efficient crosstalk cancellation filters. The filter design is formulated as a minimax problem solvable with linear programming methods. Penalty functions on crosstalk levels and detrimental reverberation are introduced, which allow control of the reverberant tails and crosstalk levels. Shorter crosstalk cancellation filters are designed, by leaving in early echoes and/or allowing a slower decay of the late reverberant tail.

Index Terms— Crosstalk cancellation, impulse response shortening, minimax optimization, reverberation, spatial audio.

1. INTRODUCTION

Crosstalk cancellation finds application in several audio problems. In spatial audio, it is applied in *virtual acoustics* imaging systems, where different signals are delivered to the left and right ear to create the spatial impression of a sound [1]. In *personal audio* applications, crosstalk cancellation could be used to supply separate sounds to separate listeners in the same listening space. In this paper, we propose a method to perform more efficient crosstalk cancellation based upon impulse response shortening filters.

Impulse response shortening is the task of determining a filter which, when convolved with a channel impulse response, shortens its length. It was originally applied in discrete multitone systems [2, 3]. The application to acoustics is motivated by the fact that only the late reverberation tail of the acoustic impulse response is detrimental to speech intelligibility [4, 5, 6]. Impulse response shortening has been used for the de-reverberation of a room impulse response [5]. A minimax formulation was proposed in [6] and solved with a steepest descent algorithm. It was applied to multiple-point equalization in [7]. A p-norm pre-optimization was added to speed the rate of convergence in [8]. Least squares and rayleigh quotient methods of multiple-point equalization were compared in [9].

We apply impulse response shortening to solve the crosstalk cancellation problem, devising a minimax approach that can be solved with standard linear programming methods. A novel weighting function is presented which shapes the reverberant tail with the exponential decay seen in real rooms. We explore the trade-offs between design parameters, the crosstalk and reverberation levels, showing that shorter crosstalk cancellation filters can be obtained than from impulse response inversion.

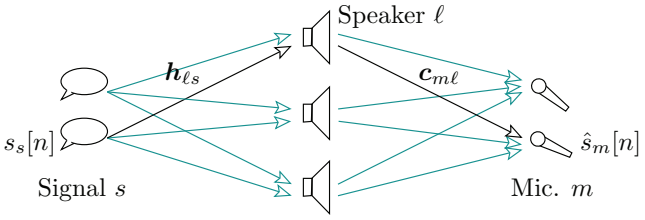


Fig. 1. A 3×2 crosstalk canceler, showing the channel impulse response $\boldsymbol{c}_{m\ell}$ between loudspeaker ℓ and microphone m, and crosstalk cancellation filter $\boldsymbol{h}_{\ell s}$ between source s and loudspeaker ℓ.

2. CROSSTALK CANCELLATION

The objective of crosstalk cancellation is to independently deliver S objective sound signals to M microphones or pressure matching points using L ($\geq S$) loudspeakers. Typically $M = S$. The problem shall be referred to as the $L \times M$ crosstalk problem. As shown in Fig. 1, crosstalk cancellation requires a bank of filters to successfully cancel the crosstalk. Let the crosstalk cancellation filter impulse response from the sth signal to the ℓth loudspeaker be summarized in vector $\boldsymbol{h}_{\ell s}$ and the acoustic impulse response from the ℓth loudspeaker to mth microphone contained in vector $\boldsymbol{c}_{m\ell}$. Assume, without loss of generality, that each impulse response $\boldsymbol{h}_{\ell s}$ is of the same length N_h and each impulse response $\boldsymbol{c}_{m\ell}$ has the same length N_c.

Formulating the problem in the time-domain, the Toeplitz convolution matrix $\boldsymbol{C}_{m\ell}$ can be defined in terms of acoustic impulse response elements $c_{m\ell}[0], \ldots, c_{m\ell}[N_c - 1]$ as

$$\boldsymbol{C}_{m\ell} = \begin{bmatrix} c_{m\ell}[0] & 0 & \ldots & 0 \\ \vdots & c_{m\ell}[0] & \ddots & \vdots \\ c_{m\ell}[N_c-1] & \vdots & \ddots & 0 \\ 0 & c_{m\ell}[N_c-1] & & c_{m\ell}[0] \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & c_{m\ell}[N_c-1] \end{bmatrix}.$$

The overall or nett impulse response r_{ms} from the sth signal to the mth microphone is

$$\boldsymbol{r}_{ms} = \sum_{l=1}^{L} \boldsymbol{C}_{m\ell} \boldsymbol{h}_{\ell s} \tag{1}$$

All nett impulse responses are then length $N_r = N_h + N_c - 1$.

