On the Equalization and Reshaping of Room Impulse Responses

Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins
Universität zu Lübeck, Institute for Signal Processing, Ratzeburger Allee 160, D-23562 Lübeck
Email: jungmann@isip.uni-luebeck.de

Abstract

In listening room compensation, the aim is to compensate for the acoustic channel between loudspeaker and listener. The degradations rendered to the source signal by reproduction in a closed room are namely reverberation and spectral distortions. A pre- or postfilter is designed in such a way that the convolution of the equalizer and the room impulse response contains better acoustic properties than the room impulse response itself. For dereverberation, the equalizer is usually designed by optimizing a time-domain based objective function. Recent methods also consider the frequency-domain representation to yield a flat overall frequency response as well. In this work, we propose to use a generalized spectral flatness measure to quantify the disparity between the actual frequency response and a desired one. This allows us to integrate the concept of auditory scales into the equalizer design, and thus, to take into account some properties of the human auditory system. Results are presented for an acoustic channel measured in a typical living room.

Introduction

The distortions rendered to an audio signal by reproduction in a closed room are reverberation and spectral distortions. The goal of room impulse response (RIR) reshaping is to design an equalizer in such a way that the convolution of the equalizer and the RIR (the global impulse response, GIR) contains better acoustic properties than the RIR itself. The problem of RIR reshaping is connected to the problem of deconvolution. However, recent approaches do not target a deconvolution but aim at a GIR that introduces no audible distortions. It was proposed recently to use a p-norm based optimality criterion that exploits some average temporal masking properties of the human auditory system [1].

RIR Reshaping by p-Norm Optimization

In [1] the common quadratic cost function for designing reshaping filters was generalized to a p-norm based optimality criterion. The optimization problem reads

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

where $\mathbf{h}$ is the target vector containing the equalizer and $\mathbf{g}_u$ and $\mathbf{g}_d$ are made up by the *unwanted* and the *desired* parts of the GIR $g(n)$. The values p_u and p_d are usually chosen as $10 \leq p_d, p_u \leq 20$, but also the case $p_d = p_u = 2$, for which a direct solution is obtained, is included in the framework. The optimization of (1) for $p_{u,d} \neq 2$ is carried out by applying a gradient-descent procedure.

Frequency Domain Based Regularization

While the reshaping according to [1] usually yields a flat overall frequency response, there may be cases in which spectral distortions occur. In [2] we proposed a p-norm based regularization term that captures the frequency-domain representation of the GIR. However, the regularization term from [2] does not take into account any properties of the human auditory system.

The *spectral flatness measure* (SFM) is defined as the ratio of the geometric and arithmetic means of the values of a power spectral density [3, 4]. For filters, the ratio of the geometric and arithmetic means of the squared magnitude frequency response is considered. In the case of a perfectly flat frequency response the SFM equals one and degrades to zero with increasing spectral distortions.

For the proposed approach, we integrate the squared absolute values of the overall frequency response into subbands before computing the geometric and arithmetic means. We call this approach the *integrated spectral flatness measure* (iSFM). In addition, we generalize the measure by introducing weighting factors for the resulting frequency bands.

With $\tilde{\mathbf{g}}$ containing the squared absolute values of the discrete Fourier transform (DFT) of the GIR, the proposed generalized measure reads

$$Q(\mathbf{h}) = \frac{\mathrm{geomean}\,(\mathbf{\Gamma W}\tilde{\mathbf{g}})}{\mathrm{arithmean}\,(\mathbf{\Gamma W}\tilde{\mathbf{g}})}, \tag{2}$$

where $\mathbf{W}$ is a $K \times L_g$ matrix that integrates the L_g elements of $\tilde{\mathbf{g}}$ into K subbands, $\mathbf{\Gamma}$ is a diagonal matrix of positive weighting factors γ_k, $k = 1, 2, \ldots, K$, and $\mathrm{geomean}(\cdot)$ and $\mathrm{arithmean}(\cdot)$ determine the geometric and arithmetic mean of the elements of a vector, respectively. The proposed method is quite flexible. When setting both $\mathbf{\Gamma}$ and $\mathbf{W}$ as identity matrices, the standard SFM is obtained. By selecting different values for γ_k, $k = 1, 2, \ldots, K$, frequency bands can be weighted differently. Matrix $\mathbf{W}$ can be adjusted to different auditory scales. In the following, we set $\mathbf{\Gamma} = \mathbf{I}$ and consider the iSFM only.

