Robust Listening Room Compensation by Optimizing Multiple p-Norm Based Criteria

Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins
Universität zu Lübeck, Institute for Signal Processing, D-23538 Lübeck
Email: {jungmann, mazur, mertins}@isip.uni-luebeck.de

Abstract

The purpose of room impulse response reshaping is to reduce reverberation and thus to improve the perceived quality of the received signal by prefiltering the source signal before it is played with a loudspeaker. The optimization of an infinity- and/or p-norm based objective function in the time domain has shown to be quite effective compared to least-squares methods. Multi-position approaches have been developed in order to increase the robustness against small movements of the listener. It has been shown recently, that it is necessary to also consider the frequency-domain representation of the reshaped impulse response, in order to avoid spectral distortions of the overall system. In this contribution we propose to jointly optimize a p-norm based criterion in the time and frequency domains for multiple positions in order to achieve a good reshaping while not affecting the perceived quality due to spectral distortions in a limited listening area.

Multi-Point Reshaping by p-Norm Optimization

By designing reshaping filters, one aims at improving the acoustic properties of a channel. Early approaches tried to invert the acoustic channel. Recent methods exploit psychoacoustic properties of the human auditory system to lighten the pressure on the equalized overall system: echoes may occur if they are below the average temporal masking limit. Furthermore, the optimization of a p-norm based optimality criterion has proven to result in a *superior* overall reshaping compared to least-squares approaches.

One problem is the lack of spatial robustness as RIRs are very sensible, even to small movements. If the loudspeaker or the listener move slightly, the perceived quality is degraded. An efficient implementation of a *multi-point* equalization scheme utilizing multiple loudspeakers is given in [2]: according to the spatial sampling theorem of RIRs, it is sufficient to design the reshaping filters for a finite number of sampled RIRs in a small volume to perform reshaping inside the whole area [2].

Considering a setup consisting of N loudspeakers and R RIRs sampled in a finite area, one aims to design one prefilter for each loudspeakers to jointly reshape all of the measured RIRs in the listening area. Let $c_\ell^{(r)}(n)$ denote the r-th sampled RIR from loudspeaker ℓ to the listening area, whereas the prefilter for the ℓ-th loudspeaker is denoted as $h_\ell(n)$. The global impulse response (GIR) at the r-th sampling point is given by

$$g^{(r)}(n) = \sum_{\ell=1}^{N} c_\ell^{(r)}(n) * h_\ell(n), \tag{1}$$

with $*$ being the convolution operator. Two window functions $w_d(n)$ and $w_u(n)$ are used to characterize the desired and the unwanted part of each GIR. The weighting window for the desired part extracts the first four milliseconds after the direct sound pulse, whereas the weighting window for the unwanted part is linked to the average temporal masking limit of the human auditory system [2].

With $g_u^{(r)}(n) = g^{(r)}(n) \cdot w_u(n)$ and $g_d^{(r)}(n) = g^{(r)}(n) \cdot w_d(n)$, the vectors $\mathbf{g}_u$ and $\mathbf{g}_d$ are constructed by stacking up the R weighted GIRs. The vector $\mathbf{h}$ is constructed by stacking up the N equalizers:

$$\mathbf{h} = \left[\mathbf{h}_1^T, \ldots, \mathbf{h}_N^T\right]^T. \tag{2}$$

The p-norm based optimization problem for designing the multi-point reshaping filters reads

$$\min_{\mathbf{h}} : f(\mathbf{h}) = \left(\frac{\|\mathbf{g}_u\|_{p_u}}{\|\mathbf{g}_d\|_{p_d}}\right) \tag{3}$$

and is solved by applying a gradient-descent procedure [2] with learning rule

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu^l \cdot \nabla_{\mathbf{h}} f\left(\mathbf{h}^l\right), \tag{4}$$

where μ^l is the adaptive step-size parameter in the l-th iteration.

Extended Objective Function

The optimization of (3) considers only the time-domain representations of the GIRs, which, in some cases, can lead to spectral distortions [1]. In this section we extend the multi-point equalization algorithm from [2] with the frequency-domain based regularization term from [1], which is adapted to the multi-point scenario.

The proposed regularization term reads $y(\mathbf{h}) = \|\mathbf{a}_f\|_{p_f}$, where $\mathbf{a}_f$ is a vector that is made up by the concatenation of the discrete Fourier transforms (DFTs) of all the equalizers or by the DFTs of all the GIRs, respectively. Depending on this choice, we demand either the equalizers (*Case 1*) or the GIRs (*Case 2*) to not contain any high spectral peaks.

The new optimization problem is given by

$$\min_{\mathbf{h}} : q(\mathbf{h}) = f(\mathbf{h}) + \alpha \cdot y(\mathbf{h}) \quad \text{s.t.} \quad \mathbf{h}^T\mathbf{h} = 1 \tag{5}$$

with regularization weight α.