In traditional crosstalk cancellation, the task is to design $\boldsymbol{h}_{\ell s}$ such that the crosstalk impulse responses are zero, i.e.

$$\boldsymbol{r}_{ms} = \mathbf{0}_{N_r}, m \neq s,$$

while the delivered impulse responses are each equal to a time delay vector,

$$\boldsymbol{r}_{ss} = [\mathbf{0}^T_{N_{\mathrm{mp}}}, 1, \mathbf{0}^T_{N_{\mathrm{eq}}}]^T,$$

where N_{mp} is the length of zero padding required to ensure the system is minimum phase, $N_{\mathrm{eq}} = N_r - N_{\mathrm{mp}} - 1$ is the length of equalized tail and $\mathbf{0}_n$ is an n-long zero vector. The nett response from every signal s to every microphone m can be summarized in the matrix equation:

$$\boldsymbol{C}\boldsymbol{H} = \boldsymbol{R}, \tag{2}$$

where $\boldsymbol{C}$ is the $N_r M \times N_h L$ block matrix representing convolution by all of the channel responses defined as $[\boldsymbol{C}]_{m\ell} = \boldsymbol{C}_{m\ell}$, $\boldsymbol{H}$ is the $N_h L \times M$ block matrix of crosstalk cancellation filters defined $[\boldsymbol{H}]_{\ell m} = \boldsymbol{h}_{\ell m}$ and $\boldsymbol{R}$ is the matrix of nett impulses defined as $[\boldsymbol{R}]_{ms} = \boldsymbol{r}_{ms}$, and $[\boldsymbol{M}]_{ij}$ is the (i,j)th sub-block of matrix $\boldsymbol{M}$. The crosstalk canceler based upon an inverse filter design is obtained by solving for the inverse filters directly through

$$\boldsymbol{H} = \boldsymbol{C}^{\dagger}\boldsymbol{R},$$

where $(\cdot)^{\dagger}$ is the pseudoinverse matrix, though a certain level of regularization is typically required to dampen slowly decaying modes.

The next section outlines the approach of crosstalk cancellation with shortened impulse responses.

3. IMPULSE RESPONSE SHORTENING

Impulse response shortening is formulated in Section 3.1 as a minimax problem which can be solved through standard linear programming methods. The design strategy is to "not care" about the structure of the early reverberation in the delivered impulse responses, zero weighting the corresponding equations. The impulse responses are shortened on the basis of psychoacoustic considerations summarized in Section 3.2 using weighting functions defined in Section 3.3.

3.1. Minimax Problem

The problem is solved using the infinity norm as a penalty measure. The solution is facilitated by stacking the columns of $\boldsymbol{H}$ and $\boldsymbol{R}$ into vectors $\boldsymbol{h} = \overrightarrow{\boldsymbol{H}}$ and $\boldsymbol{r} = \overrightarrow{\boldsymbol{R}}$ respectively (where $\overrightarrow{\cdot}$ represents the vectorization operation), causing the matrix equation (2) to become

$$\boldsymbol{\mathcal{C}}\boldsymbol{h} = \boldsymbol{r},$$

where $\boldsymbol{\mathcal{C}} = \boldsymbol{I}_M \otimes \boldsymbol{C}$, $\boldsymbol{I}_M$ is the $M \times M$ identity matrix, $\boldsymbol{r}$ is the vector of desired overall responses and $\otimes$ is the Kronecker product[1]. The task is to match the vector of realized overall impulse responses

