JOINT TIME- AND FREQUENCY-DOMAIN RESHAPING 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. For dereverberation, the filter is usually designed by optimizing a solely time-domain based cost function and, therefore, may introduce spectral distortions. Recent methods also consider the frequency-domain representation of the overall system and aim at yielding a flat overall frequency response. However, in some cases, it may be desirable to follow a predefined frequency response rather than obtaining a flat one. Hearing impaired persons, for example, may prefer or even need an amplification of certain frequency ranges. In this work, we propose a new method to jointly shape the time- and the frequency-domain representations of the overall acoustic channel according to prescribed curves. Furthermore, we integrate the concept of auditory scales into the filter design.

Index Terms— room impulse responses, equalization, optimization, reshaping, spectral flatness.

1. INTRODUCTION

Equalization of an acoustic channel is usually carried out by designing a filter, in such a way that the global impulse response (GIR, that is the convolution of the filter and the channel response) contains better acoustic properties than the room impulse response (RIR) itself.

Early approaches aimed at minimizing the quadratic distance between the GIR and a desired target response [1]. The target response is usually chosen as a delayed and/or bandpass filtered unit pulse. Techniques originating from the field of data transmission have been used for *shortening* of room impulse responses [2] by concentrating its energy in a certain time frame. However, in [2] it was shown that a *shaping* is

preferable over a shortening in practice. By shaping an acoustic channel, the masking properties of the human auditory system can be exploited efficiently. The optimization of a time-domain based optimality criterion may lead to spectral distortions of the overall system. In [2] it was proposed to compensate for the spectral distortions by applying a postfilter that is designed for the GIR.

In [3], the optimality criterion from [2] was generalized by replacing the quadratic norm with a general p-norm. By choosing *high* values for p, a very even shaping of the reverberant tail is be achieved while, in general, solutions with one dominant peak are preferred, resulting in a flat frequency response. However, in the case of long reverberation time, even the p-norm based method may lead to spectral distortions. In [4] it was proposed to explicitly consider the frequency response of the overall system for the filter design. The regularization term from [4] is also based on the p-norm. In [5], a new regularization term was proposed to yield a flat overall frequency response. The term from [5] is based on the spectral flatness measure, which is used to quantify the *flatness* of a spectrum [6].

In this work, we extend the method from [5] by formulating a *generalized spectral flatness measure*. The new measure is considered during filter design by a new regularization term. The proposed method allows us to define exact demands on the time- and frequency-domain representations of the overall acoustic channels. This allows us, for example, to yield reshaping filters that are designed to compensate for specific hearing loss of hearing impaired persons.

This paper is organized as follows. In Section 2, we review the p-norm based reshaping method from [3] and the regularization from [4]. Our new measure and the new optimization problem is described in Section 3. Experiments and results are described in Section 4. Finally, conclusions are drawn in Section 5.

Notation: Vectors and matrices are denoted by boldface small and capital letters, respectively. The p-norm of a vector is denoted by $\|\cdot\|_p$, $\mathrm{diag}\{\cdot\}$ transforms a vector into a diagonal matrix, and $\mathrm{sign}\{\cdot\}$ returns the sign of its argument, where the sign of a complex number is defined as its projection onto the unit circle. The superscript ^{H} denotes the Hermitian transpose of a matrix, and $*$ denotes convolution.

2. ROOM IMPULSE RESPONSE RESHAPING

A brief overview of the p-norm based filter design method from [3] is given in this section. With $c(n)$ denoting the RIR of length L_c, and $h(n)$ denoting the equalizer of length L_h, the GIR of length $L_g = L_c + L_h - 1$ is given by $g(n) = c(n) * h(n)$. Usually, two window functions $w_d(n)$ and $w_u(n)$ are specified to define a *desired* and an *unwanted* part of the GIR, respectively. The desired part is given by $g_d(n) = w_d(n) g(n)$, and the unwanted part by $g_u(n) = w_u(n) g(n)$.

2.1. p-Norm Based Reshaping

The p-norm based optimization problem from [3] reads

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

where $\boldsymbol{g}_u$ and $\boldsymbol{g}_d$ are vectors containing the unwanted and desired parts $g_u(n)$ and $g_d(n)$, of the GIR, respectively. By choosing appropriately high values for p_d and p_u (usually $10 \leq p_d, p_u$), a very even shaping of the GIR can be achieved [3]. The weighting windows are, as in [3], chosen to capture the compromise temporal masking limit of the human auditory system [7]. The vector $\boldsymbol{h} = [h_1, \ldots, h_{L_h}]^T$ contains the equalizer's impulse response $h(n)$, i.e., $h_n = h(n-1)$.

