ROBUST COMBINED CROSSTALK CANCELLATION AND LISTENING-ROOM COMPENSATION

Jan Ole Jungmann, Radoslaw Mazur, Markus Kallinger, and Alfred Mertins*

University of Lübeck, Institute for Signal Processing, 23562 Lübeck, Germany

ABSTRACT

Crosstalk cancellation is a well-known technique to deliver virtual 3D sound to a listener via two or more loudspeakers. In this method, the binaural source signals are processed with a network of prefilters prior to loudspeaker reproduction in order to ensure that only the prescribed source signals reach the corresponding ears of the listener, such that all acoustic crosstalk is cancelled out and no significant reverberation is present. Since listeners might slightly move their heads within a certain range while listening, robust designs are needed. In this paper, we propose a method for the robust design of crosstalk cancellers in which we replace known least-squares techniques by a p-norm optimization that allows us to explicitly control the amount of crosstalk and *shape* the remaining reverberation effects according to a desired decay.

Index Terms— Crosstalk, deconvolution, equalizers, robustness, spatial filters

1. INTRODUCTION

Three-dimensional audio reproduction with loudspeakers in a room requires the use of a prefilter network that processes the source signals prior to rendering in such a way that the individual signals arrive only at the designated ears of the listener. Thus, all acoustic crosstalk needs to be cancelled out, and no spectral distortion or reverberation should be introduced along the signal paths. Early approaches assumed symmetric propagation paths and aimed at the equalization of head related transfer functions (HRTFs) and the cancellation of crosstalk [1]. Later designs considered the individual transmission paths and tried to tackle the above mentioned equalization problem in more detail as described below.

Signal propagation from N loudspeakers to the two ears of a listener can be described by a $2 \times N$ matrix $\boldsymbol{C}(z)$ that is composed of system functions $C_{m\ell}(z)$ which describe the transmission from loudspeaker ℓ to ear m. Given two sources, the preprocessing network is then given by an $N \times 2$ matrix $\boldsymbol{H}(z)$ of system functions $H_{\ell q}(z)$ that determine the transmission from source q to loudspeaker ℓ. The goal is to obtain an overall system $\boldsymbol{G}(z) = \boldsymbol{C}(z)\boldsymbol{H}(z)$ in which the off-diagonal terms $G_{12}(z)$ and $G_{21}(z)$, which describe the crosstalk, both vanish, and the diagonal terms $G_{11}(z)$ and $G_{22}(z)$ do not introduce audible distortion.

Assuming just two loudspeakers, an ideal prefilter network can be written as $\boldsymbol{H}_{ideal}(z) = \frac{z^{-n_0}}{\det \boldsymbol{C}(z)} \text{adj}\,\{\boldsymbol{C}(z)\}$, where adj $\{\boldsymbol{C}(z)\}$ is the adjugate matrix of $\boldsymbol{C}(z)$ and ensures perfect crosstalk cancellation. The term $\frac{z^{-n_0}}{\det \boldsymbol{C}(z)}$ denotes the system function of a prefilter that has to be applied to both channels in order to remove spectral distortion with a delay n_0 which allows the filter to be causal. In particular, the task of equalizing the determinant is found to be very demanding, because it contains the difference of two products of system functions and has many zeros on or close to the unit circle of the z-plane [2, 3, 4]. Nelson et al. [5] proposed a least-squares design that aimed to achieve both equalization and crosstalk cancellation in one step. This method has been extended by Ward [6], who simultaneously considered multiple head positions in order to achieve good spatial robustness. Kallinger and Mertins [7] tried to achieve spatial robustness with a least-squares method that considered perturbations from the measured systems based on statistical knowledge of the acoustic transfer functions [8]. In this paper, we extend the impulse-response shaping method from [9] to the design of robust crosstalk cancellers and keep control of the amount of crosstalk due to small head movements and the audibility of spectral distortion and reverberation.

The paper is organized as follows: In Section 2, a brief overview of the design of crosstalk cancellers is given, and the proposed approach is derived. In Section 3, we present the results of applying the proposed method to measured room impulse responses, and in Section 4, we give some conclusions.

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 $\text{diag}[\cdot]$ turns a vector into a diagonal matrix, and $\|\cdot\|_p$ returns the ℓ_p-norm of a vector. The lengths of FIR filters are denoted as L_c and L_h for filters $c(n)$ and $h(n)$, respectively. Given a vector $\boldsymbol{c}$ containing an impulse response $c(n)$, the operator $\text{convmtx}(\boldsymbol{c}, L_h)$ generates a convolution matrix of size $(L_h + L_c - 1) \times L_h$.

