A Stochastic Approach for Robust Listening Room Compensation

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

The purpose of room impulse response (RIR) 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. 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. The perturbations introduced by small movements can be described by a stochastic error term. However, only quadratic penalty terms have been considered so far.

In this contribution we propose a third method to improve the robustness against spatial misalignment. We combine the two approaches by feeding multiple realizations of the distorted RIR into the multi-position algorithm. We propose a simple, yet effective model to capture the perturbations and give comparative results against other state of the art methods.

Introduction

The purpose of room impulse response (RIR) reshaping is to suppress the audible distortions that are rendered to an audio signal by playback over loudspeakers in a closed room. The distortions, namely reverberation, are usually not canceled out but suppressed in such a way that they are inaudible to a human listener. For the equalizer design, the joint optimization of multiple p-norm based criteria has shown to yield a uniform reshaping while retaining a flat overall frequency response [1].

One problem is the lack of spatial robustness, as RIRs are very sensible to movement. Even slight movements of the listener result in a degradation of the perceived quality. There are mainly two approaches to design equalizers that are robust against small movements. The methods rely either on stochastic regularization terms or on multiple measurements of additional RIRs in the vicinity of the reference position; *state-of-the-art* implementations of both approaches are described in [1]. While the stochastic error term from [1] demands to keep a huge amount of data in memory, the multi-position approach requires extensive measurements of RIRs. In this contribution we present a third method that is based on artificially perturbed RIRs. We derive certain properties that we demand on the perturbations. We then randomly generate multiple instances of perturbed RIRs and feed them into the multi-position reshaping algorithm from [1].

Multi-Position RIR Reshaping by p-Norm Optimization

The setup under investigation consists of N_s loudspeakers and R sampled RIRs in the listening area. The acoustic channel from the ℓ-th speaker to the r-th position is denoted by $c_\ell^{(r)}(n)$. The global impulse response (GIR) at the r-th sampling point is given by $g^{(r)}(n) = \sum_{\ell=1}^{N_s} c_\ell^{(r)}(n) * h_\ell(n)$, where $h_\ell(n)$ is the prefilter for the ℓ-th loudspeaker and $*$ denotes convolution. Two window functions $w_d(n)$ and $w_u(n)$ are used to characterize the desired and the unwanted part of each GIR. The desired part consists of four milliseconds starting with the direct sound pulse, whereas the weighting window for the unwanted part captures the temporal masking properties of the human auditory system [1].

The vector $\mathbf{g}_d$ is constructed by stacking up the wanted parts $g_d^{(r)}(n) = w_d(n)\, g^{(r)}(n)$ of all GIRs ($\mathbf{g_u}$, accordingly), and the vector $\mathbf{h}$ contains the concatenation of all prefilters $\mathbf{h}_\ell$. Furthermore, $\mathbf{g}_f$ contains the discrete Fourier transforms of all GIRs. The overall optimization problem is then given by

$$\min_{\mathbf{h}} \; : \; \log\left(\frac{\|\mathbf{g}_u\|_{p_u}}{\|\mathbf{g}_d\|_{p_d}}\right) + \alpha \,\|\mathbf{g}_f\|_{p_f} \,, \text{ s.t. } \mathbf{h}^\top \mathbf{h} = 1, \quad (1)$$

where the additional regularization term (weighted by α) guarantees a flat overall frequency response [1]. The optimization is carried out by a gradient-descent procedure with an adaptive step-size.

Robust Reshaping Using Artificial Perturbations

In this section we derive a simple method to generate additive perturbations to model the distortions of a RIR caused by spatial mismatch. The presented algorithm is then used to generate multiple instances of perturbed RIRs.

Perturbation Properties in the Time Domain

To derive our model we describe a perturbed RIR $\widehat{c}(t)$ by

$$\widehat{c}(t) = \gamma c(t) + Ap(t) \,, \quad (2)$$

where $c(t)$ is the RIR in the reference position and $p(t)$ describes the perturbations of the RIR in the case of spatial mismatch; γ and A are weighting factors (with $0 \leq \gamma < 1$ and $0 \leq A$) to ensure that the energy of a perturbed RIR can be equal to the RIR in the reference position.

It is a common approach to model the perturbations due to measurement errors or misalignment as stationary white noise and considering them during the optimization process. Although increasing the robustness, the assumption of stationarity does not hold in practice due to the exponentially decaying behavior of RIRs.

In the model from Polack [2] a RIR is described as one realization of a non-stationary stochastic process. In this model a RIR is given by

$$c(t) = \begin{cases} \epsilon(t)\, e^{-\Delta t}, & t \geq 0 \\ 0, & t < 0, \end{cases} \tag{3}$$

where $\epsilon(t)$ is stationary Gaussian noise with zero mean (for the sake of simplicity we assume unit variance) and Δ is linked to the reverberation time T_{60} by $\Delta \hat{=} \frac{3\ln(10)}{T_{60}}$ [2]. With (3), the energy envelope of a RIR is given by $\mathrm{E}\left\{c^2(t)\right\} = e^{-2\Delta t}$.

