PERTURBATION OF ROOM IMPULSE RESPONSES AND ITS APPLICATION IN ROBUST LISTENING ROOM COMPENSATION

Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins

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

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 filter design is usually carried out by solving an optimization problem.

There are, in general, two possibilities to improve the robustness of the equalizers against small movements of the listener and/or receiver; namely multi-position approaches or the utilization of a regularization term. Multi-position approaches suffer from the extensive effort of measuring multiple room impulse responses. Stochastic models may describe the average system error due to spatial mismatch, but only quadratic penalty terms have been considered so far.

In this contribution we propose a third method to improve robustness against spatial misalignment. We combine the two approaches by generating multiple realizations of distorted room impulse responses and feeding them into the multi-position algorithm. Based on our previous work, we propose a model to capture the perturbations with respect to the assumed displacement.

***Index Terms*—** room impulse response, RIR reshaping, p-norm, spatial robustness.

1. INTRODUCTION

In listening room compensation (LRC) one aims at neutralizing the convolutional distortions that are added to an audio signal by reproduction in a closed room. For that purpose, a filter is placed in front of the loudspeaker to preprocess the audio signal. The goal is to reduce the influence of the room impulse response (RIR) in order to obtain a signal that is hardly distinguishable from the source signal by a human listener [1]. The prefilters are designed in such a way that the global impulse response (GIR, that is the convolution of the RIR and the equalizer) satisfies certain requirements. Early approaches minimized the mean squared error between the GIR and a desired target system [2].

More recent approaches take into account the psychoacoustic properties of the human auditory system and aim at a *shaping* of the GIR [3]. In [4] the least-squares method has been generalized to a p-norm based optimality criterion. The method has been further extended to explicitly control the frequency response of the overall system [5].

Unfortunately, all of these approaches lack spatial robustness. In the case of small spatial mismatch (e.g. due to the listener moving his head slightly) the performance of the equalizer degrades greatly [6]. There are, in general, two approaches to improve spatial robustness which are discussed in [7]. The first one is the *multi-position* approach [8]. The equalizers are designed to achieve reshaping at multiple positions inside the listening area. If the spatial sampling of the RIRs is dense enough and the reshaping is successful, then the listener is allowed to move inside the listening area without perceiving a degraded quality. The second method is to consider the system errors in the optimization problem by introducing an additional regularization term [9]. In [7] a stochastic model with an arbitrary weighting for the reverberant tail was presented. Furthermore, in [7] the approach from [5] has been extended to the robust design methods in order to guarantee a flat overall frequency response.

In this paper we propose a third method to achieve robust reshaping filters. Based on one measured RIR, we generate multiple instances of the perturbed RIR and feed them into the multi-position algorithm. The perturbation term fits both the spectral and temporal properties we demand.

This paper is organized as follows. In Section 2 we give an overview of the multi-position p-norm based reshaping algorithm and the frequency-domain based regularization term. In Section 3 we briefly review the model to capture the system perturbations in case of spatial mismatch from [6] and present our algorithm to generate perturbed instances of the RIR. Results are given in Section 4. Finally, we give some conclusions in Section 5.

Notation: Lowercase boldface characters denote vectors. The asterisk $*$ denotes convolution, $\|\cdot\|_p$ returns the ℓ_p-norm of a vector, and $\mathrm{E}\{\cdot\}$ is the expectation operator.

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

2. ROOM IMPULSE RESPONSE RESHAPING

In this section we give a brief overview of the multi-position reshaping algorithm from [7]. In a setup consisting of N_s loudspeakers and R measured RIRs in the listening area, we denote the r-th sampled RIR of length L_c from loudspeaker ℓ by $c_\ell^{(r)}(n)$; the prefilter for the ℓ-th loudspeaker is denoted by $h_\ell(n)$ and is of length L_h. The r-th GIR $g^{(r)}(n)$ of length $L_g = L_c + L_h - 1$ is given by $g^{(r)}(n) = \sum_{\ell=1}^{N_s} h_\ell(n) * c_\ell^{(r)}(n)$. The reshaping filters are designed by defining two window functions $w_d(n)$ and $w_u(n)$ to determine the *desired* and the *unwanted* parts of the GIRs. The desired parts are given by $g_d^{(r)}(n) = g^{(r)}(n)\, w_d(n)$, and the unwanted parts accordingly.