The optimization of (1) is usually carried out iteratively by applying a gradient-descent procedure [3].

2.2. Frequency Domain Based Regularization

In [4] it has been shown that a joint optimization of the time- and frequency-domain representations of the overall system yields, in general, good reshaping in the time domain, while the frequency response contains no high spectral peaks. The p-norm based regularization term from [4] reads

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

where the vector $\boldsymbol{g}_f$ contains the discrete Fourier transform (DFT) of the overall impulse response. By choosing appropriately high values for p_f (e.g., $p_f = 8$ in [4]), the overall frequency response is prohibited from containing any high spectral peaks.

3. PROPOSED METHOD

The proposed measure is based on a generalized form of the *spectral flatness measure* (SFM) [6]. The SFM is used to capture the *flatness* of a spectrum and can, of course, also be used to quantify the flatness of a frequency response. The new regularization term is based on the proposed measure and replaces the regularization term (2) from [4]. In [5] it was proposed to use a regularization term that is directly based on the SFM for filter design.

3.1. Generalized Spectral Flatness Measure

For the proposed measure, we generalize the SFM by introducing an integration of the frequency response into subbands and an individual weighting for each of the subbands. We call this measure the *generalized spectral flatness measure* (gSFM). With

$$G_s = \sum_{k=1}^{K} w_{sk} \left|G\left(e^{j\omega_k}\right)\right|^2 \tag{3}$$

the proposed measure is calculated as

$$\text{gSFM}_{g(n)} = \frac{\left(\prod_{s=1}^{S} \gamma_s G_s\right)^{\frac{1}{S}}}{\frac{1}{S} \sum_{s=1}^{S} \gamma_s G_s}, \tag{4}$$

where the coefficients w_{sk}, $k = 1, \ldots, K$, $s = 1, \ldots, S$ integrate the squared absolute values of the length-K-DFT of $g(n)$ (denoted by $G\left(e^{j\omega_k}\right)$) into S subbands. Each subband is then weighted by a positive weighting factor γ_s, $s = 1, \ldots, S$ before the arithmetic and the geometric means are computed.

The gSFM achieves its maximum of gSFM $= 1$ when all terms $\gamma_s G_s$ in (4) are equal, i.e.,

$$\gamma_1 G_1 = \gamma_2 G_2 = \ldots = \gamma_S G_S. \tag{5}$$

By maximizing the gSFM given in (4), one yields

$$\gamma_s G_s \approx \nu \quad \text{for} \quad s = 1, \ldots, S \tag{6}$$

with some $\nu > 0$. This property allows us to prescribe a desired frequency response by choosing suitable weighting factors γ_s in the gSFM. Examples of this will be presented in Section 4.

For the special case of $S = K$ (i.e., when no subband integration takes place) and with the choices $\gamma_s = 1$ for $s = 1, \ldots, S$ and $w_{sk} = \delta_{sk}$ with δ_{sk} being the Kronecker-delta, defined as

$$\delta_{sk} = \begin{cases} 1, & s = k, \\ 0, & s \neq k, \end{cases} \tag{7}$$

the gSFM (4) reduces to the SFM known from literature [6].

3.2. The New Optimization Problem

The proposed optimization problem reads

$$\min_{\boldsymbol{h}} \quad : \quad f(\boldsymbol{h}) + \alpha s(\boldsymbol{h}), \tag{8}$$

where $f(\boldsymbol{h})$ is given in (1), and the regularization term $s(\boldsymbol{h})$ is given by

$$s(\boldsymbol{h}) = -\log\left(\text{gSFM}_{g(n)}\right) \tag{9}$$

with gSFM defined in (4). The prefilter $h(n)$ is designed by solving (8) with a gradient-descent procedure. The $\log$ operation in (9) allows for a simpler expression of the required gradients. The gradient $\nabla_{\boldsymbol{h}} f(\boldsymbol{h})$ is given in [3]; the gradient $\nabla_{\boldsymbol{h}} s(\boldsymbol{h})$ is derived in the following.