[1] For the example of 3×2 crosstalk cancellation, the system of equations is:

$$\begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \boldsymbol{C}_{13} & 0 & 0 & 0 \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \boldsymbol{C}_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \boldsymbol{C}_{13} \\ 0 & 0 & 0 & \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \boldsymbol{C}_{23} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_{11} \\ \boldsymbol{h}_{21} \\ \boldsymbol{h}_{31} \\ \boldsymbol{h}_{12} \\ \boldsymbol{h}_{22} \\ \boldsymbol{h}_{32} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{11} \\ \boldsymbol{r}_{21} \\ \boldsymbol{r}_{12} \\ \boldsymbol{r}_{22} \end{bmatrix}.$$

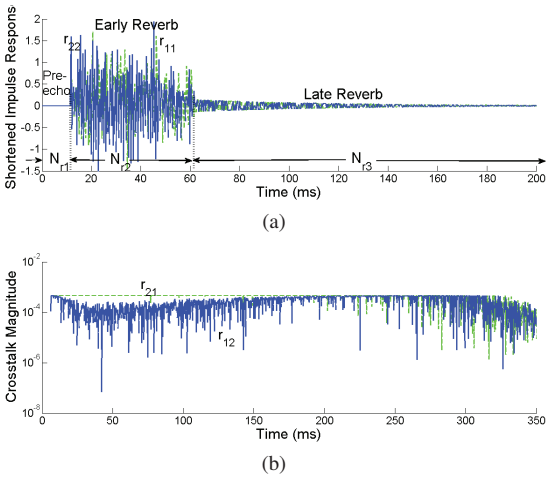


Fig. 2. Crosstalk canceled impulse responses resulting from a 3×2 system showing (a) the delivered impulse responses showing pre-echo, early and late reverberation and (b) the magnitude of crosstalk.

$\hat{\boldsymbol{r}} = \boldsymbol{\mathcal{C}}\boldsymbol{h}$ to a desired vector $\boldsymbol{r}$ for all impulse response taps. The deviations of the taps in $\hat{\boldsymbol{r}}$ from $\boldsymbol{r}$ are penalized according to a weighting vector $\boldsymbol{w}$. The objective problem hence can be written:

$$\min_{\boldsymbol{h}} \left\| \boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{C}}\boldsymbol{h} - \boldsymbol{r}) \right\|_{\infty}, \tag{3}$$

where the weighting matrix $\boldsymbol{\mathcal{W}} = \mathrm{Diag}(\boldsymbol{w})$ corresponds to the diagonal matrix with $[\boldsymbol{\mathcal{W}}]_{ii} = \boldsymbol{w}_i$, $\boldsymbol{r}$ is the vector of desired overall responses and $\| \cdot \|_{\infty}$ is the infinity norm operator. The weighting vector is constructed from a weighting matrix $\boldsymbol{W}$ defined in Section 3.3, as $\boldsymbol{w} = \overrightarrow{\boldsymbol{W}}$.

The problem can be written in epigraph form [10], by introducing bounding variable t as

$$\begin{aligned} \min_{t,\boldsymbol{h}} \quad & t \\ \text{subject to} \quad & [\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{C}}\boldsymbol{h} - \boldsymbol{r})]_n \leq t, \\ & [\boldsymbol{\mathcal{W}}(\boldsymbol{\mathcal{C}}\boldsymbol{h} - \boldsymbol{r})]_n \geq -t, \qquad n = 1 \ldots N, \end{aligned} \tag{4}$$

where $N = M^2 N_r$. The epigraph form hence recasts the infinity norm problem as a linear program. Problem (4) is then solved using SeDuMi [11]. Note the following features of the method:

1. This linear program can be solved numerically either using the simplex algorithm or interior point methods. The former is numerically efficient for medium scale problems, whilst the latter is more efficient for large-scale problems. Interior point methods converge quadratically, and hence more rapidly than the linear convergence rate of the steepest descent approach in [6]. Unlike [6], where the *step-size* update parameter must be prechosen, conventional convex optimization methods choose this parameter automatically.
2. For room impulse responses the problem is large scale, being both numerically intensive and memory hungry. The approach is currently not practical for impulse responses of more than about 1000 taps. The program possesses $2N$ inequality constraints yet in a reverberant room, impulse responses are thousands of taps longs. For a 3×2 system when reverberation time is $T_{60} = 250$ ms, impulse responses sampled at 8 kHz

are $N_c = 2000$ taps long. With shortening filters of length $N_h = N_c$, 24000 linear inequality constraints are required solving.

