JOINT TIME-DOMAIN RESHAPING AND FREQUENCY-DOMAIN EQUALIZATION OF ROOM IMPULSE RESPONSES

Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins

Institute for Signal Processing
University of Lübeck
Ratzeburger Allee 160, 23562 Lübeck, Germany

ABSTRACT

In listening room compensation, the aim is to compensate for the degradations that are rendered to an audio signal by transmission in a closed room. Due to multiple reflections of the soundwaves, the listener receives a superposition of delayed and attenuated versions of the source signal. A filter is designed so that the convolution of the room impulse response and the equalizer contains better acoustic properties than the original acoustic channel. Common approaches for dereverberation optimize only the time-domain representation of the overall impulse response and may introduce distortions in the frequency domain. Equalization of the frequency response, on the other hand, often does not consider the time-domain behavior in an ideal way. In this paper, we propose a novel method to jointly consider both the time- and frequency-domain behavior. It outperforms the methods known from literature in terms of dereverberation and equalization performance. Results are presented for a room impulse response measured in a real living room.

***Index Terms*—** room impulse response, listening room compensation, optimization, spectral flatness

1. INTRODUCTION

In listening room compensation (LRC) the goal is to compensate for an acoustic channel, described by the room impulse response (RIR), in order to make the received signal hardly distinguishable from the original source signal by a human listener [1]. Classically, the LRC filter is designed in such a way that the difference between the global impulse response (GIR, the convolution of the room impulse response with the LRC filter) and a given target system is minimized in a least-squares sense [2]. The target system is usually chosen as a bandpass-filtered and/or delayed unit pulse.

More relaxed approaches originate from the field of channel shortening. In channel shortening, the effective length of an impulse response is reduced by concentrating most of the energy of the GIR in a certain time interval. For example, by concentrating the energy in the first 50 ms after the first peak, the psychoacoustic D-50 measure [3], used to quantify speech intelligibility, is maximized.

In [4] it was shown that a *shaping* is preferable over a *shortening* in practice. The method for reshaping an impulse response has been further developed in [5]. The least-squares optimality criterion was generalized to a p-norm based measure. It could be shown that the p-norm based optimality criterion allows for a much better control of the reshaping than least-squares based methods. By preferring solutions with one dominant peak, the GIR obtained by p-norm optimization according to [5] usually shows a reasonably flat frequency response. An efficient implementation for CUDA-enabled hardware was developed in [6]. However, for rooms with large reverberation time, it was shown in [7] that it is preferable to also consider the frequency-domain representation of the GIR during optimization, in order to guarantee a flat frequency response.

In this work, we present a new method to jointly reshape and equalize a RIR. It achieves a time-domain reshaping that uniformly follows any pre-defined decay curve (e.g., the average temporal masking curve of human listeners according to [8]) by optimizing a p-norm based optimality criterion. In addition, a flat overall frequency response is obtained by combining the time-domain criterion with the spectral flatness measure [9, 10], here applied to the squared magnitude frequency response of the overall system. Experiments show that the proposed method yields superior results compared to other methods such as the one proposed in [7].

This work is organized as follows. In Section 2 we give a short overview of the p-norm based method to design reshaping filters. In Section 3 we briefly review the spectral flatness measure and integrate it into the optimization problem. Reshaping results for a measured room impulse response are given in Section 4. Finally, some conclusions are drawn in Section 5.

Notation: Vectors and matrices are denoted by lowercase and uppercase boldface letters, respectively. The asterisk $*$ denotes convolution, and $\|\cdot\|_p$ returns the p-norm of a vector. The superscript ^{H} denotes the Hermitian transpose of a matrix, the superscript * denotes the complex conjugate of a

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

complex number. The sign of a complex number is defined as its projection onto the unit circle.

2. ROOM IMPULSE RESPONSE RESHAPING

In this section we give a brief overview of the p-norm based reshaping method from [5]. Let $c(n)$ denote the room impulse response of length L_c. With the reshaping filter $h(n)$ being of length L_h, the global impulse response is given by $g(n) = h(n) * c(n)$ with length $L_g = L_c + L_h - 1$. Usually, two window functions $w_d(n)$ and $w_u(n)$ are used to define a *desired* and an *unwanted* part of the GIR. In channel shortening approaches, the ratio of the energies of the desired and the unwanted parts is maximized [11, 4].