3.3. Derivation of the Gradients

To derive the required gradients, we use matrix-vector notation. Let $\boldsymbol{g} = [g_1, \ldots, g_K]^T$ be a vector containing the K values of the DFT of the GIR. The vector $\boldsymbol{g}$ is computed via

$$\boldsymbol{g} = \boldsymbol{M}\boldsymbol{h} \quad \text{with} \quad \boldsymbol{M} = \widetilde{\boldsymbol{\mathcal{F}}}\boldsymbol{C}, \tag{10}$$

where $\boldsymbol{C}$ is the convolution matrix of the RIR $c(n)$, and $\widetilde{\boldsymbol{\mathcal{F}}}$ is a modified DFT matrix that contains just the rows for the discrete frequencies ω_k, $k = 1, \ldots, K$ under consideration. Furthermore, by defining a matrix $\boldsymbol{W}$ of dimension $S \times K$ that contains the values w_{sk}, $s = 1, \ldots, S$, $k = 1, \ldots, K$ as its entries, and a diagonal matrix $\boldsymbol{\Gamma}$ of size $S \times S$ with the values γ_s, $s = 1, \ldots, S$ on its main diagonal, the gSFM can be expressed by

$$\text{gSFM}_{g(n)} = \frac{\text{geomean}(\boldsymbol{\Gamma}\boldsymbol{W}\widetilde{\boldsymbol{g}})}{\text{arithmean}(\boldsymbol{\Gamma}\boldsymbol{W}\widetilde{\boldsymbol{g}})}, \tag{11}$$

where $\widetilde{\boldsymbol{g}}$ contains the squared absolute values of the DFT of $g(n)$. The subband-integration and weighting can be captured by a single matrix

$$\boldsymbol{\Psi} = \boldsymbol{\Gamma}\boldsymbol{W}, \tag{12}$$

that is of dimension $S \times K$. Its entries are denoted by ψ_{sk}.

Applying the logarithmic laws, the regularization term is given by

$$s(\boldsymbol{h}) = B(\boldsymbol{h}) - A(\boldsymbol{h}), \tag{13}$$

where

$$B(\boldsymbol{h}) = \log\left(\prod_{s=1}^{S}\left(\sum_{k=1}^{K} \psi_{sk} |g_k|^2\right)^{\frac{1}{S}}\right) \tag{14}$$

captures the geometric mean, and

$$A(\boldsymbol{h}) = \log\left(\frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K} \psi_{sk} |g_k|^2\right) \tag{15}$$

captures the arithmetic mean of (4).

The computation of the gradient $\nabla_{\boldsymbol{h}}s(\boldsymbol{h}) = \nabla_{\boldsymbol{h}}B(\boldsymbol{h}) - \nabla_{\boldsymbol{h}}A(\boldsymbol{h})$ is given in the following.

3.4. Gradient for the Arithmetic Mean Part

The derivation of $A(\boldsymbol{h})$ with respect to a coefficient h_n of the target vector $\boldsymbol{h}$ is calculated by applying the chain rule. The partial derivative is then given by

$$\frac{\partial A(\boldsymbol{h})}{\partial h_n} = \zeta_A \cdot \left(\sum_{k=1}^{K} g_k m_{kn} \sum_{s=1}^{S} \psi_{sk}\right), \tag{16}$$

with

$$\zeta_A = \frac{2}{\sum_{s=1}^{S}\sum_{k=1}^{K} \psi_{sk} |g_k|^2}, \tag{17}$$

and m_{kn} denoting the element in the k-th row and n-th column of matrix $\boldsymbol{M}$, given by (10).

By defining a vector $\widetilde{\boldsymbol{\psi}} = \left[\widetilde{\psi}_1, \ldots, \widetilde{\psi}_K\right]^T$, whose elements are given by

$$\widetilde{\psi}_k = \sum_{s=1}^{S} \psi_{sk}, \tag{18}$$

the gradient is computed via

$$\nabla_{\boldsymbol{h}}A(\boldsymbol{h}) = \zeta_A \boldsymbol{M}^H \text{diag}\left\{\widetilde{\boldsymbol{\psi}}\right\}\boldsymbol{g}. \tag{19}$$