3. The size of the problem can be reduced by excising the "don't care" rows of $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mathcal{C}}$ and the corresponding elements in $\boldsymbol{r}$. $M(N_{r2} - 1)$ "don't care" taps are present corresponding to the early reverberation, elements $N_{r1} + 2$ through $N_{r1} + N_{r2}$, of each vector $\boldsymbol{r}_{ss}$.

The problem could also be solved directly using a least squares approach as in [9]. Such optimization is more computationally efficient and hence applicable for the larger problems. This optimization can be carried out directly $\boldsymbol{r} = (\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{C}})^{\dagger}\boldsymbol{h}$ and regularization applied to determine filters with a smaller 2-norm energy. Reducing the 2-norm energy improves the robustness of the filter taps to perturbation.

3.2. Psychoacoustic Aspects

The idea underlying the crosstalk cancellation approach is to care less about the amount of reverberation in the delivered impulse responses than the levels of crosstalk. Following [6], each shortened impulse response is divided into three separate regions shown in Fig. 2(a):

Pre-echo should be small enough not to be audible.

Early reverberation is used by the human ear to reinforce the sound volume of the impulse response. The first 50 ms is known to contribute beneficially to speech intelligibility [4].

Late reverberation is said to cause syllabic blurring and reduce the speech intelligibility, especially the later reverberant tail.

Here pre-echo is N_{r1} taps, early reverberation is N_{r2} taps, late reverberation is N_{r3} taps and $N_{r1} + N_{r2} + N_{r3} = N_r$. The task of impulse response shortening as presented by [6, 8] is to choose the shortening filter impulses to reinforce the beneficial early reverberation of delivered impulse responses whilst attenuating the detrimental late reverberation.

The extension of [6, 8] to the crosstalk problem requires an explicit penalty on the pre-echo. Different loudspeakers have different propagation times to each microphone. N_{r1} is set equal to the *maximum* propagation time of all loudspeaker-microphone pairs. The penalty is required to prevent sound from arriving before this time.

3.3. Weighting Functions

In light of the psychoacoustic considerations, weighting functions on the shortened delivered impulse responses and crosstalk are defined. S^2 windows are defined penalizing the difference between the nett impulse responses $\hat{\boldsymbol{r}}$ and the objective impulses $\boldsymbol{r}$. For the S shortened delivered impulse responses $\boldsymbol{r}_{ss}$, the weighting vector is:

$$\boldsymbol{w}_{ss} = [\mathbf{1}_{N_{r1}}^T, \boldsymbol{w}_{\mathrm{d}}^T, \boldsymbol{w}_{\mathrm{u}}^T]^T \tag{5}$$

where $\mathbf{1}_{N_{r1}}$ is a vector penalizing pre-echo, $\boldsymbol{w}_{\mathrm{d}}$ is an N_{r2}-long vector penalizing deviations in the early reverberation from a desired response and $\boldsymbol{w}_{\mathrm{u}}$ is an N_{r3}-long vector penalizing the late reverberant tail.

The design strategy is to not care about the details in the first N_{r2} taps of the delivered impulse responses. However, rewarding at least one tap in this region not to be zero is necessary to avoid the trivial solution $\hat{\boldsymbol{r}} = \mathbf{0}$. We unity-weight the first tap of the early reverberation to be equal to one, whilst not caring about the rest i.e. $\boldsymbol{w}_{\mathrm{d}} = [1, \mathbf{0}_{N_{r2}-1}^T]^T$. The non-zero weight in $\boldsymbol{w}_{\mathrm{d}}$ corresponds to the position $N_{r1} + 1$ of the unity tap in $\boldsymbol{r}_{ss}$. A simple iteration of the optimization can be used to adjust these weights to ensure all crosstalk levels are the same, though the details are omitted here.

This approach contrasts with that of [6] where an additional weight vector rewards the early reverberation. We expect slight degradation over this method caused by imposing conditions on a single tap of each delivered impulse response. However for long impulse responses with many early reverberation taps N_{r2}, the decrease in performance due to lost degrees of freedom is small.