2. CROSSTALK-CANCELLER DESIGN

In the following we describe the crosstalk canceller for two loudspeakers as depicted in Figure 1, but the extension to more loudspeakers is straight forward.

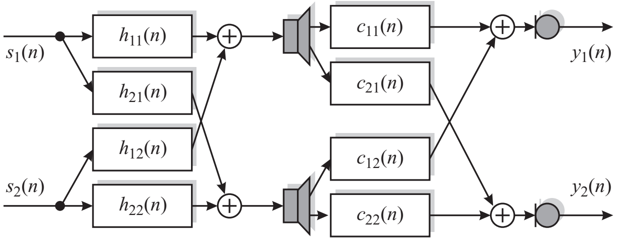

Figure 1: 2×2 setup of a crosstalk canceller.

This work has been supported by the German Research Foundation under Grant No. ME1170/3-1.

*Markus Kallinger is now with the Fraunhofer IIS, Germany, and contributed to the work while being with the Signal Processing Group, University of Oldenburg, Germany.

Assuming FIR filters with system functions $C_{m\ell}(z)$ and $H_{\ell q}(z)$, respectively, the elements of the 2×2 global system can be written as

$$G_{mq}(z) = C_{m1}(z)H_{1q}(z) + C_{m2}(z)H_{2q}(z), \quad m, q = 1, 2. \tag{1}$$

Because the filters for the two sources can be designed independent of each other, we only describe the transmission of source one in the following. The traditional aim is to have no transmission through the crosstalk system

$$G_{21}(z) = C_{21}(z)H_{11}(z) + C_{22}(z)H_{21}(z), \tag{2}$$

and to achieve ideal transmission through the system

$$G_{11}(z) = C_{11}(z)H_{11}(z) + C_{12}(z)H_{21}(z). \tag{3}$$

Perfect crosstalk cancellation (i.e., $G_{21}(z) = 0$) is achieved if $H_{11}(z) = C_{22}(z)F(z)$ and $H_{21}(z) = -C_{21}(z)F(z)$, where $F(z)$ is an arbitrary system function. Ideal transmission means that the system $G_{11}(z)$ satisfies $G_{11}(z) = [C_{11}(z)C_{22}(z) - C_{12}(z)C_{21}(z)]F(z) \approx D_1(z)$, where $D_1(z)$ is a bandpass system that does not cause audible distortion and has an appropriate delay (for example, it has to take the delays by the systems $C_{m\ell}(z)$ into account). The main task then is to find an appropriate filter $F(z)$ and possibly define an appropriate system $D_1(z)$.

By assuming FIR systems, representing the impulse responses $h_{\ell q}(n)$ and $d_1(n)$ by vectors $\boldsymbol{h}_{\ell q}$ and $\boldsymbol{d}_1$, respectively, and forming convolution matrices

$$\boldsymbol{C}_{m\ell} = \operatorname{convmtx}\big([c_{m\ell}(0), \ldots, c_{m\ell}(L_c - 1)]^T, L_h\big),$$

we can write the general problem as

$$\boldsymbol{C}\boldsymbol{h} = \boldsymbol{d} \tag{4}$$

where

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_{11} \\ \boldsymbol{h}_{21} \end{bmatrix}, \quad \boldsymbol{d} = \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{0} \end{bmatrix}. \tag{5}$$

The linear system of equations (4) can, for example, be solved in a least-squares sense. In order to increase the spatial robustness, Ward [6] extended the system of equations by introducing multiple microphone positions. Using the notation from [7], in which the deviations from the ideal responses are described by stochastic perturbation systems $P_{m\ell}(z)$, this can be expressed as

$$\text{minimize}_{\boldsymbol{h}} : \quad E\left\{d([\boldsymbol{C} + \boldsymbol{P}]\boldsymbol{h}, \boldsymbol{d})\right\} \tag{6}$$

where $d(\cdot, \cdot)$ is an appropriate distance measure, $E\{\cdot\}$ denotes the expected value, and $\boldsymbol{P}$ contains the stochastic convolution matrices $\boldsymbol{P}_{m\ell}$. If $d(\cdot, \cdot)$ is chosen as the Euclidean distance, depending on the way the perturbation is dealt with, one either obtains the solution from [6] or from [7].