3.5. Gradient for the Geometric Mean Part

The gradient for the geometric-mean part is, again, computed by applying the chain rule and deriving the partial derivatives for $B(\boldsymbol{h})$ with respect to the target coefficients h_n. The partial derivative is given by

$$\frac{\partial B(\boldsymbol{h})}{\partial h_n} = \frac{2}{S}\sum_{s=1}^{S}\left(\zeta_{B_s}\sum_{k=1}^{K} \psi_{sk}\text{sign}\{g_k\}m_{kn}\right) \tag{20}$$

with

$$\zeta_{B_s} = \frac{1}{\sum_{k=1}^{K} \psi_{sk} |g_k|}. \tag{21}$$

By defining a vector $\widehat{\boldsymbol{\psi}} = \left[\widehat{\psi}_1, \ldots, \widehat{\psi}_K\right]^T$, whose elements are given by

$$\widehat{\psi}_k = \sum_{s=1}^{S} \zeta_{B_s}\psi_{sk}, \tag{22}$$

the gradient $\nabla_{\boldsymbol{h}}B(\boldsymbol{h})$ is finally given by

$$\nabla_{\boldsymbol{h}}B(\boldsymbol{h}) = \frac{2}{S}\boldsymbol{M}^H\text{diag}\left\{\widehat{\boldsymbol{\psi}}\right\}\text{sign}\{\boldsymbol{g}\}. \tag{23}$$

3.6. Overall Update Rule

The overall update rule for the target vector $\boldsymbol{h}$ reads

$$\boldsymbol{h}^{l+1} = \boldsymbol{h}^l - \mu^l(\nabla_{\boldsymbol{h}}f(\boldsymbol{h}) + \alpha\nabla_{\boldsymbol{h}}s(\boldsymbol{h})), \tag{24}$$

where μ^l is the adaptive step-size in iteration l. The gradient $\nabla_{\boldsymbol{h}}f(\boldsymbol{h})$ is given in [3], and $\nabla_{\boldsymbol{h}}s(\boldsymbol{h}) = \nabla_{\boldsymbol{h}}B(\boldsymbol{h}) - \nabla_{\boldsymbol{h}}A(\boldsymbol{h})$ with $\nabla_{\boldsymbol{h}}A(\boldsymbol{h})$ and $\nabla_{\boldsymbol{h}}B(\boldsymbol{h})$ given in (19) and (23), respectively. The step-size μ^l is chosen so that the value of the cost function decreases in every iteration [8].

4. RESULTS

For the results we measured a RIR using a Cortex MK-2 manikin in a typical office room using exponential sine sweeps [9]. The sampling rate was chosen as $f_s = 16\text{kHz}$, and the RIR was limited to $L_c = 4000$ taps. The frequency response of the outer ear was compensated by applying a short prediction-error filter. The resulting RIR and its frequency response are shown in Fig. 1.
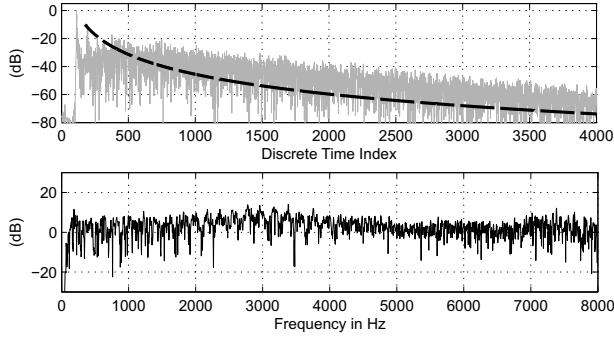
Fig. 1. Magnitude of measured RIR on a logarithmic scale (upper plot) and its frequency reponse (lower plot). The dashed line is the average temporal masking limit.

To quantify the amount of dereverberation we utilize the *normalized perceivable reverberation quantization* measure (nPRQ) measure that has been proposed in [10]. The nPRQ is defined as the average overshot of the time-domain coefficients over the compromise temporal masking limit [7] and above $-60\,\mathrm{dB}$ on a logarithmic scale. If no tap exceeds the temporal masking limit, then no reverberation should be perceivable by a human listener and thus nPRQ = $0\,\mathrm{dB}$.

For the filter design, we aimed at reshaping the impulse response in the time domain while jointly compensating for a frequency-dependent hearing loss. The desired frequency dependent amplification is derived from a fictitious audiogram, which is depicted in Fig. 2. For the experiments, the filters were designed with a length of $L_h = 4000$ taps and the parameters for the p-norm were $p_d = 20$, $p_u = 10$.
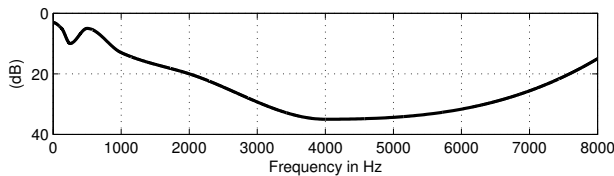
Fig. 2. Fictitious audiogram showing the frequency-dependent hearing loss. The desired amplification for each frequency is derived from this audiogram.

