OPTIMIZED GRADIENT CALCULATION FOR ROOM IMPULSE RESPONSE RESHAPING ALGORITHM BASED ON P-NORM OPTIMIZATION

Radoslaw Mazur, Jan Ole Jungmann, and Alfred Mertins

Institute for Signal Processing
University of Lübeck
23562 Lübeck, Germany

ABSTRACT

By using room impulse response shortening and reshaping it is possible to reduce the reverberation effects and therefore improve the perceived quality. This may be achieved by a prefilter that modifies the overall impulse response to have a faster decay. The traditional filter shortening approach using least-squares methods is fast and directly computable, but it suffers from late echoes. Newer approaches using the p-norm overcome this drawback but are computationally very demanding, as the optimization process uses a gradient-descent approach with slow convergence. In this work we propose a modification to this approach that results in a significantly faster convergence. With this modification, the algorithm is less likely to be trapped in a local minimum and therefore also leads to a better convergence point. The method will be demonstrated on simulated and real-world room impulse responses.

Index Terms— Room impulse response (RIR), shortening, p-norm, gradient method, optimization

1. INTRODUCTION

In order to reduce the influence of the reverberation in acoustic scenes and to improve the perceived quality the concepts of room impulse response (RIR) shortening and reshaping have been introduced [1, 2, 3]. Using a filter to either preprocess the loudspeaker signal or postprocess the recorded microphone signal, the overall impulse response is equalized [4].

It is not necessary to invert the channel and recover the exact source signal [5, 6, 7]. As proposed in [2, 3] it is sufficient to shape the overall RIR with respect to the human auditory system. Therefore, for reducing the reverberation, it is sufficient to equalize the RIR in a such way that the audible echoes are removed, while the inaudible ones may stay unaffected. As a further benefit, this approach lightens the pressure of designing the prefilter.

The method in [3] exploits the fact that echoes may remain in the signal and are unperceivable if they fall below the temporal masking curve of the human auditory system. As shown in [8] the exact temporal masking curve is signal dependent, but an average signal-independent masking curve has been found in [9] that is triggered by the direct sound and was used in [3] to design the prefilter.

This single-channel approach from [3] is not designed for spatial robustness, and small movements of speaker or microphones may result in substantially changed RIRs and a reduced overall performance. Based on the spatial sampling principle [10] a multichannel extension has been proposed in [11]. It allows for an equalization of multiple points in space by using several loudspeakers and microphones. When the sampling points are chosen dense enough according to the spatial sampling principle an entire area can be equalized. This approach is more robust, as it allows for small movements of the listener inside this area. As these methods are computationally very demanding, a CUDA implementation using modern graphic hardware was presented in [12].

The methods for prefilter design in [2, 3] are gradient-descent approaches, which minimize either an Euclidean norm, an ℓ_p-norm, the ∞-norm or a combination of them. These norms are necessary, as the least-squares methods based on the ℓ_2-norm do not allow for a precise control of the very small coefficients in the tail of the reshaped RIR. Furthermore, the errors are distributed non-uniformly across the time coefficients, which results in late audible echoes. As shown in [3], using the high non-linearity of the ∞-norm the error distribution becomes uniform which allows for a much more precise error control and great reduction of audible echoes. Although computationally little more demanding, the ℓ_p-norm, with $10 \leq p \leq 20$, inherits these benefits and allows for a much faster convergence, as every iteration of the gradient descent takes multiple time coefficients into account.

In this work we propose to speed up calculations for the ℓ_p-norm by a modification of the gradient, as the simple approach is not optimal due to the windowing using the inverse of the average temporal masking curve and the non-quadratic properties of the ℓ_p-norm. Using this modification, the gradient-descent method needs significantly less iterations and allows for a better solution, as it will usually not be trapped in local minima so often. For the sake of conciseness we present the modification for the single-channel case, as the multi position and multichannel extension is quite straightforward.

In the next section we briefly summarize the approach based on the ℓ_p-norm optimization of the time-domain representation of the global RIR. In Section 3 we present the proposed modification of the gradient calculations, which allows for the speed up. In Section 4 we show some results on simulated and real-world data and, finally, in Section 5 we give some short conclusions.