We here follow the approach from [9] with the multichannel extension from [10] and use a p-norm based design criterion that takes the average temporal masking threshold of the human auditory system into account by using appropriate weighting windows. For the algorithm development we assume we have R perturbed versions of the channel matrix, $\boldsymbol{C}^{(r)} = \boldsymbol{C} + \boldsymbol{P}^{(r)}$, $r = 1, 2, \ldots, R$, at our disposal, where $\boldsymbol{P}^{(r)}$ means the r-th perturbation. The corresponding global responses computed via $\boldsymbol{C}^{(r)}\boldsymbol{h}$ are denoted by $\boldsymbol{g}_{11}^{(r)}$ and $\boldsymbol{g}_{21}^{(r)}$. In accordance with [9, 10], we define the desired parts of the global responses as

$$g_{11,d}^{(r)}(n) = w_d(n)g_{11}^{(r)}(n). \tag{7}$$

Undesired parts are given by

$$g_{11,u}^{(r)}(n) = w_u(n)g_{11}^{(r)}(n) \tag{8}$$

and by the weighted crosstalk

$$g_{21,u}^{(r)}(n) = w_c(n)g_{21}^{(r)}(n). \tag{9}$$

The window $w_d(n)$ extracts and weights the first T_d milliseconds after the direct pulse of $g_{11}^{(r)}(n)$, and the window $w_u(n)$ is given by the reciprocal of the desired energy decay curve, which was linked in [9] to the average temporal masking curve. The approximate masking limit starts at -10 dB at 4 ms after the direct sound impulse and then decays exponentially to -70 dB at 200 ms after the direct pulse on the logarithmic scale.

Finally, we define the window for weighting the crosstalk as

$$w_c(n) = \max[w_0, w_u(n)]. \tag{10}$$

The value of w_0 directly determines the attenuation of the crosstalk component in comparison to the desired path. The max operator ensures that the tail of the crosstalk path does not exceed the reverberation tail of the signal path $g_{11}(n)$.

Analogue to [9], the proposed optimization problem reads

$$\text{minimize}_{\boldsymbol{h}} : \quad f(\boldsymbol{h}) = \log\left(\frac{f_u(\boldsymbol{h})}{f_d(\boldsymbol{h})}\right) \tag{11}$$

with

$$f_d(\boldsymbol{h}) = \|\boldsymbol{g}_d\|_{p_d} \tag{12}$$

and

$$f_u(\boldsymbol{h}) = \|\boldsymbol{g}_u\|_{p_u} \tag{13}$$

where $\boldsymbol{g}_d = [\boldsymbol{g}_{11,d}^{(1)\ T}, \ldots, \boldsymbol{g}_{11,d}^{(R)\ T}]^T$ and $\boldsymbol{g}_u = [\boldsymbol{g}_{11,u}^{(1)\ T}, \boldsymbol{g}_{21,u}^{(1)\ T}, \ldots, \boldsymbol{g}_{11,u}^{(R)\ T}, \boldsymbol{g}_{21,u}^{(R)\ T}]^T$. The log operation in (11) is used in view of obtaining a simple expression for the gradient. The optimization of (11) is done by applying a gradient descent procedure; the learning rule reads

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^l - \mu(l)\,\nabla_{\boldsymbol{h}} f(\boldsymbol{h}^l) \tag{14}$$

with $\mu(l)$ being an adaptive positive step-size parameter. The optimization can start with a relatively large value for $\mu(l)$, and by observing the value of the objective function during optimization, it can be reduced whenever the objective function starts to increase instead of decrease. The gradients $\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h})$ and $\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h})$ are derived in the next equations.

The gradient for $f_d(\boldsymbol{h})$ is calculated as

$$\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h}) = \zeta_d(\boldsymbol{h}) \cdot \nabla_{\boldsymbol{h}} \phi_{f_d}(\boldsymbol{h}) \tag{15}$$

where

$$\zeta_d(\boldsymbol{h}) = \left(\sum_{r=1}^{R} \sum_{n=0}^{L_g-1} |g_{11,d}^{(r)}(n)|^{p_d}\right)^{\frac{1}{p_d}-1} \tag{16}$$

and

$$\nabla_{\boldsymbol{h}} \phi_{f_d}(\boldsymbol{h}) = \begin{bmatrix} \sum_{r=1}^{R} \left(\boldsymbol{C}_{11}^{(r)}\right)^T \boldsymbol{b}_{11,d}^{(r)} \\ \sum_{r=1}^{R} \left(\boldsymbol{C}_{12}^{(r)}\right)^T \boldsymbol{b}_{11,d}^{(r)} \end{bmatrix} \tag{17}$$

and $\boldsymbol{b}_{11,d}^{(r)}$ given by

$$\boldsymbol{b}_{11,d}^{(r)} = \operatorname{diag}[\operatorname{sgn}[\boldsymbol{g}_{11,d}^{(r)}]]\operatorname{diag}[\boldsymbol{w}_d]|\boldsymbol{g}_{11,d}^{(r)}|^{(p_d-1)}. \tag{18}$$