Starting with $\mathrm{E}\left\{c^2(t)\right\} = \mathrm{E}\left\{\widehat{c}^2(t)\right\}$, in other words assuming that a RIR and a nearby RIR have the same decay behavior, it can be shown that the energy envelope of the perturbation is given by

$$\mathrm{E}\left\{A^2 p^2(t)\right\} = \left(1-\gamma^2\right) e^{-2\Delta t}. \tag{4}$$

With respect to (4) we finally end up with the following demand for the energy envelope of the perturbations

$$\mathrm{E}\left\{p^2(t)\right\} = \begin{cases} e^{-2\Delta(t-t_0)}, & t \geq t_0 \\ 0, & t < t_0, \end{cases} \tag{5}$$

where t_0 denotes the time taken by the direct sound.

Proposed Model

Assuming band limited signals that fulfill the sampling theorem, we replace the continuous-time signals and impulse responses with their discrete-time equivalents. We then perform the following steps to generate an instance $p^{(r)}(n)$ of the perturbations:

1. Generate zero-mean Gaussian white noise $b^{(r)}(n)$ with unit variance.
2. Multiply $b^{(r)}(n)$ with $\mathrm{E}\left\{p^2(nT)\right\}^{\frac{1}{2}}$ from (5) to form $p^{(r)}(n)$, where $T = 1/f_s$ and f_s is the sampling frequency.

The final distorted RIR from speaker ℓ to the microphone is then computed by

$$c_\ell^{(r)}(n) = \gamma c_\ell(n) + A p^{(r)}(n)\,, \tag{6}$$

where γ and A are chosen so that the energy of $c_\ell^{(r)}(n)$ equals the energy of $c_\ell(n)$ and that the normalized system misalignment $M_{\mathrm{dB}} = -10\log_{10}\left(\frac{\gamma^2 \mathbf{c}^\top \mathbf{c}}{A^2 \mathbf{p}^{(r)\top}\mathbf{p}^{(r)}}\right)$ reaches a prescribed value.

Results

For the reshaping experiments we measured the impulse responses from $N_s = 4$ loudspeakers to a microphone in a typical office room. The sampling rate was $f_s = 16$ kHz and the RIRs had a length of $L_c = 4000$ taps. The reshaping filters had a length of $L_h = 5000$ taps. For all experiments we used $p_d = 20$, $p_u = 10$ and $p_f = 8$; α was chosen so that we achieved an acceptable frequency response. To investigate the robustness we measured 40 more RIRs in the vicinity of the reference position and computed the corresponding GIRs. To quantify the results, we use the nPRQ and the *spectral flatness measure* (SFM) [1]. The nPRQ captures the average overshot of the time coefficients over the temporal masking limit on a logarithmic scale, while the SFM evaluates the frequency response.

Table 1: Average values for nPRQ and SFM for the non-robust and the proposed reshaping method.

Setup	nPRQ [dB]	SFM
unreshaped	9.93	0.63
non-robust, $\alpha = 40$	10.50	0.64
$R = 26$, $M_{\mathrm{dB}} = -10$ dB, $\alpha = 1$	**4.35**	**0.60**

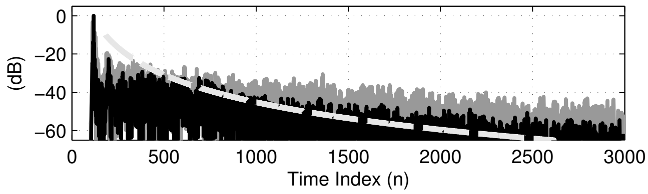

Figure 1: GIR in the case of small spatial mismatch for the non-robust (gray) and the proposed design method (black). The dashed line is the average temporal masking limit.

According to the algorithm from the previous section we generated $R = 26$ perturbed versions of the N_s reference RIRs. We then used the $N_s{\cdot}R$ perturbed, and the N_s reference RIRs as a basis for the equalizer design. Comprehensive results are given in Table 1. A comparison of a GIR in the case of spatial mismatch for the non-robust and the proposed method is depicted in Figure 1.

To investigate the reproducibility of the results, we ran the experiment ten times. The nPRQ was always between 4.35 dB and 4.57 dB while the SFM was always between 0.59 and 0.62.

Conclusions

We presented a simple, yet effective method to increase the spatial robustness of RIR reshaping filters. We could show that the proposed method delivers good results independent of the state of the pseudorandom number generator.

References

[1] 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*, 20(6):1829–1842, Aug. 2012.

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