2. ROOM IMPULSE RESPONSE RESHAPING

Let $c(n)$ denote the RIR with the length L_c. With $h(n)$ being a prefilter of length L_h, the overall system $g(n)$ is given by

$$g(n) = h(n) * c(n) = \mathbf{Ch}, \tag{1}$$

with $\mathbf{C}$ being a $L_g \times L_h$ convolution matrix of $c(n)$. The length of $g(n)$ is given by $L_g = L_h + L_c - 1$. The task of filter reshaping is to design $h(n)$ in such a way, that $g(n)$ contains no audible echoes.

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

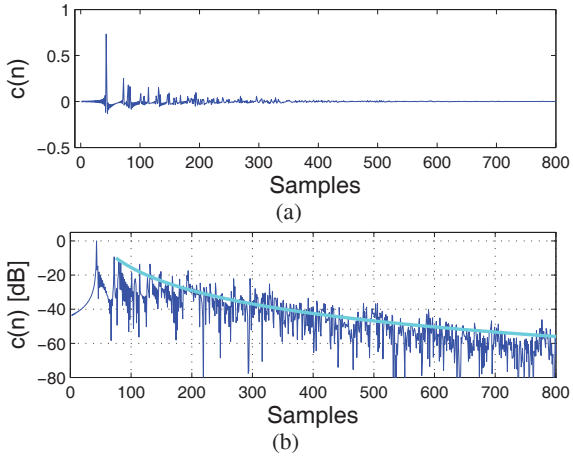

Fig. 1. (a) Simulated RIR $c(n)$ with 800 taps. (b) Energy of the coefficients of $c(n)$. The light blue line indicates the average temporal masking curve.

Here, we follow [2, 3] and define a desired and unwanted part of the global RIR using two windows $w_d(n)$ and $w_u(n)$ as $g_d(n) = g(n) \cdot w_d(n)$ and $g_u(n) = g(n) \cdot w_u(n)$ respectively. As proposed in [3], we use

$$\mathbf{w}_d = [\underbrace{0, 0, \ldots, 0}_{N_1}, \underbrace{1, 1, \ldots, 1}_{N_2} \underbrace{0, 0, \ldots, 0}_{N_3}]^T \tag{2}$$

and

$$\mathbf{w}_u = [\underbrace{0, 0, \ldots, 0}_{N_1+N_2}, \underbrace{\mathbf{w}_0^T}_{N_3}]^T \tag{3}$$

with $N_1 = t_0 \cdot f_s$, $N_2 = 0.004\text{s} \cdot f_s$ and $N_3 = L_g - N_1 - N_2$, where f_s is the sampling frequency and t_0 the time taken by the direct sound. The window $\mathbf{w}_0$ is defined using the inverse of the average temporal masking curve as

$$w_0(n) = 10^{\frac{3}{\log(N_0/(N_1+N_2))} \log(n/(N_1+N_2))+0.5} \tag{4}$$

with $N_0 = (0.2\text{s}+t_0) \cdot f_s$ and time index n ranging from N_1+N_2+1 to $L_g - 1$. In Fig. 1(a) a simulated RIR and in Fig. 1(b) the energy of its coefficients is shown. Additionally, in Fig. 1(b) the inverse of $\mathbf{w}_0$, the average temporal masking curve, is shown. All coefficients above the average temporal masking curve are perceived as echoes.

The goal of RIR reshaping is to maximize some function of $|g_d(n)|$ while minimizing some other function of $|g_u(n)|$. For example, this could be a maximization of the energy of $g_d(n)$ while the energy of $g_u(n)$ would be minimized. A least-squares criterion (i. e. ℓ_2-norm) leads to the minimization problem

$$\begin{cases} \text{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \mathbf{g}_u^T \mathbf{g}_u \\ \text{S.T.} : \mathbf{g}_d^T \mathbf{g}_d = \text{constant} \end{cases} \tag{5}$$