In order to obtain a simpler expression of the required gradient we take the logarithm of (2) with $\mathbf{\Gamma} = \mathbf{I}$. The resulting criterion reads

$$s(\mathbf{h}) = -\log Q(\mathbf{h}) = B(\mathbf{h}) - A(\mathbf{h})\,, \tag{3}$$

where

$$B(\mathbf{h}) = \log\left(\left(\prod_{k=1}^{K}\left(\sum_{\ell=1}^{L_g} w_{k\ell}\,|g_\ell|^2\right)\right)^{\frac{1}{K}}\right) \tag{4}$$

captures the geometric mean and

$$A(\mathbf{h}) = \log\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{\ell=1}^{L_g} w_{k\ell}\,|g_\ell|^2\right) \tag{5}$$

captures the arithmetic mean; g_ℓ denotes the ℓ-th element of the DFT of the GIR, computed as $\mathbf{g} = \mathbf{Mh}$ with $\mathbf{M} = \mathbf{FC}$,

where $\mathbf{F}$ is the DFT matrix, and $\mathbf{C}$ is the convolution matrix made up by $c(n)$ of compatible size; $w_{k\ell}$ denotes the element in the k-th row and ℓ-th column of the weighting matrix $\mathbf{W}$.

With (1) and (3) the overall optimization problem is now given by

$$\min_{\mathbf{h}} \quad : \quad f(\mathbf{h}) + \alpha s(\mathbf{h})\,. \tag{6}$$

For the calculation of the gradient $\nabla_{\mathbf{h}} s(\mathbf{h})$ the terms $\nabla_{\mathbf{h}} A(\mathbf{h})$ and $\nabla_{\mathbf{h}} B(\mathbf{h})$ are required. Due to the limited space, just the partial derivatives of $A(\mathbf{h})$ and $B(\mathbf{h})$ with respect to the n-th entry of the target vector $\mathbf{h}$ (denoted by h_n) are given here. With $m_{\ell n}$ being the element of $\mathbf{M}$ in the ℓ-th row and n-th column, we obtain

$$\frac{\partial A(\mathbf{h})}{\partial h_n} = \zeta_A \cdot \left(\sum_{k=1}^{K} \sum_{\ell=1}^{L_g} w_{k\ell} g_\ell m_{\ell n} \right) \tag{7}$$

with

$$\zeta_A = \frac{2}{\sum_{k=1}^{K} \sum_{\ell=1}^{L_g} w_{k\ell} \left| g_\ell \right|^2} \tag{8}$$

and

$$\frac{\partial B(\mathbf{h})}{\partial h_n} = \frac{2}{K} \sum_{k=1}^{K} \left(\zeta_{B_k} \cdot \sum_{\ell=1}^{L_g} w_{k\ell} \operatorname{sign}\{g_\ell\} m_{\ell n} \right) \tag{9}$$

with

$$\zeta_{B_k} = \frac{1}{\sum_{\ell=1}^{L_g} w_{k\ell} \left| g_\ell \right|}. \tag{10}$$

According to (6) the final learning rule reads

$$\mathbf{h}^{l+1} = \mathbf{h}^{l} - \mu^{l} \left(\nabla_{\mathbf{h}} f\left(\mathbf{h}^{l}\right) + \alpha \nabla_{\mathbf{h}} s\left(\mathbf{h}^{l}\right) \right) \tag{11}$$

with $\nabla_{\mathbf{h}} s(\mathbf{h}) = \nabla_{\mathbf{h}} A(\mathbf{h}) - \nabla_{\mathbf{h}} B(\mathbf{h})$ and $\nabla_{\mathbf{h}} f\left(\mathbf{h}^{l}\right)$ from [1].