2.1. p-Norm Based Reshaping

In [5] it was proposed to generalize the quadratic optimality criterion usually considered for dereverberation filter design (e.g. [4]) to a p-norm based criterion. The optimization problem is given by

$$\min_{\mathbf{h}} \;:\; f(\mathbf{h})\,, \quad f(\mathbf{h}) = \log\left(\frac{\|\mathbf{g}_u\|_{p_u}}{\|\mathbf{g}_d\|_{p_d}}\right), \tag{1}$$

where $\mathbf{g}_u$ is the vector made up by the *unwanted* part of the global impulse response $g_u(n) = w_u(n)\,g(n)$, and the *desired* part $\mathbf{g}_d$ is defined accordingly. The vector $\mathbf{h} = [h_1, h_2, \ldots, h_{L_h}]^T$ contains the equalizer impulse response $h(n)$, i.e., $h_n = h(n-1)$. By choosing appropriately high values for p_d and p_u (typically $10 \leq p_d, p_u \leq 20$) and proper windows, one achieves a very even shaping of the time-domain coefficients of the unwanted part of the global impulse response that, for example, closely follows a given computational model for the average temporal masking curve of human listeners [5]. The optimization of (1) is carried out by applying a gradient-descent procedure.

2.2. Frequency-Domain Based Regularization

In [4] it has been shown that one needs to consider the spectral distortions that can be introduced by designing the equalizer just for the time-domain representation of the RIR. In [4] the spectral distortions were corrected by an additional postfilter that was designed for the GIR. It has been proposed to consider the frequency-domain representation of the global impulse response during optimization in [7]. For this, a p-norm based optimality criterion in the frequency domain was presented. The regularization term from [7] is given by

$$y(\mathbf{h}) = \|\mathbf{g}_f\|_{p_f}\,, \tag{2}$$

where $\mathbf{g}_f$ is the vector containing the discrete Fourier transform of the global impulse response. The regularization term with $p_f \approx 8$ forces the overall system to show no high spectral peaks.

3. THE PROPOSED REGULARIZATION

In this work, we replace the regularization term (2) from [7] by the *spectral flatness measure* (SFM) known from literature [9, 10]. The gradient of the new extended objective function is explicitly derived to allow for an efficient filter design.

3.1. Spectral Flatness Measure

The spectral flatness of an impulse response $g(n)$ can be measured as the ratio of the geometric and arithmetic means of squared samples of its frequency response $G\big(e^{j\omega}\big)$:

$$\mathrm{SFM}_g = \frac{\left(\prod_{k=1}^{K} \left|G\big(e^{j\omega_k}\big)\right|^2\right)^{\frac{1}{K}}}{\frac{1}{K}\sum_{k=1}^{K}\left|G(e^{j\omega_k})\right|^2}, \tag{3}$$

where ω_k, $k = 1, 2, \ldots, K$ are the discrete frequencies under consideration. Typically, the frequency response is computed by a length-K discrete Fourier transform (DFT) of $g(n)$. In the case of a perfectly flat frequency response, no spectral distortions occur, and the SFM is equal to one. With increasing spectral distortions, the SFM degrades down to zero.

3.2. Proposed Method

To design the reshaping filter, we formulate an extended objective function (as in [7]) that jointly considers the time- and the frequency-domain representations of the global impulse response. The latter is introduced through the SFM, which penalizes both spectral peaks and notches of the overall magnitude frequency response. This is in contrast to the p-norm based criterion from [7] that mainly measured peaks in the magnitude frequency response. In order to take into account properties of the measurement and/or reproduction system (e.g. the lowpass filter used in the D/A converter and the inability of acoustic setups to reproduce a DC signal), we allow the SFM to be computed solely for a predefined frequency range during the optimization process.

The proposed optimization problem is given by

$$\min_{\mathbf{h}} \quad : \quad f(\mathbf{h}) + \alpha s(\mathbf{h})\,, \tag{4}$$

where $f(\mathbf{h})$ is given in equation (1), α is a positive weighting factor for the regularization term

$$s(\mathbf{h}) = -\log\left(\frac{\left(\prod_{k=1}^{K} \left|G\big(e^{j\omega_k}\big)\right|^2\right)^{\frac{1}{K}}}{\frac{1}{K}\sum_{k=1}^{K}\left|G(e^{j\omega_k})\right|^2}\right) \tag{5}$$

with K being the number of discrete frequencies in the range between ω_1 and ω_K. The log operation in (5) allows for an easier expression of the gradient $\nabla_{\mathbf{h}} s(\mathbf{h})$.