and can be solved by using a generalized eigenvalue decomposition [3]. As this approach suffers from late echoes, due to poor control of the tail of $g(n)$ with some very small values, in [3] a criterion based on the ℓ_p-norm has been proposed. The corresponding optimization problem is given by

$$\text{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \tag{6}$$

with

$$f_u(\mathbf{h}) = ||\mathbf{g}_u||_{p_u} = \left(\sum_{n=0}^{L_g-1} |g_u(n)|^{p_u}\right)^{\frac{1}{p_u}} \tag{7}$$

and

$$f_d(\mathbf{h}) = ||\mathbf{g}_d||_{p_d} = \left(\sum_{n=0}^{L_g-1} |g_d(n)|^{p_d}\right)^{\frac{1}{p_d}} \tag{8}$$

with p_u and p_d being integers. The learning rule for the gradient-descent reads

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu(l) \cdot \nabla_{\mathbf{h}} f(\mathbf{h}^l) \tag{9}$$

with

$$\nabla_{\mathbf{h}} f(\mathbf{h}) = \frac{1}{f_u(\mathbf{h})^{p_u}} \mathbf{C}^T \mathbf{b}_u - \frac{1}{f_d(\mathbf{h})^{p_d}} \mathbf{C}^T \mathbf{b}_d \tag{10}$$

where

$$b_u(n) = w_u(n) \cdot \text{sgn}(g_u(n)) \cdot |g_u(n)|^{p_u-1}, \tag{11}$$

$$b_d(n) = w_d(n) \cdot \text{sgn}(g_d(n)) \cdot |g_d(n)|^{p_d-1}, \tag{12}$$

and $\mu(l)$ being an adaptive step-size parameter.

3. MODIFIED GRADIENT APPROACH

In order to justify the new approach, we first take a closer look on some details of the windowing process $g_u(n) = g(n) \cdot w_u(n)$, the window is given in (4) and (3), and the calculation of the gradient, given in (9) and (10).

The window $w_0(n)$ in (4) is the inverse of the temporal masking curve. The ratio of the smallest and biggest values is in the order of several magnitudes. The windowing process $g_u(n) = g(n) \cdot w_u(n)$ amplifies the small values in the tail of the RIR, so that the overshoot of the RIR over the average temporal masking curve becomes linearized. On a linear scale, the highest peaks contribute to the highest perceived reverberation. This allows for a simple formulation of the objective function in (6), whose minimization also minimizes these highest coefficients in the global RIR and therefore reduces the perceived reverberation. In (9) a gradient-descent approach with the Euclidean gradient is used for the minimization process.

The windowing process can be also interpreted as a highly nonlinear scaling of the individual dimensions of $\mathbf{h}$. The gradient in (10) is therefore not optimal. An example is shown in Fig. 2. In Fig. 2(a), a reshaping filter for the already mentioned example from Fig. 1, is given. The energy decay, as shown in Fig. 2(b), has also an attenuation of several magnitudes. The gradient $\nabla_{\mathbf{h}} f(\mathbf{h}^l)$, as calculated using (10), is shown in 2(c) and has coefficients of the same magnitude in all parts. When making a step of the gradient descent in (9), only very small step-size μ is allowed, in order to achieve convergence in all parts of the RIR.

We propose therefore, a modified gradient update rule

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu(l) \cdot \mathbf{w}_r \odot \nabla_{\mathbf{h}} f(\mathbf{h}^l) \tag{13}$$

with $\odot$ being a point-wise multiplication of two vectors and $w_r(n)$ a window that approximately inverts the scaling property of the initial windowing process. For example, $w_r(n)$ can be obtained by taking the reciprocal of $w_0(n)$.

With this modification the step-size can be larger, and this leads to faster convergence, and the gradient-descent procedure is less likely to be trapped in a local minimum.

The modification in (13) may be interpreted as a multiplication with a diagonal approximation of the Hessian of $f(\mathbf{h})$. Unfortunately, the limited space of this work does not allow for a detailed derivation of this approach.