The gradient for the undesired part $f_u(\boldsymbol{h})$ is calculated as

$$\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h}) = \zeta_u(\boldsymbol{h}) \cdot \nabla_{\boldsymbol{h}} \phi_{f_u}(\boldsymbol{h}) \tag{19}$$

where

$$\zeta_u(\boldsymbol{h}) = \left(\sum_{r=1}^{R}\sum_{n=0}^{L_g-1} |g_{11,u}^{(r)}(n)|^{p_u} + |g_{21,u}^{(r)}(n)|^{p_u}\right)^{\frac{1}{p_u}-1} \tag{20}$$

and

$$\nabla_{\boldsymbol{h}}\phi_{f_u}(\boldsymbol{h}) = \begin{bmatrix} \sum_{r=1}^{R}\left(\boldsymbol{C}_{11}^{(r)}\right)^T \boldsymbol{b}_{11,u}^{(r)} + \left(\boldsymbol{C}_{21}^{(r)}\right)^T \boldsymbol{b}_{21,u}^{(r)} \\ \sum_{r=1}^{R}\left(\boldsymbol{C}_{12}^{(r)}\right)^T \boldsymbol{b}_{11,u}^{(r)} + \left(\boldsymbol{C}_{22}^{(r)}\right)^T \boldsymbol{b}_{21,u}^{(r)} \end{bmatrix}. \tag{21}$$

The vectors $\boldsymbol{b}_{11,u}^{(r)}$ and $\boldsymbol{b}_{21,u}^{(r)}$ are given by

$$\boldsymbol{b}_{11,u}^{(r)} = \mathrm{diag}[\mathrm{sgn}[\boldsymbol{g}_{11,u}^{(r)}]]\mathrm{diag}[\boldsymbol{w}_u]|\boldsymbol{g}_{11,u}^{(r)}|^{(p_u-1)} \tag{22}$$

and

$$\boldsymbol{b}_{21,u}^{(r)} = \mathrm{diag}[\mathrm{sgn}[\boldsymbol{g}_{21,u}^{(r)}]]\mathrm{diag}[\boldsymbol{w}_c]|\boldsymbol{g}_{21,u}^{(r)}|^{(p_u-1)}. \tag{23}$$

Finally, the gradient of $f(\boldsymbol{h})$ is given by

$$\nabla_{\boldsymbol{h}} f(\boldsymbol{h}) = \frac{1}{f_u(\boldsymbol{h})}\nabla_{\boldsymbol{h}} f_u(\boldsymbol{h}) - \frac{1}{f_d(\boldsymbol{h})}\nabla_{\boldsymbol{h}} f_d(\boldsymbol{h}). \tag{24}$$

Due to the special structure of the convolution matrices, the calculations can be performed in the frequency domain with the fast Fourier transform and the inverse fast Fourier transform; so the algorithm (14) is computationally efficient.

3. SIMULATION RESULTS

For the simulations we measured impulse responses in an office room of size $6.85\ \mathrm{m} \times 5.3\ \mathrm{m} \times 3\ \mathrm{m}$. The reverberation time was estimated as $T_{60} = 0.7\ \mathrm{s}$. The room impulse responses were measured using an exponential sine-sweep method at a sampling rate of $48\ \mathrm{kHz}$ and were then downsampled to $16\ \mathrm{kHz}$. For the measurements we used two Klein&Hummel M52 loudspeakers as sound sources. Both loudspeakers had a distance of $1.1\ \mathrm{m}$ to the back wall and $0.9\ \mathrm{m}$ to each other. We used a custom-made dummy head with two Beyerdynamics MM1 microphones inside the ears to record the signals. The dummy head was placed $2\ \mathrm{m}$ in front of the loudspeakers, facing toward them.

To get the different realizations of the acoustic channels, the dummy head was mounted on a linear stage with a positioning accuracy in the sub-millimeter magnitude. We moved the position of the head inside a $2\ \mathrm{cm} \times 2\ \mathrm{cm} \times 2\ \mathrm{cm}$ volume with a spatial sampling of $1\ \mathrm{cm}$ on every axis; so we finally ended up with 2×27 measured impulse responses per ear.