2.1. Multi-Position Reshaping by p-Norm Optimization

The time-domain representation of the GIRs is optimized by solving the optimization problem given by

$$\min_{\mathbf{h}} \quad f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \tag{1}$$

with

$$f_d(\mathbf{h}) = \|\mathbf{g}_d\|_{p_d} = \left(\sum_{r=1}^{R} \sum_{n=0}^{L_g-1} \left|g_d^{(r)}(n)\right|^{p_d}\right)^{\frac{1}{p_d}} \tag{2}$$

and $f_u(\mathbf{h}) = \|\mathbf{g}_u\|_{p_u}$, accordingly. The vectors $\mathbf{g}_d$ and $\mathbf{g}_u$ are constructed by stacking up all wanted and unwanted parts of the GIRs, respectively. The target vector $\mathbf{h} = \left[\mathbf{h}_1^\top, \ldots, \mathbf{h}_{N_s}^\top\right]^\top$ is made up by the concatenation of the prefilters for the N_s loudspeakers. The optimization is carried out by applying a gradient-descent procedure.

In comparison to common least-squares methods, the advantage of (1) is that by choosing appropriately large values for p_d and p_u (usually between 10 and 20), a very even shaping of the GIRs according to the prescribed decay behavior is achieved. For the weighting we use window functions from [4] that capture the temporal masking effect of the human auditory system.

2.2. Frequency Domain Based Regularization

In [5] it has been shown that one has to consider both the time- and frequency-domain representations of the GIRs to achieve a *good* reshaping without degrading the perceived quality through spectral distortions. In [7] the method from [5] has been extended to arbitrary multi-channel setups. The regularization term from [7] is given by

$$y(\mathbf{h}) = \|\mathbf{g}_f\|_{p_f}\,, \tag{3}$$

where $\mathbf{g}_f$ is constructed by stacking up the discrete Fourier transforms of the GIRs. The regularization term forces the overall system to not contain any high spectral peaks.

2.3. Comprehensive Objective Function

By combining the two optimality criteria presented in this section, a comprehensive optimization problem is given as in [7] by

$$\min_{\mathbf{h}} \quad f(\mathbf{h}) + \alpha y(\mathbf{h}) \quad \text{s.t.} \quad \mathbf{h}^\top\mathbf{h} = 1. \tag{4}$$

The factor α weights the demand on the frequency response against the reshaping of the time-domain coefficients of the GIRs. The derivation of the required gradient is given in [7].

3. ROBUST RESHAPING USING PERTURBED ROOM IMPULSE RESPONSES

In this section we derive a model to generate additive perturbations that describe the distortions of the RIR caused by microphone movement. In the case of spatial mismatch from the reference position, the perturbed RIR $\hat{c}(t)$ is expressed by

$$\hat{c}(t) = c(t) + p(t)\,, \tag{5}$$

where $c(t)$ is the RIR in the reference position and $p(t)$ is the perturbation caused by microphone movement.

The perturbations are modeled as random signals with specific properties in the frequency and time domains.

3.1. Perturbation Properties in the Frequency Domain

The problem of designing an equalizer for a reference position and then moving the microphone away has been studied by Radlović et al. [6]. In their work they derived a frequency dependent error term for the system perturbations of the equalized overall system.

Let $\omega = 2\pi f$ denote the radial frequency and let $C(\omega)$, $P(\omega)$ and $H(\omega)$ be the Fourier transforms of the RIR $c(t)$, its perturbation $p(t)$ caused by microphone movement, and the equalizer $h(t)$, respectively. The frequency-dependent error term is then given as in [6] by

$$Q(\omega) = \mathrm{E}\left\{\left|\left[C(\omega) + P(\omega)\right] H(\omega) - 1\right|^2\right\}. \tag{6}$$

Being in the far field in reverberant environments and assuming perfect equalization at the reference position (e.g. $H(\omega) = 1/C(\omega)$), the distance measure (as derived in [6]) is given by

$$Q(\omega) \cong \frac{\mathrm{E}\left\{|P(\omega)|^2\right\}}{|C(\omega)|^2} = 2 - 2\frac{\sin(\omega D/v)}{\omega D/v}, \tag{7}$$

where D is the distance to the reference position in meters and v is the speed of sound. Solving (7) for $\mathrm{E}\left\{|P(\omega)|^2\right\}$ yields

$$\mathrm{E}\left\{|P(\omega)|^2\right\} = |C(\omega)|^2\left(2 - 2\frac{\sin(\omega D/v)}{\omega D/v}\right). \tag{8}$$