A linearly increasing penalty weight for the undesired part is used in [6, 8] whilst a constant weight is used in [9]. We propose an exponential penalty function for the penalty vector, defined as:

$$[\boldsymbol{w}_{\mathrm{u}}]_n = e^{\beta(n-N_{r3})/N_{r3}},$$

where β controls the degree to which late reverberation is penalized. This choice creates an exponential decay of the late reverberant tail that is seen in real rooms. It penalizes the last tap in the reverberant tail by the same level as the pre-echo.

Define a single constant window to penalize the $S^2 - S$ crosstalk impulse responses equally:

$$\boldsymbol{w}_{ms} = \rho\mathbf{1}_{N_r}, m \neq s \tag{6}$$

where ρ is a penalty factor corresponding to the desired attenuation level of the crosstalk. We choose $\rho = 1$ to set the crosstalk at the same inaudible level as the pre-echo.

The penalty vectors are summarized into the $SN_r \times M$ block matrix $\boldsymbol{W}$, defined as $[\boldsymbol{W}]_{ms} = \boldsymbol{w}_{ms}$.

4. SIMULATION

A 3×2 crosstalk canceler is designed using the proposed approach and tested with different design parameters. The performance metrics quantified are the early-to-late ratio:

$$\mathrm{ELR} = \frac{\sum_{n=N_{r1}+1}^{N_{r1}+N_{50}} |h_{ss}[n]|^2}{\sum_{n=N_{r1}+N_{50}+1}^{N_r} |h_{ss}[n]|^2},$$

the crosstalk level of $\max_{n,m,s\neq m} |h_{ms}[n]|$ and T_{30} and T_{40} reverberation times. N_{50} is the number of taps for a 50 ms duration.

Performance is investigated for a set of synthetic impulse responses. Reverberant impulse responses of $T_{60} = 250$ ms duration are generated with exponentially decaying independent Gaussians and sampled at 4 kHz. These responses represent the case that microphones and loudspeakers are spaced far apart for all frequencies of interest. A small random time delay was introduced at the start of each impulse, modeling the propagation time from each loudspeaker and microphone up to a maximum distance of 4 m.

Varying the reverberant tail penalty slope β permits shaping of the reverberant tail. Choosing $\beta = 0$ forces all of the reverb into the "don't care" region, suppressing late reverberation. Choosing $\beta = 10$ creates a gradual decay of the late reverberant tail. A typical design is shown in Fig. 2 for $\beta = 5.7$, $N_h = 125$ ms and $N_{r1} = 50$ ms. Here the crosstalk level is -67 dB and $T_{40} = 130$ ms.

Fig. 3 shows contours of cross talk achieved for various shortening filter lengths N_h, and lengths of the "don't care" region N_{r2}. The crosstalk cancellation improves with increasing N_h, with increasing N_{r2}, and with increasing β. Introducing a 50 ms "don't care" region, the required length of the shortening filter to yield a -50 dB crosstalk is reduced by 75 ms for $\beta = 0$.

Fig. 4 shows the trade-off between the crosstalk cancellation and early-to-late ratio as the reverberant tail penalty is varied. The

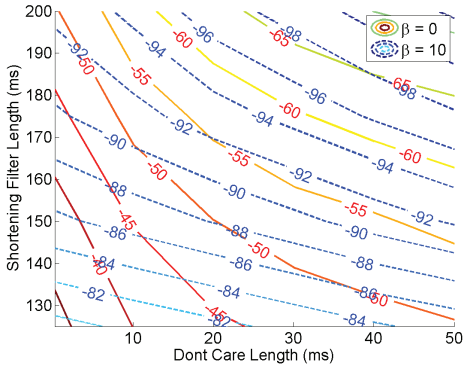
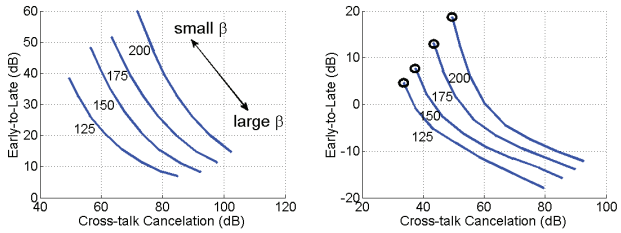


Fig. 3. Contours of crosstalk (in dB) achieved for various shortening filter lengths, and lengths of the "don't care" region. The solid contours are for tail penalty slope $\beta = 0$, whereas the dashed contours are for $\beta = 10$.