The lengths of the room impulse responses were limited to $L_c = 4000$ taps; exemplary, one of the measured RIRs is shown in Figure 2.

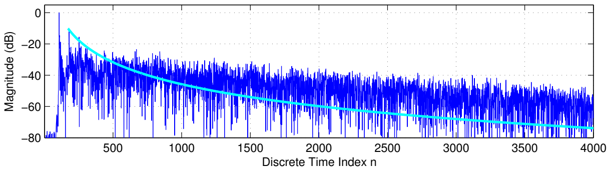

Figure 2: One of the measured RIRs on a logarithmic scale. The light blue line is the average masking limit.

For the reshaping procedure, we set the weighting windows for the direct sound component as proposed in [9] ($T_d = 4$ ms); the weighting window for the unwanted part of the crosstalk component was chosen as defined by (10) with $w_0 = 180$, $p_d = 20$ and $p_u = 10$ for all experiments. With increasing w_0, the crosstalk attenuation will increase, but with very large w_0, spectral coloration will occur. With the value of 180, spectral distortion is still negligible. To visualize the effect of the reshaping approach, we overlayed the RIR from Figure 2 with its reshaped realization and depicted the result in Figure 3.

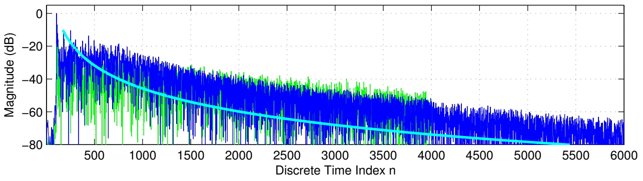

Figure 3: One of the measured RIRs (green) overlayed with its reshaped version (blue). The light blue line is the average masking limit.

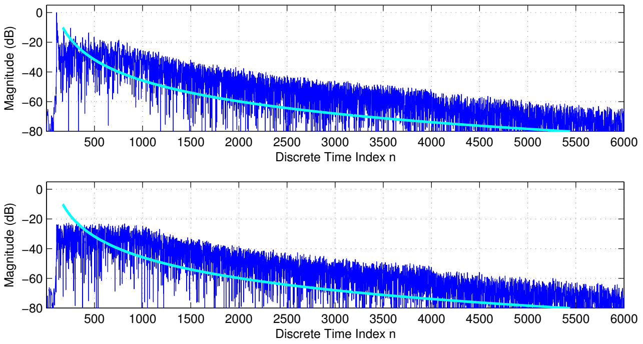

Figure 4: Reshaped impulse response of the desired component of 10th realization of the channel (upper plot). The lower plot shows the reshaped crosstalk path for the 10th realization of the channel. In both plots the light blue line is the temporal masking curve, defined by the direct pulse of the desired path.

To investigate the reshaping of the signal path for the crosstalk component, we take the 10th realization of the channel and depict the reshaped impulse response for the desired component and the crosstalk component in Figure 4. The impulse responses have been normalized for the direct pulse of the desired signal path to be 1. The energy of the direct pulse is $22.6\ \mathrm{dB}$ higher for the reshaped desired path than for the reshaped crosstalk path; in addition, the reverberation tail of the crosstalk component does not exceed the reverberation tail of the desired component.

To demonstrate the spatial robustness of the proposed approach, we depict the average impulse response of the desired component in comparison to the average impulse response of the crosstalk component in Figure 5; both impulse responses were normalized for the direct pulse of the average desired path to be 1. The observations that were made for a specific realization of the acoustic channels are also valid for the average impulse responses over all realizations: the energy of the direct pulse is $20.7\ \mathrm{dB}$ higher for the average reshaped desired path than for the average reshaped crosstalk path while the reverberation tail of the crosstalk component does not exceed the reverberation tail of the desired component. Before reshaping, the magnitude of the direct pulse was just $5.2\ \mathrm{dB}$ higher for the average desired path than for the average crosstalk path.