Derivation of the Gradient

To derive the required gradient, we begin by formulating the regularization term in matrix-vector notation. Let $\mathbf{F}$ be the DFT matrix of compatible size and

$$\widetilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{F} \end{bmatrix}. \tag{6}$$

Furthermore, we define $\mathbf{T}$ to be either the identity matrix (*Case 1*) or $\widetilde{\mathbf{C}}$ (*Case 2*), with

$$\widetilde{\mathbf{C}} = \left[\widehat{\mathbf{C}}^{(1),T}, \ldots, \widehat{\mathbf{C}}^{(R),T}\right]^T \tag{7}$$

being made up by the matrices $\widehat{\mathbf{C}}^{(r)} = \left[\mathbf{C}_1^{(r)}, \ldots, \mathbf{C}_N^{(r)}\right]$, that contain the individual convolution matrices for each RIR.

These definitions allow us to rewrite the regularization term as

$$y(\mathbf{h}) = \|\mathbf{a}_f\|_{p_f} = \left\|\widetilde{\mathbf{F}}\mathbf{T}\mathbf{h}\right\|_{p_f}. \tag{8}$$

The gradient for the regularization term is calculated as

$$\nabla_{\mathbf{h}} y(\mathbf{h}) = \zeta(\mathbf{h}) \cdot \Re\left\{\left(\widetilde{\mathbf{F}}\mathbf{T}\right)^H \mathbf{b}_\mathbf{f}\right\} \tag{9}$$

with $\zeta(\mathbf{h})$ being

$$\zeta(\mathbf{h}) = \left(\sum_{k=0}^{L_a-1} |a_f(k)|^{p_f}\right)^{\frac{1}{p_f}-1}, \tag{10}$$

$\mathbf{b}_f$ given by

$$\mathbf{b}_f = \mathrm{diag}\{\mathrm{sign}\{\mathbf{a}_f\}\} \cdot |\mathbf{a}_f|^{p_f-1}, \tag{11}$$

and $\Re\{\cdot\}$ returning the real part of its input vector.

Finally, the gradient for the objective function reads

$$\nabla_{\mathbf{h}} q(\mathbf{h}) = \nabla_{\mathbf{h}} f(\mathbf{h}) + \alpha \cdot \nabla_{\mathbf{h}} y(\mathbf{h}), \tag{12}$$

where the derivation of $\nabla_{\mathbf{h}} f(\mathbf{h})$ is given in [2].

Results

To test the algorithm, we measured multiple RIRs using four loudspeakers according to the spatial sampling theorem. We measured 27 RIRs of length $L_c = 4000$ taps for each loudspeaker inside a [2 cm × 2 cm × 2 cm] volume on which we designed the equalizers of length $L_h = 5000$ taps. We then measured 40 more RIRs inside the volume to test the equalizers in the case of small spatial mismatch. To quantify the dereverberation performance, we utilize the nPRQ measure, which captures the average overshot of the time coefficients over the temporal masking limit that is above −60 dB on a logarithmic scale. The nPRQ measure is a normalized version of the pRQ measure, introduced in [1]. The frequency-domain representation is captured by the *spectral-flatness* (SF) measure [1]. The values for the SF and the nPRQ were averaged over all 40 test GIRs. The results are given in a comprehensive form in Table 1. For all experiments, we chose $p_d = 20$, $p_u = 10$ and $p_f = 8$. Besides that, we depict the frequency response of a reshaped impulse response in the case of small spatial mismatch with and without regularization in Figure 1.

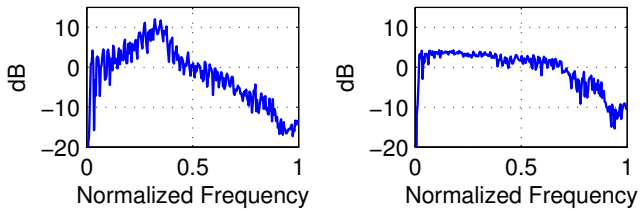

Figure 1: Magnitude frequency response of a GIR in the case of spatial mismatch. Left plot: without regularization. Right plot: with regularization (GIR constrained, $\alpha = 5$).

Table 1: Average values for the nPRQ and SF measures before and after reshaping using the different algorithms. EQ means, that we constrained the DFTs of the equalizers (*Case 1*), while GIR means that we constrained the DFTs of the global impulse responses (*Case 2*).

Setup	nPRQ [dB]	SF
unreshaped	9.93	0.63
$\alpha = 0$	0.18	0.3
EQ, $\alpha = 1$	0.92	0.54
EQ, $\alpha = 2$	3.22	0.56
GIR, $\alpha = 5$	0.78	0.62
GIR, $\alpha = 10$	1.58	0.77

Conclusions

In this work we extended the time-domain based objective function from [2] by a p-norm based regularization term in the frequency-domain. The frequency-domain based regularization term was introduced in [1] for the single-channel scenario. Simulations showed that our approach attenuates spectral distortions at the expense of a slightly degraded dereverberation performance. As in the single channel case [1], constraining the DFTs of the overall impulse responses yields better results than constraining just the DFTs of the equalizers.

References

[1] 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*, pages 1658–1662, Barcelona, Spain, Aug. 2011.

[2] 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)*, pages 305–308, New Paltz, New York, USA, Oct. 2011.