(a) (b)

Fig. 4. Plots of early-to-late ratio versus crosstalk cancellation levels as reverberant tail penalty slope β is varied from 0 to 10, for (a) a 50 ms "don't care" region and (b) no "don't care" region. Shortening filters were designed with filter lengths $N_h \in [125, 150, 175, 200]$ ms. Circles in (b) represent ideal minimax inverse filters.

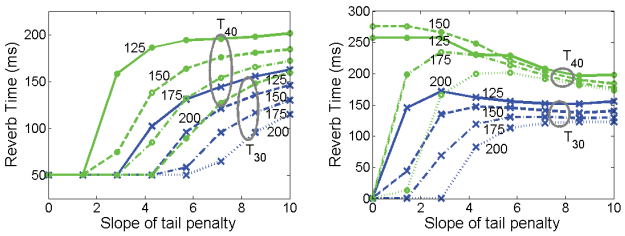
crosstalk cancellation is better when more energy is allowed to leak into the late reverberant tail. Introducing a 50 ms "don't care" region (Fig. 4(a)) improves both the early-to-late ratio and the crosstalk suppression.

The minimax analog to the ideal inverse filter of Section 2 corresponds to no "don't care" region and a flat penalty function and is shown by the circles in Fig. 4(b). Better crosstalk cancellation is achieved by leaking reverberation into the tails of the delivered impulse responses.

Fig. 5 plots the shortening of delivered impulse responses achieved as measured by T_{30} and T_{40} reverberation times against the reverberant tail penalty slope. Fig. 5(a) shows that for $N_{r2} = 50$ ms, reverberation time is increased from 50 ms by increasing β and decreasing N_h. Fig. 5(b) shows that for no "don't care" region, the early reverberant decay can also reliably be controlled by varying β and N_h.

5. CONCLUSION

A crosstalk cancellation approach is presented using impulse response shortening filters. By permitting early reverberation, crosstalk canceling filters are shown to require less taps than ideal impulse inversion. Further, allowing a slow late reverberant decay



(a) (b)

Fig. 5. The shortening achieved, as measured by the T_{30} and T_{40} reverberation times, shown for (a) 50 ms "don't care" region and (b) no "don't care" region. Shown for shortening filters lengths $N_h \in [125, 150, 175, 200]$ ms.

further reduces the filter length requirements. An impulse shaping method is presented, where key parameters are the length of the early reverberation "don't care" region and a reverberant tail slope parameter. The crosstalk levels and the reverberation time of the shortened impulse responses are controlled by these parameters.

6. REFERENCES

[1] B. Atal and M. R. Schroeder, "Apparent sound source translator," U.S. Patent 3 236 949, Feb. 1966.

[2] P. Melsa and R. Younce, "Impulse response shortening for discete multitone transceivers," *IEEE Trans. Communications*, vol. 44, pp. 1662 – 1672, Dec. 1996.

[3] G. Arslan, B. Evans, and S. Kiaei, "Optimum finite-length equalization for multicarrier transceivers," *IEEE Trans. Signal Processing*, vol. 49, pp. 3123 – 3135, Feb. 2001.

[4] R. Thiele, "Richtungsverteilung und zeitfolge der schallrckwrfe in rumen," *Acustica*, vol. 3, pp. 291–302, 1953.

[5] M. Kallinger and A. Martins, "Room impulse response shortening for acoustic listening room compensation," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, 2005, pp. 197–200.

[6] T. Mei, A. Martins, and M. Kallinger, "Room impulse response shortening with infinity-norm optimization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3745–3748.

[7] T. Mei and A. Martins, "On the robustness of room impulse response reshaping," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aug. 2010.

[8] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity- and p -norm optimization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

[9] W. Zhang, E. Habets, and P. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Aug. 2010.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[11] I. Pólik, "Addendum to the SeDuMi user guide," June 2005, Version 1.1.