We compared the results of the proposed approach with the method from [6]. We choose the desired system to be a bandpass filtered unit pulse; for the filtering we used a 10th order Butterworth

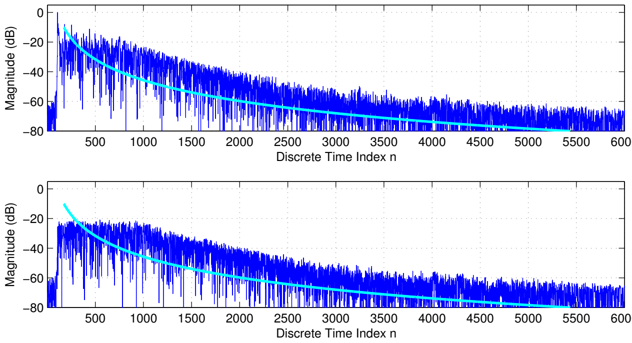

Figure 5: Average reshaped impulse response of the desired component (upper plot). The lower plot shows the average reshaped crosstalk path. In both plots the light blue line is the temporal masking curve, defined by the direct pulse of the desired path.

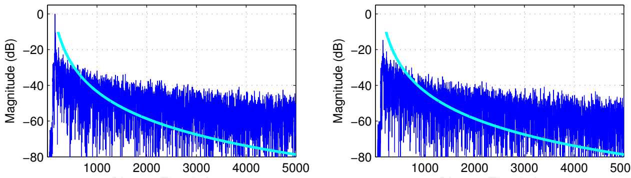

Figure 6: Average impulse response for the desired (left) and the crosstalk component by using the robust least-squares method from [6] for the design of the prefilters. The light blue line is the average masking curve, defined by the direct pulse of the desired path.

filter with the cutoff frequencies at 200 Hz and 5500 Hz. In Figure 6 we depict the average shaped impulse response for the desired and the crosstalk component. In comparison to our approach (the upper plot of Figure 5) it can be seen that the reverberation tail does not follow the decay of the masking limit. Besides that, the direct pulse of the average desired path exceeds the direct pulse of the average crosstalk path by only 14.6 dB.

In addition, we compared the performance of the proposed method and the least-squared method by designing filters of different lengths. We then measured the minimum amount of crosstalk attenuation, defined as the minimum ratio of the largest tap of the signal paths $g_{11}^{(r)}(n)$ and the largest tap of the corresponding crosstalk $g_{21}^{(r)}(n)$, for all 27 realizations of the channels. The results are plotted in Figure 7. As one can see, the guaranteed crosstalk attenuation is much larger for the proposed method.

4. CONCLUSIONS

In this paper we proposed a novel way to combine the spatially robust listening room compensation by optimizing a p-norm based criteria with the problem of crosstalk cancellation. Simulations were performed using measured impulse responses from a reverberant office room. With a joint design for multiple head positions, we could show that the proposed method significantly reduces the main peak of the crosstalk component while keeping the reverberant part of the crosstalk at the same level as the reverberation of the desired component. Besides that, the amount of perceivable reverberation could be lowered.

There are still parameters that need to be tuned in order to find an optimal equilibrium between crosstalk cancellation and listening room compensation; in addition, experiments with more than two loudspeakers need to be done. Besides the qualitative evaluation performed in this paper, alternative measures are needed to capture the performance of combined listening-room compensation and crosstalk-compensation algorithms.

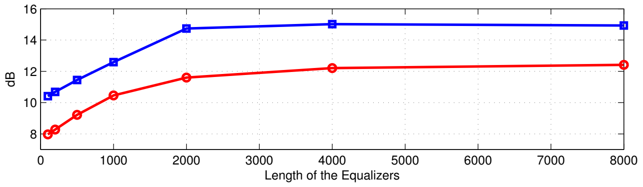

Figure 7: Minimum crosstalk attenuation over all 27 channel realizations for different prefilter lengths, measured as the minimum ratio of the direct pulse of the desired component and the crosstalk component path, plotted in dB. The upper curve (blue) represents the results of the proposed approach, while the lower curve (red) represents the least-squares approach from [6].

5. REFERENCES

[1] P. Damaske, "Head-Related Two-Channel Stereophony with Loudspeaker Reproduction," *Journal of the Acoustical Society of America (JASA)*, vol. 50, pp. 1109–1115, 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)*, vol. 3, Adelaide, Australia, Apr. 1994, 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] Y. Kahana, P. A. Nelson, and S. Yoon, "Experiments on the Synthesis of Virtual Acoustic Sources in Automotive Interiors," in *AES 16th International Conference: Spatial Sound Reproduction*, 1999.

[5] 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.

[6] 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.

[7] M. Kallinger and A. Mertins, "A spatially robust least squares crosstalk canceller," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. 1, 15-20 April 2007, pp. 177–180.

[8] 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.

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

[10] T. Mei and A. Mertins, "On the robustness of room impulse response reshaping," in *Proc. International Workshop on Acoustic Echo and Noise control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.