3.2. Perturbation Properties in the Time Domain

The distribution of energy across the time coefficients of a RIR has been studied by Polack. In [10] a RIR is modeled as one realization of a non-stationary stochastic process. In this model, a RIR is described by a stationary random noise process that is weighted by an exponentially decaying function. The decay of the exponential function is directly linked to the reverberation time, T_{60}, of a room. The RIR is given by

$$c(t) = \begin{cases} 0, & t < 0 \\ b(t)\, e^{-\Delta t}, & t \geq 0 \end{cases} \tag{9}$$

with $b(t)$ being a white zero-mean Gaussian stationary noise with variance σ^2 and Δ given by $\Delta \hat{=} \frac{3 \ln(10)}{T_{60}}$; for the sake of simplicity we assume $\sigma^2 = 1$.

Given (9), the energy envelope of the RIR is expressed by

$$\mathrm{E}\left\{c^2(t)\right\} = e^{-2\Delta t}. \tag{10}$$

By assuming a RIR and a nearby RIR having the same decay behavior, expressed by

$$\mathrm{E}\left\{c^2(t)\right\} = \mathrm{E}\left\{\hat{c}^2(t)\right\} = e^{-2\Delta t}, \tag{11}$$

we need to slightly modify the additive model from (5) to

$$\hat{c}(t) = \gamma c(t) + A p(t)\,, \tag{12}$$

with $0 \leq \gamma < 1$ and $0 \leq A$ being weighting factors to guarantee that the energy and the energy envelope of the reference RIR and a nearby RIR can be equivalent.

Assuming the RIRs and perturbation being mutually independent and $\mathrm{E}\{p(t)\} = 0$, the energy envelope of the perturbation is computed as

$$\begin{aligned} \mathrm{E}\left\{c^2(t)\right\} &= \mathrm{E}\left\{\hat{c}^2(t)\right\} \\ &= \mathrm{E}\left\{\left(\gamma c(t) + A p(t)\right)^2\right\} \\ &= \mathrm{E}\left\{\gamma^2 c^2(t)\right\} + \mathrm{E}\left\{A^2 p^2(t)\right\}. \end{aligned} \tag{13}$$

By inserting (11) into (13) we obtain

$$\mathrm{E}\left\{A^2 p^2(t)\right\} = \left(1 - \gamma^2\right) e^{-2\Delta t}. \tag{14}$$

With the energy decay behavior of the perturbation given by (14) we further refine our model by considering the assumed spatial displacement. Denoting the time taken by the direct sound by t_0, the assumed spatial displacement by D, and v being the speed of sound, the energy envelope of the perturbation is finally given by

$$\mathrm{E}\left\{p^2(t)\right\} = \begin{cases} 0, & t < t_0 - \frac{D}{v} \\ 1, & t_0 - \frac{D}{v} \leq t < t_0 + \frac{D}{v} \\ e^{-2\Delta\left(t - t_0 - \frac{D}{v}\right)}, & t \geq t_0 + \frac{D}{v}, \end{cases} \tag{15}$$

where we normalized $\mathrm{E}\left\{p^2(t)\right\}$ to have a maximum value of one. Equation (15) captures the decay behavior of the time coefficients of the perturbations as well as the time taken by the direct sound pulse in correspondence to the assumed spatial displacement.

3.3. Proposed Model

In Sections 3.1 and 3.2 we have derived the spectral and temporal properties of the perturbations. Assuming a band-limited input signal with a maximum radial frequency of ω_c and fulfilling the sampling theorem, the continuous-time signals, impulse responses and envelopes can be replaced by their discrete-time equivalents; accordingly, (8) is sampled at discrete frequencies. To generate a single realization of the perturbation, we perform the following steps:

1. Generate a zero-mean Gaussian white noise $p^{(r)}(n)$ with unit variance.
2. Multiply the DFT of $p^{(r)}(n)$ by $\mathrm{E}\left\{|P(\omega)|^2\right\}^{\frac{1}{2}}$.
3. Apply the IDFT and multiply the result by $\mathrm{E}\left\{p^2(nT)\right\}^{\frac{1}{2}}$ from (15) to shape the sequence according to the desired decay, where $T = 1/f_s$ with f_s being the sampling frequency.