In the following, we will use vector notation for the derivation of the required gradient. For this, a vector $\mathbf{g}$ is made

up by the K considered entries of the frequency response of the global system with impulse response $g(n)$ in the form $\mathbf{g} = [g_1, g_2, \ldots, g_k]^T$ with $g_k = G(e^{j\omega_k})$. With $\mathbf{C}$ denoting the convolution matrix of $c(n)$ and $\widetilde{\mathbf{F}}$ being a modified DFT matrix that just contains the rows capturing the discrete frequencies ω_k, $k = 1, 2, \ldots, K$, the vector $\mathbf{g}$ can be expressed as

$$\mathbf{g} = \mathbf{M}\mathbf{h} \quad \text{with} \quad \mathbf{M} = \widetilde{\mathbf{F}}\mathbf{C}. \tag{6}$$

By exploiting the logarithmic laws, (5) can be rewritten as

$$s(\mathbf{h}) = A(\mathbf{h}) - B(\mathbf{h}), \tag{7}$$

with

$$A(\mathbf{h}) = \log\left(\frac{1}{K}\sum_{k=1}^{K}|g_k|^2\right) \tag{8}$$

capturing the arithmetic mean part of the calculation of the SFM and

$$B(\mathbf{h}) = \frac{2}{K}\sum_{k=1}^{K}\log(|g_k|) \tag{9}$$

capturing the part which calculates the geometric mean of the SFM.

Considering (7), the gradient $\nabla_{\mathbf{h}} s(\mathbf{h})$ is given by

$$\nabla_{\mathbf{h}} s(\mathbf{h}) = \nabla_{\mathbf{h}} A(\mathbf{h}) - \nabla_{\mathbf{h}} B(\mathbf{h}). \tag{10}$$

The derivation of $\nabla_{\mathbf{h}} A(\mathbf{h})$ and $\nabla_{\mathbf{h}} B(\mathbf{h})$ is given in the following.

3.2.1. Gradient for the Arithmetic Mean

By applying the chain rule, the partial derivative of $A(\mathbf{h})$ with respect to a coefficient h_n is given by

$$\frac{\partial A(\mathbf{h})}{\partial h_n} = \zeta_a \sum_{k=1}^{K}\frac{2}{K}|g_k|\,\mathrm{sign}\{g_k\}\,m_{kn}, \tag{11}$$

where m_{kn} denotes the entry in the k-th row and n-th column of $\mathbf{M}$ and

$$\zeta_a = \frac{1}{\frac{1}{K}\sum_{k=1}^{K}|g_k|^2}. \tag{12}$$

By simplifying (11), the gradient for the part capturing the arithmetic mean is given by

$$\nabla_{\mathbf{h}} A(\mathbf{h}) = \frac{2}{\sum_{k=1}^{K}|g_k|^2}\mathbf{M}^H\mathbf{g}. \tag{13}$$

3.2.2. Gradient for the Geometric Mean

By applying the chain rule, the partial derivative of $B(\mathbf{h})$ with respect to h_n is given by

$$\frac{\partial B(\mathbf{h})}{\partial h_n} = \frac{2}{K}\sum_{k=1}^{K}\frac{1}{|g_k|}\mathrm{sign}\{g_k\}\,m_{kn}. \tag{14}$$

By defining a vector $\widetilde{\mathbf{g}}$ whose components $\widetilde{g}_k$ are given by

$$\widetilde{g}_k = \frac{1}{g_k^*}, \tag{15}$$

the gradient of $\nabla_{\mathbf{h}} B(\mathbf{h})$ can be expressed quite compact in vector notation:

$$\nabla_{\mathbf{h}} B(\mathbf{h}) = \frac{2}{K}\mathbf{M}^H\widetilde{\mathbf{g}}. \tag{16}$$

3.2.3. Overall Update Rule

The optimization problem in (4) is solved by applying a gradient-descent procedure with an adaptive step-size computed by line search [12]. The update rule reads:

$$\mathbf{h}^{l+1} = \mathbf{h}^l - \mu^l\left(\nabla_{\mathbf{h}} f(\mathbf{h}^l) + \alpha\nabla_{\mathbf{h}} s(\mathbf{h}^l)\right), \tag{17}$$

where μ^l is the adaptive step-size in iteration l, $\nabla_{\mathbf{h}} f(\mathbf{h})$ is given in [5] and $\nabla_{\mathbf{h}} s(\mathbf{h})$ is the gradient derived above; μ^l is chosen so that the value of the objective function decreases in every iteration.

