Parameter Estimation for the NPRQ-Measure with Sub-Sample Precision

Radoslaw Mazur, Fabrice Katzberg, and Alfred Mertins*
Institute for Signal Processing, University of Lübeck, 23562 Lübeck, Germany
Email: {mazur, katzberg, mertins}@isip.uni-luebeck.de

Abstract

Sounds played in closed rooms often suffer from added reverberation. For reducing these effects, the methods of room impulse response equalization may be employed. In order to render the echoes inaudible at a given position, a prefilter is used to modify the played signal. This prefilter is designed according to the properties of the human auditory system, such as temporal masking. Typically, the average temporal masking curve is used to describe the audible reverberation in terms of the nPRQ-Measure and to derive the cost function for the filter design. In this work, we propose to estimate the parameters of this curve with sub-sample precision in order to achieve a better estimation of the perceived reverberation. This higher precision is necessary for comparing the results in case of spatial mismatch.

Introduction

When a sound is played in a closed room, it is reflected multiple times at walls and the objects inside the room. These reflections lead to the sound arriving multiple times at the receiver. The different delays and scalings of these reflections are usually modeled by a convolution with the room impulse response (RIR). The audible result is often a degraded perceived quality for a human listener. In order to combat these distortions, a prefilter can be used. This prefilter is designed in such a way that the audible echoes of the global impulse response (GIR, the convolution of the prefilter and RIR) are reduced or even inaudible [1].

The first approaches [2, 3] designed a prefilter so that the GIR approximates a unit pulse. The employed quadratic cost functions leads to the unwanted part of the GIR being minimized. Unfortunately, even small values lead to clearly audible echoes in the signal. In order to remove these, the properties of the human auditory system need to be considered. The use of the compromise temporal masking curve from [4] allows for a more relaxed formulation. Instead of trying to remove all echoes, it is sufficient to render them inaudible for a human listener. For controlling the late echoes, the quadratic cost function needs to be modified to a p-norm or infinity-norm based criterion as in [5].

When targeting a human listener, who is not able to hold completely still, it is necessary to make the approach spatially robust. Even small movements lead to changes of the RIR and the performance of the prefilter is substantially degraded. In this case, the prefilter may even add reverberation [6]. Different approaches for spatially robust designs have been proposed. Generally, there are two different approaches. The first group uses the multiposition approach, where the prefilters are designed in such a way that multiple points in the listening area are equalized [7, 8]. If the positions fulfill the time-space sampling theorem, the whole area is equalized [8]. In order to achieve satisfactory results, multiple loudspeakers are needed for bigger volumes. This MIMO-approach is usually very demanding, since the RIRs from all loudspeakers to all positions have to be known. The measurement burden can be reduced by employing moving microphones [9, 10]. The second group of algorithms uses regularizers, such as constraining the length of the filters [11], or generating hypothetical RIRs in the vicinity of known positions [12].

In this paper we will revise the window design process based on the temporal masking curve. The discretized method, as in [7], introduces some artifacts and is therefore not suitable for comparing results at different points in the listening area. The proposed method estimates the parameters with sub-sample precision and overcomes this drawback.

RIR reshaping

For the reshaping method from [7] the RIRs $c^{(i)}(n)$ of length L_c from a loudspeaker to the i-th position in space have to be known. With the prefilter $h(n)$ of length L_h, the overall impulse responses are given by

$$g^{(i)}(n) = h(n) * c^{(i)}(n). \tag{1}$$

The reshaping is carried out according to the desired and unwanted parts of the RIR, which are defined using the windows $w_d(n)$ and $w_u(n)$. The desired part is given by

$$g_d^{(i)}(n) = w_d(n)\, g^{(i)}(n) \tag{2}$$

and analogously with index $_u$ for the unwanted part. The design of $w_d(n)$ and $w_u(n)$ will be discussed in more detail in the following section.

The prefilter is obtained by solving the optimization problem given by

$$\mathrm{MIN}_h : f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \tag{3}$$

*This work has been supported by the German Research Foundation under Grant Nos. ME 1170/8-1 and ME 1170/10-1.

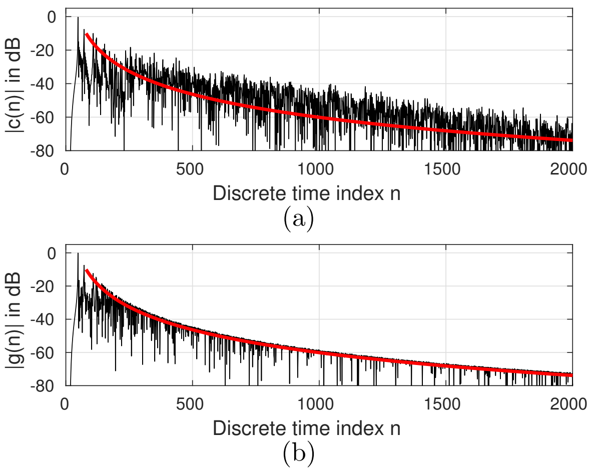


Figure 1: Equalization of an RIR. (a) The original RIR in logarithmic representation. Additionally, temporal masking limit is shown in red. (b) GIR after equalization.

with

$$f_d(\mathbf{h}) = ||\mathbf{g}_d||_{p_d} = \left(\sum_{i=1}^{N_m} \sum_{k=0}^{L_g-1} |g_d^{(i)}(k)|^{p_d} \right)^{\frac{1}{p_d}} \tag{4}$$

and $f_u(\mathbf{h}) = ||\mathbf{g}_u||_{p_u}$, accordingly. N_m is the number of target points in the listening area. The vectors $\mathbf{g}_d$ and $\mathbf{g}_u$ consist of stacked wanted and unwanted parts of the N_m global RIRs. For the solution, a gradient based optimization is used [7].