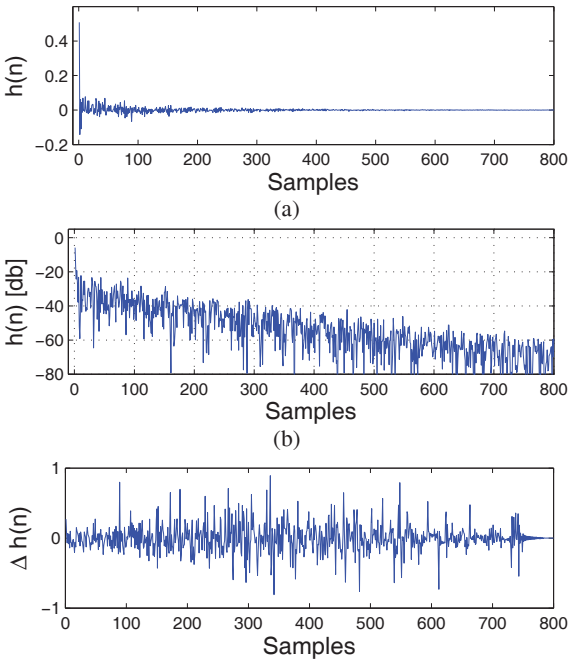

Fig. 2. (a) Reshaping filter $h(n)$ with the same length as $c(n)$ after 5000 iterations. (b) Energy of the coefficients of $h(n)$. (c) The gradient $\Delta_h(n)$ using the normal gradient.

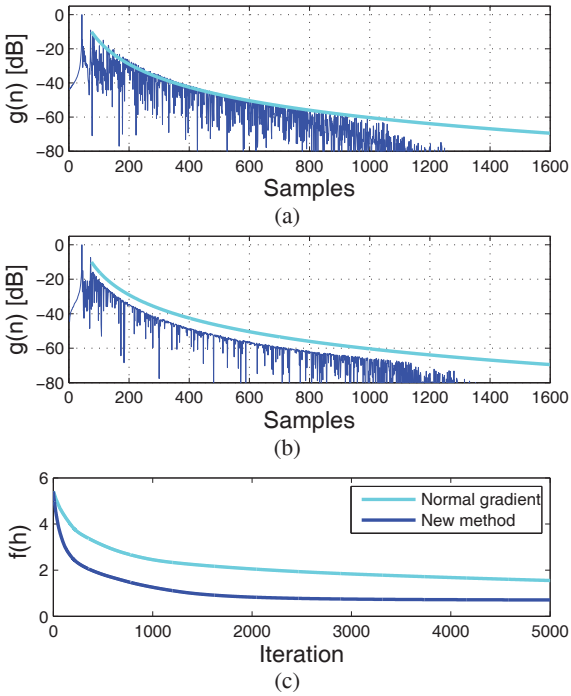

Fig. 3. (a) The overall impulse response $g(n) = c(n) * h(n)$ using the normal gradient with $L_c = L_h = 800$. (b) $g(n)$ using the modified gradient. (c) The value of the objective function $f(\mathbf{h})$ for both approaches.

4. SIMULATIONS

Both gradient methods have been compared on simulated and real RIRs. The parameters have been set as in [13], e. g. $p_d = 10$ and $p_u = 20$. Besides the convergence speed and the objective value, the results have also been examined in terms of reducing the reverberation using the *normalized perceivable reverberation quantization* measure nPRQ, which is derived from the pRQ measure from [13]. This measure captures the average magnitude of the impulse response taps that overshoot the temporal masking limit on a logarithmic scale and that is above -60 dB compared to the direct sound. It is calculated as

$$\mathrm{nPRQ} = \frac{1}{||\mathbf{g}_E||_0} \cdot \sum_{n=N_0}^{L_g-1} g_e(n) \tag{14}$$

when $||\mathbf{g}_E||_0 > 0$ and 0 otherwise, with

$$g_e(n) = 20 \log_{10}(|g(n)| \cdot w_u(n)) \tag{15}$$

for $|g(n)| > \frac{1}{w_u(n)}$, $n \geq N_1 + N_2$ and 0 otherwise with $||\mathbf{g}_E||_0$ denoting ℓ_0 pseudo norm, which counts the number of nonzero elements of a vector. With perfect reshaped RIR $\mathrm{nPRQ} = 0$. Otherwise it measures the average overshoot in dB.