The resulting signal $p^{(r)}(n)$ approximates the desired properties in both the time- and frequency-domain. By introducing additional normalization stages, Steps 2 and 3 can be iterated for a better approximation of the desired properties.

4. RESULTS

In a first step we investigate the quality of the generated perturbations without iterating Steps 2 and 3 from Section 3.3. For that we generated $R = 5000$ instances $p^{(r)}(n)$ and give plots for the average power spectrum and the average energy of the time coefficients in Fig. 1. For comparison purposes we also depict the desired power spectrum and energy envelope as given by (8) and (15).

For the reshaping experiments we used $N_s = 4$ loudspeakers for playback in a typical office room. We measured four impulse responses $c_\ell(n)$ of length $L_c = 4000$ taps with a sampling frequency $f_s = 16$ kHz. The reshaping filters were designed with a length of $L_h = 5000$ taps. The additional parameters were chosen as $p_d = 20$, $p_u = 10$, and $p_f = 8$ for all experiments. We utilize the nPRQ [7] and the *spectral flatness measure* (SFM) [11] to quantify the amount of audible reverberation and spectral distortions. The nPRQ measure captures the average overshot of the time coefficients of an impulse response above the average temporal masking limit and above -60 dB; the SFM is one in the case of a *flat* frequency response and degrades to zero with increasing spectral distortions.

According to the algorithm from Section 3.3 we generated R perturbed versions of the N_s RIRs with $c_\ell^{(0)}(n) = c_\ell(n)$. The perturbed RIRs were generated according to

$$c_\ell^{(r)}(n) = \gamma c_\ell(n) + A p^{(r)}(n)\,, \tag{16}$$

where A and γ were chosen so that the normalized system misalignment $M_{\mathrm{dB}} = -10 \log_{10}\left(\frac{\gamma^2 \mathbf{c}^\top \mathbf{c}}{A^2 \mathbf{p}^{(r)\top} \mathbf{p}^{(r)}}\right)$ achieved a

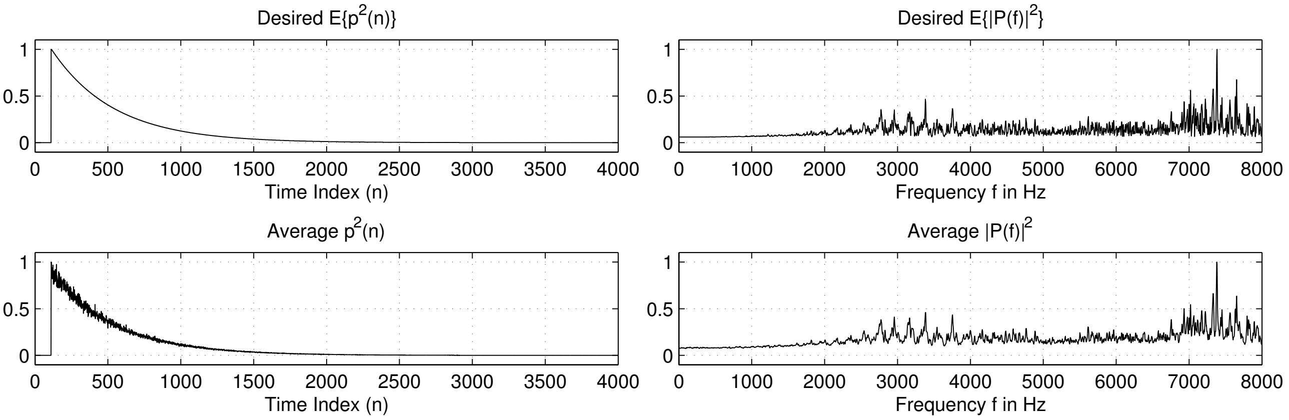

Fig. 1. Desired (top) and average (bottom) temporal and spectral properties of the perturbations. The average was calculated over 5000 random instances.

Table 1. Average values for nPRQ and SFM for different reshaping algorithms.