For the actual implementation of (7) and (9), one needs to take into account the symmetry of the DFT. The computations can be implemented efficiently using the fast Fourier transform and the Hadamard product.

Results

For the experiment, a RIR was measured using a high-quality audio system in a typical living room. The RIR was measured by playing back an exponential sine sweep and using a Beyerdynamics MM1 microphone for recording. The sampling rate for playback and recording was set as $f_s = 44.1$ kHz.

The weighting matrix $\mathbf{W}$ was chosen to contain 27 non-overlapping equivalent rectangular bandwidth (ERB) bands up to a maximum frequency of 20.25 kHz. This frequency limit takes care of the effect of the reconstruction lowpass filter of the playback equipment, which is quite visible in Figure 1 (lower plot). Results for the proposed method ($\alpha = 0.5$, length of the filter: 8000 taps) are depicted in Figure 1. The nPRQ measure [5] that aims to quantify the audible reverberation could be reduced from 12.9 dB for the RIR to 1.3 dB for the GIR. Moreover, unlike the original RIR, the obtained overall system has a reasonably flat frequency response, and, in particular, it no longer contains the high spectral peaks for the very low frequencies.

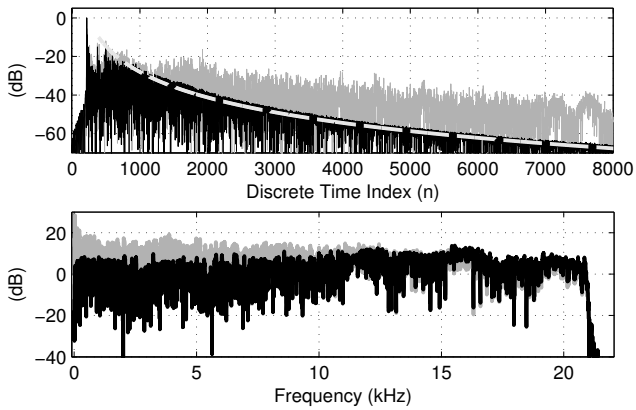

Figure 1: *Upper plot:* Coefficient magnitudes of the measured RIR (light gray) and the reshaped one (black); the dashed line is the average temporal masking limit of the human auditory system according to [6]. *Lower plot:* Frequency response before (light gray) and after reshaping (black).

Conclusions

In this work, we proposed to integrate the squared absolute values of the frequency-domain representation of the GIR into several bands before computing the SFM. The weighting matrix that defines the subbands was chosen with respect to the concept of critical bands. With the proposed method we could yield a good reshaping in the time domain while spectral distortions could be reduced. In future work, we will further investigate the influence of different weighting matrices of the generalized spectral flatness measure in order to further increase the quality of the reshaping results.

References

[1] 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*, 18(2):249–259, February 2010.

[2] 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. European Signal Processing Conference*, pages 1658–1662, Barcelona, Spain, August 2011.

[3] J. Makhoul and J. Wolf. Linear Prediction and the Spectral Analysis of Speech. Techn. Report, No. 2304, Cambridge, Mass. Bolt, Beranek und Newman, Inc., 1972.

[4] J. D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE J. Selected Areas in Communications*, 6(2):314–323, February 1988.

[5] J. O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. Mertins. Combined acoustic mimo channel crosstalk cancellation and room impulse response reshaping. *IEEE Trans. Audio, Speech, and Language Processing*, 20(6):1829–1842, August 2012.

[6] L. D. Fielder. Practical Limits for Room Equalization. AES Convention, vol. 111, preprint no. 5481. New York, NY, USA, September 2001.