The results of the RIR from Fig. 1 are compared in Fig. 3. In Fig. 3(a) the overall RIR $g(n)$ after 5000 iterations using the normal gradient is shown. Here, the RIR could be almost perfectly reshaped, with only a few coefficients above the temporal masking limit. The nPRQ is reduced from 3.71 to 0.75 as shown in Table 1.

In Fig. 3(b) the results of the modified gradient are presented. Here, we achieve a perfect reshaping with $\mathrm{nPRQ} = 0$, and the coefficients are well below the temporal masking limit, which results in a quite robust design for a large class of signals, for which the reverberation becomes inaudible. In 3(c) the development of the objective function is given. The new method converges significantly faster and converges to a smaller value, which results in the already mentioned much better overall RIR.

Table 1. Comparison of the perceived reverberation using nPRQ-measure in [dB] for different configurations. (S) denotes simulated RIR, and (M) a measured RIR in a real room.

Problem Size (Taps)	800 (S)	2000 (S)	4000 (M)
Unmodified RIR	3.71	4.35	10.62
Standard method	0.75	2.71	5.75
New method	0.00	0.08	0.00

In Fig. 4 the results of a simulated RIR with $L_c = 2000$ are given. With the longer RIR, we can see the faster convergence and a better result for the modified gradient again. The nPRQ measure is reduced from 4.35 to 2.71 and 0.08, respectively. Only the modified gradient gives an almost perfect reshaping.

In Fig. 5 the results for a real RIR, measured in a seminar room with $L_c = 4000$, are given. The room has a high reverberation with $\mathrm{nPRQ} = 10.62$. The standard method is able to reduce the nPRQ by 5, but the overall RIR still contains quite audible reverberations. The modified gradient is able to remove all echoes with $\mathrm{nPRQ} = 0$. The standard method gets trapped in a local minimum.

5. CONCLUSIONS

In this work we have proposed a modification to the gradient based calculation for room impulse response reshaping algorithm based on p-norm optimization. This modification allows for a faster convergence. Additionally, the new algorithm is able to achieve better results, as it does not get trapped in local minima so often. The validity of the approach has been shown on simulated and measured room impulse responses.

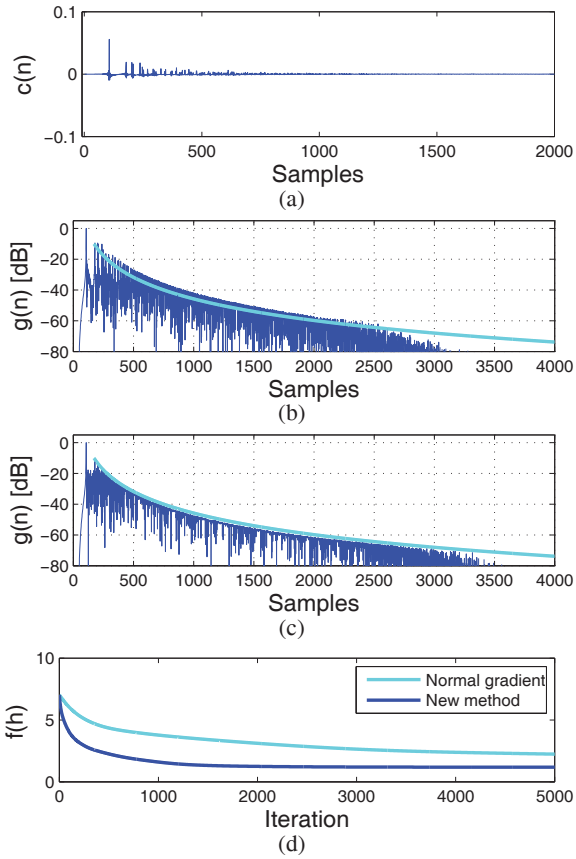

Fig. 4. (a) A simulated RIR with the length 2000. (b) The overall impulse response $g(n) = c(n) * h(n)$ using the normal gradient with $L_h = 2000$. (c) $g(n)$ using the modified gradient. (d) The values of the objective function $f(\mathbf{h})$ for both approaches.