Setup	nPRQ [dB]	SFM
unreshaped	9.93	0.63
non-robust, $\alpha = 40$	10.50	0.64
multi-position, $\alpha = 10$	1.23	0.70
stat-robust, $\beta = 5 \cdot 10^{-4}$, $\alpha = 5$	3.92	0.64
$R = 14$, $M_{\mathrm{dB}} = -10$ dB, $\alpha = 1$	$\mathbf{3.91}$	$\mathbf{0.62}$
$R = 29$, $M_{\mathrm{dB}} = -15$ dB, $\alpha = 1$	$\mathbf{3.47}$	$\mathbf{0.66}$

prescribed value and all perturbed RIRs contained the same energy as the reference RIR. We then used the $N_s \cdot (R + 1)$ RIRs given by $c_\ell^{(r)}(n)$, $0 \leq r \leq R$ as a basis for the equalizer design.

To investigate the spatial robustness we averaged the nPRQ and SFM values over 40 microphone positions in the vicinity of the reference position. To compare the proposed method with the multi-position approach from [7], we measured 26 more RIRs around the reference position according to the spatial sampling theorem of RIRs and used them for the multi-position method. A comparison of the results for the non-robust, the multi-position, and the stochastic-penalty-term method (denoted by "stat-robust", weighted by β) from [7] with the proposed method with different values for R and M_{dB} is given in Table 1. For the experiments we assumed a displacement of $D = 2$ cm.

The factor α for the weighting of the frequency-domain based regularization term (3) from [7] has been chosen so that the resulting equalizers did not introduce significant spectral distortions according to the SFM.

It can be seen that the proposed approach improves the nPRQ measure by about 6.5 dB. Of course, the performance of the multi-position approach, which requires 26 RIR measures around the reference position, can not be reached. However, the proposed approach is superior in terms of nPRQ and SFM values compared to the utilization of the quadratic penalty term from [7]. The proposed method requires to store $N_s \cdot (R + 1)$ RIRs of length L_c while the method from [7] keeps N_s matrices of size $L_h \times L_h$ in memory. A depiction of a GIR in the case of spatial mismatch for the non-robust and the proposed design method is given in Fig 2.

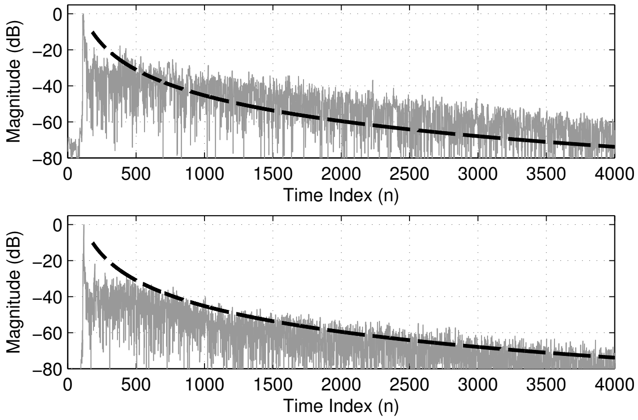

Fig. 2. GIR in the case of small spatial mismatch for the non-robust (top) and the proposed design method (bottom). The dashed line is the average temporal masking limit.

5. CONCLUSIONS

In this contribution we proposed a new method to improve the spatial robustness of LRC filters without the need of extensive measurements. Results show that our method gives results that are comparable to other state-of-the-art algorithms while having a low memory footprint. In future work we will refine this method in order to further improve the robustness. Compared to previous work, this contribution can be seen as a continuation of the methods we presented in [7].

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] 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.

[4] 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, 2010.

[5] Jan Ole Jungmann, Tiemin Mei, Stefan Goetze, and Alfred Mertins, "Room impulse response reshaping by joint optimization of multiple p-norm based criteria," in *Proc. 19th European Signal Processing Conference (EUSIPCO 2011)*, Barcelona, Spain, Aug. 2011, pp. 1658–1662.

[6] Biljana D. Radlović, R. C. Williamson, and Rodney A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.

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

[8] Tiemin Mei and Alfred Mertins, "On the robustness of room impulse response reshaping," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC 2010)*, Tel Aviv, Israel, Aug. 2010.

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

[10] Jean-Dominique Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.

[11] 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.