Due to the special structure of the matrices $\widetilde{\mathbf{F}}$ and $\mathbf{C}$, for uniformly spaced frequencies ω_k, the gradient can be computed efficiently using the fast Fourier transform (FFT) and its inverse.

4. RESULTS

We tested the proposed approach in a real-world scenario. The RIR under investigation was measured with an exponential sine sweep [13] in a living room with a total area of 44 square meters (including the attached open kitchen) using a high-quality audio system. For the recordings we used a Beyerdynamics MM1 microphone. The sampling rate for playback and recording was $f_s = 44.1$ kHz, and the RIR has been limited to a length of $L_c = 8000$ taps. The measured RIR and its frequency response are depicted in Fig. 1. The frequency response is shown in the range from 0 Hz up to 20 kHz while we are considering only the range from 20 Hz to 20 kHz for equalization and evaluation.

For all experiments, the sampling rate was set as 44.1 kHz, and the parameters were chosen as $p_d = 20$, $p_u = 10$ (and $p_f = 8$ for the p-norm based regularization term [7]). For the weighting windows $w_d(n)$ and $w_u(n)$ we used the functions proposed in [5] that are designed to capture the average temporal masking properties of the human auditory system.

To quantify the quality of the reshaping in the time domain, we utilize the *normalized perceivable reverberation quantization* (nPRQ) measure [14]. The nPRQ captures the average overshot of the time coefficients of an impulse response that exceed the model for the average temporal masking limit according to [8] on a logarithmic scale. In the case of all time coefficients being below the average temporal masking limit or below -60 dB, we have $\mathrm{nPRQ} = 0$. The -60 dB limit

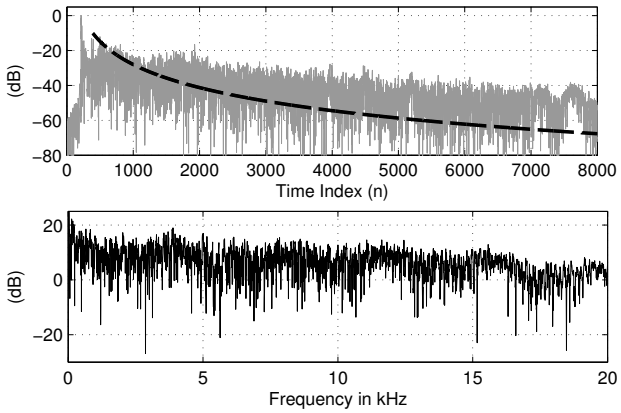

Fig. 1. The time-domain (upper plot) and frequency-domain (lower plot) representation of the measured RIR. The dashed line represents the average temporal masking limit.

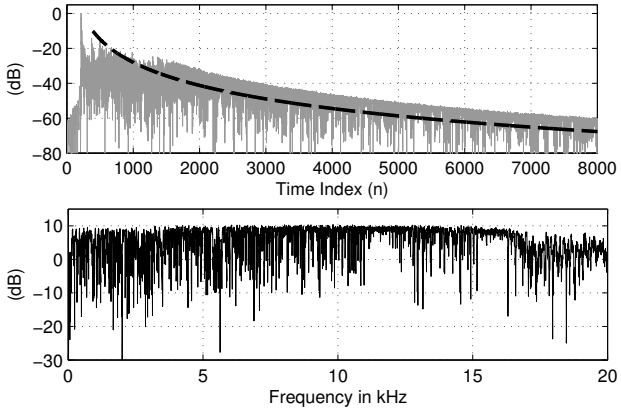

Fig. 2. The time-domain (upper plot) and frequency-domain (lower plot) representation of the reshaped impulse response using the p-norm based regularization method from [7]. The dashed line represents the average temporal masking limit.

is motivated by the definition of the reverberation time T_{60} from room acoustics [14]. To quantify the spectral distortions, we evaluate the SFM [10] in a frequency range from $f_{\text{min}} = 20$ Hz up to $f_{\text{max}} = 20$ kHz.

For all experiments, the weighting factor α was chosen empirically to yield both good overall reshaping and equalization.

For the experiment with the p-norm based regularization term from [7], we chose an equalizer length of $L_h = 8000$ and $\alpha = 15$, which yielded and acceptable SFM and still decreased the nPRQ value; the resulting overall impulse response is depicted in Fig. 2.