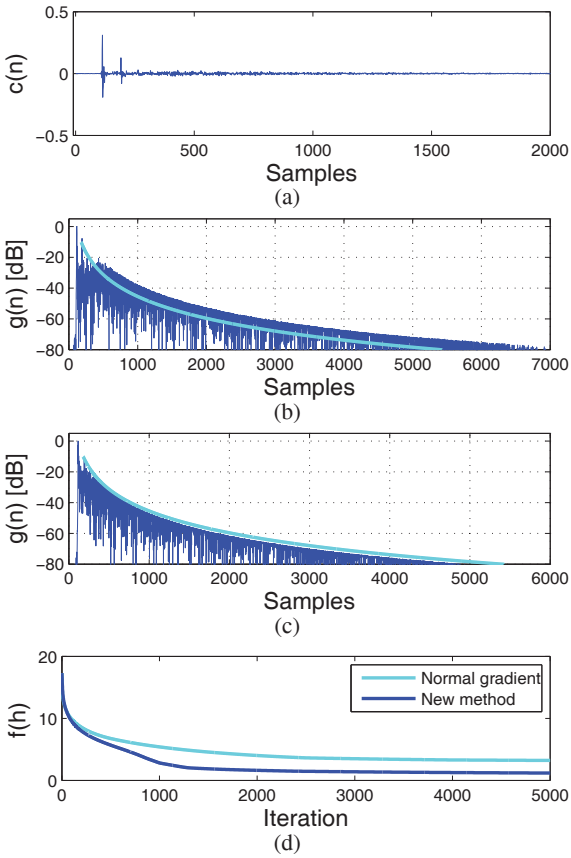

Fig. 5. (a) A measured RIR with the length 4000. (b) The overall impulse response $g(n) = c(n) * h(n)$ using the normal gradient with $L_h = 5000$. (c) $g(n)$ using the modified gradient. (d) The values of the objective function $f(\mathbf{h})$ for both approaches.

6. REFERENCES

[1] Markus Kallinger and Alfred Mertins, "Impulse response shortening for acoustic listening room compensation," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005, pp. 197–200.

[2] M. Kallinger and A. Mertins, "Room impulse response shortening by channel shortening concepts," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA , USA, Oct. 30 - Nov. 2 2005, pp. 898–902.

[3] Alfred Mertins, Tiemin Mei, and Markus 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.

[4] Wancheng Zhang, Emanuel A. P Habets, and Patrick A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2010.

[5] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoustical Society of America*, vol. 68, pp. 165–169, July 1979.

[6] B. D. Radlovic and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[7] Ahfir Maamar, Izzet Kale, Artur Krukowski, and Berkani Daoud, "Partial equalization of non-minimum-phase impulse responses," *EURASIP Journal on Applied Signal Processing*, pp. 1–8, 2006.

[8] Jens Blauert, John Mourjopoulos, and Jorg Buchholz, "Room masking: Understanding and modelling the masking of reflections in rooms," in *Audio Engineering Society Convention 110*, 5 2001.

[9] Louis D. Fielder, "Practical limits for room equalization," in *Audio Engineering Society Convention 111*, 11 2001, pp. 1–19.

[10] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *Signal Processing, IEEE Transactions on*, vol. 54, no. 10, pp. 3790 –3804, oct. 2006.

[11] J. O. Jungmann, R. Mazur, M. Kallinger, and A. Mertins, "Robust combined crosstalk cancellation and listening-room compensation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2011)*, Mohonk, New Paltz, USA, Oct. 2011.

[12] R. Mazur, J. O. Jungmann, and A. Mertins, "On cuda implementation of a multichannel room impulse response reshaping algorithm based on p-norm optimization," in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, Oct. 2011.

[13] J. O. Jungmann, T. Mei, S. Goetze, and A. Mertins, "Room impulse response reshaping by joint optimization of multiple p-norm based criteria," in *Proc. EUSIPCO 2011*, Barcelona, Spain, Aug. 2011, pp. 1658–1662.