For a first experiment we chose W as the identity matrix, i.e., no integration into subbands. The weighting factors γ_s have been chosen as the reciprocal of the frequency-dependent hearing loss (see Fig. 2) and $\alpha = 2$. The resulting overall system is depicted in Fig. 3. The nPRQ could be reduced from $11.6\,\mathrm{dB}$ for the unreshaped RIR to $4.7\,\mathrm{dB}$.

For a second experiment, the matrix W was chosen to sum up the squared frequency response into $S = 27$ subbands according to the ERB scale. The factors γ_s have been chosen to weight each subband according to the desired amplification of its center frequency, once again, aiming at compensating the hearing loss. The weighting factor α was set to $\alpha = 0.25$. The results are depicted in Fig. 4. One can see that the compensa-
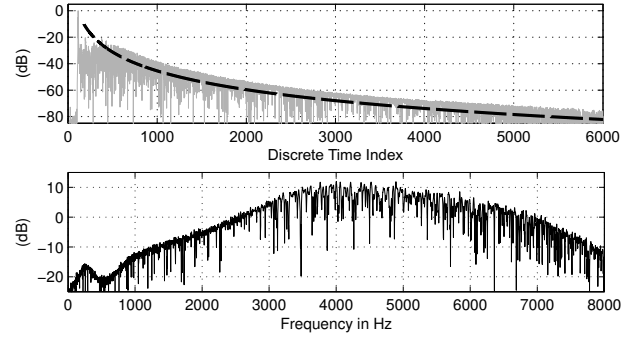
Fig. 3. Global impulse response (upper plot) and its frequency response (lower plot). The equalizer has been designed according to the proposed method with W as the identity matrix, i.e., no integration into subbands occured.

tion of the hearing loss could still be achieved, however, with some more fine-granular variation of the frequency response. The nPRQ could be reduced to $2.6\,\mathrm{dB}$.
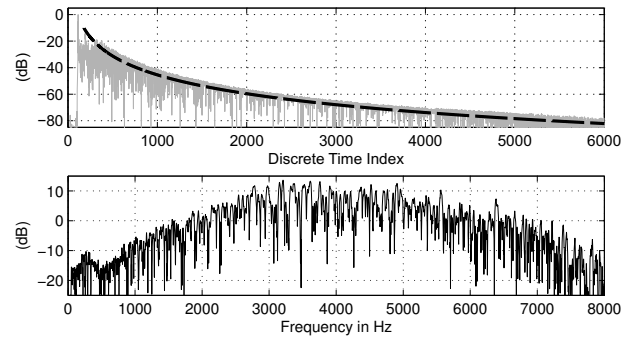
Fig. 4. Global impulse response (upper plot) and its frequency response (lower plot). The equalizer has been designed according to the proposed method. The matrix W was set up in such a way that it expressed an integration into $S = 27$ subbands according to the ERB scale.

5. CONCLUSION

In this work we proposed a new method to capture and utilize the frequency-domain representation of an impulse response during filter design. The method is based on a so-called *generalized spectral flatness measure*, which has been integrated as a regularization term into the reshaping filter design. It was shown that the proposed method allows one to design reshaping filters in such a way that the overall acoustic channel precisely fulfills predefined demands in the time and frequency domains. Furthermore, it was shown that the integration into subbands increases the quality of the time-domain shaping by relaxing the demands on the frequency-domain representation. We will pursue this approach to further exploit psychoacoustic properties of the human auditory system.

6. REFERENCES

[1] S. J. Elliott and P. 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.

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

[3] A. Mertins, T. Mei, and M. 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.

[4] J. O. Jungmann, T. Mei, S. Goetze, and A. 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. [Online]. Available: http://www.eusipco2011.org/

[5] J. O. Jungmann, R. Mazur, and A. Mertins, "Joint time-domain reshaping and frequency-domain equalization of room impulse responses," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florence, Italy, May 2014, pp. 6692–6696.

[6] J. I. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

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

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

[9] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Convention of the Audio Engineering Society*, Vienna, Austria, May 2007. [Online]. Available: http://pcfarina.eng.unipr.it/Public/

[10] J. O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. 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.