In Fig. 3 the result for the novel proposed approach with $L_h = 8000$ and $\alpha = 2$ is depicted. With these parameters, the proposed method achieves a reduction of the nPRQ measure from 12.91 dB down to 2.08 dB. The spectral flatness of the frequency range under investigation could be increased from 0.52 up to 0.8.

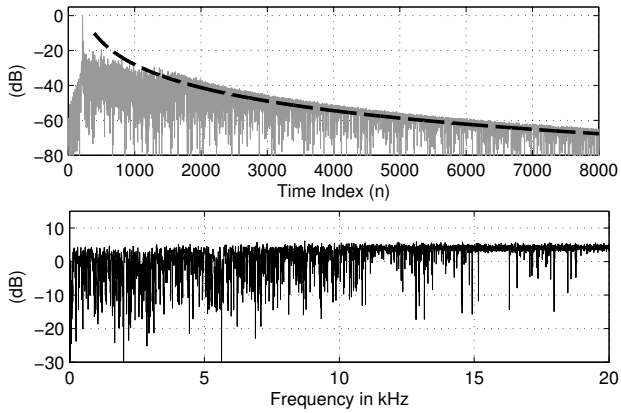

Fig. 3. The time-domain (upper plot) and frequency-domain (lower plot) representation of the reshaped impulse response using the proposed approach. The dashed line represents the average temporal masking limit.

Table 1. Results for the experiments. The SFM was computed in a range between 20 Hz and 20 kHz.

Algorithm	nPRQ [dB]	SFM
unreshaped	12.91	0.52
p-norm reg ($L_h = 8000$, $\alpha = 15$)	6.23	0.71
SFM reg ($L_h = 4000$, $\alpha = 1.5$)	5.07	0.74
SFM reg ($L_h = 8000$, $\alpha = 2$)	2.08	0.80

More results for both methods are given in Table 1. The result achieved with the p-norm based method is given in the line denoted by "p-norm reg", the results achieved with the proposed method are given in the lines denoted by "SFM reg". In comparison to the p-norm based regularization method, the experiments showed that the proposed method yields superior results in terms of nPRQ and SFM, even with much shorter filters ($L_h = 4000$ for the proposed method and $L_h = 8000$ for the p-norm based method).

5. CONCLUSIONS

In this contribution we developed a new method to regularize time-domain based reshaping algorithms in the frequency domain to yield a flat overall frequency response. In comparison to former methods, we directly optimize an established measure to capture the flatness of a frequency response. Experiments showed that the proposed method yields superior results in terms of dereverberation and equalization performance in comparison to other state-of-the-art reshaping algorithms. In addition, it allows for an equalization in a predefined frequency range. In future works we will investigate the integration of auditory scales and individual frequency weighting into the optimization problem.

6. REFERENCES

[1] John N. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[2] Stephen J. Elliott and Philip A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[3] Heinrich Kuttruff, *Room Acoustics*, Spon Press, 4 edition, 2000.

[4] Markus Kallinger and Alfred 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.

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

[6] Radoslaw Mazur, Jan Ole Jungmann, and Alfred Mertins, "On cuda implementation of a multichannel room impulse response reshaping algorithm based on p-norm optimization," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, Oct. 2011, pp. 305–308.

[7] Jan Ole Jungmann, Tiemin Mei, Stefan Goetze, and Alfred Mertins, "Room impulse response reshaping by joint optimization of multiple p-norm based criteria," in *Proceedings of the European Signal Processing Conference*, Barcelona, Spain, Aug. 2011, pp. 1658–1662.

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

[9] J. Makhoul and J . Wolf, "Linear prediction and the spectral analysis of speech," Tech. Rep., Report No. 2304, Cambridge, Mass. : Bolt, Beranek and Newman, Inc., August 1972.

[10] James D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 2, pp. 314–323, Feb. 1988.

[11] Peter J. W. Melsa, Richard C. Younce, and Charles E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Transactions on Communications*, vol. 44, no. 12, pp. 1662–1672, Dec. 1996.

[12] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[13] Angelo Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Audio Engineering Society Convention*, Vienna, Austria, May 2007.

[14] Jan Ole Jungmann, Radoslaw Mazur, Markus Kallinger, Tiemin Mei, and Alfred Mertins, "Combined acoustic mimo channel crosstalk cancellation and room impulse response reshaping," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1829–1842, Aug. 2012.