In contrast to the least-squares methods [2, 3], with the described algorithm and large values for p_d and p_u (typically between 10 and 20), a very smooth shaping without outliers can be achieved. In Figure 1 an example of an RIR and its reshaped version together with the average temporal masking curve are shown.

Window design

In [5, 7] the authors proposed to design the desired und unwanted windows $w_d(n)$ and $w_u(n)$ according to the properties of the human auditory system.

The desired window $w_d(n)$ is supposed to capture the direct-path part of the RIR and allow for integrating up to 4 ms of the subsequent part:

$$w_d(n) = \begin{cases} 1, & N_1 \le n \le N_2 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

with $N_1 = t_0 \cdot f_s$ and $N_2 = (t_0 + T_d) \cdot f_s$. Here, f_s is the sampling frequency, t_0 the time taken by the direct sound from the loudspeaker to the receiver, and $T_d = 4\,\text{ms}$. Optionally, T_d can be also set to $1/f_s$ to create a one-point window, which represents the ideal GIR, the unit pulse. The estimation of t_0 is performed by taking the position and magnitude of the largest coefficient of $c(n)$.

The unwanted window $w_u(n)$ models the reverberant tail of the RIR and is calculated according to the average temporal masking curve

$$w_u(n) = \begin{cases} 10^{\frac{3}{\log(N_0/N2)} \log(\frac{n}{N2})+0.5}, & N_2 \le n < L_g \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

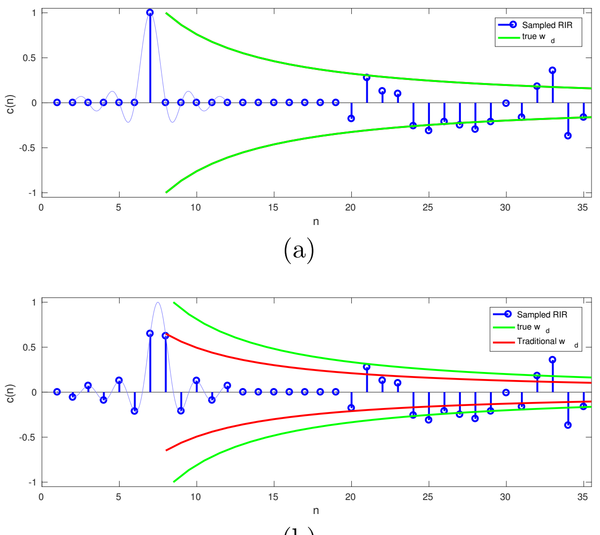


Figure 2: The main peak and the reverberant tail of an RIR. (a) The delay of the main peak is aligned with a sample point. (b) The delay of the main peak is a fraction.

with $N_0 = (0.2\,\text{s} + t_0)$. In [7], there is an implicit assumption of the main peak being scaled to one, or $w_u(n)$ being otherwise scaled accordingly.

Parameter estimation

The design of the windows $w_d(n)$ and $w_u(n)$ as proposed in [7] includes discretization and normalization. The estimated time taken by the sound from the loudspeaker to the receiver is assumed to be a multiple of the sampling interval. Furthermore, the magnitude of the main peak is underestimated in the most cases.

In Figure 2(a) an ideal situation is shown. Here, t_0 is a multiple of the sampling interval $T_0 = \frac{1}{f_s}$, and therefore the direct path is represented by a single coefficient. With the correct delay and magnitude, the unwanted window $w_u(n)$ has the correct scaling and delay.

In Figure 2(b) t_0 is a fraction of T_0. Due to the sampling process, which typically includes a lowpass filtering, the main peak is represented by a sinc-function with $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$ and becomes smeared out. In such a case, taking the largest coefficient leads to an error of up to $\frac{T_0}{2}$ for t_0 and

$$\text{sinc}\,(0.5) \approx 0.6366 \tag{7}$$

for the magnitude. The wrong scaling changes the weighting of the reverberant tail as shown in Figure 2 (b). The reverberations are overestimated.

In order to estimate the correct parameters for the main peak, we propose to use an upsampled version of the RIR [13]. Typically, an upsampling factor of at least 8 allows the magnitude error to become less then 1% as $\text{sinc}(0.5/8) \approx 0.9936$.

For the single-channel case, the scaling factor in (7) is not relevant as it becomes a constant offset for the optimization problem in (3). In case of comparing results for different points in the listening area, this error is distorting the results and hides all the fine details, as shown in the next section.

Experiments

The experiments are based on simulated RIRs [14] of length $L_c = 2000$ in an office-sized room with the dimensions of $5 \times 6 \times 4$ meters. The reverberation time was set to $t_{60} = 300\,\text{ms}$, which leads to clearly audible echoes. The sampling frequency was chosen as $f_s = 8\,\text{kHz}$.

The evaluation is based on the nPRQ-Measure, which quantifies the perceived reverberation and is again based on the temporal masking curve. It calculates the overshoot above the temporal masking curve, with a lower bound of -60dB of the main peak

$$g_{\text{os}}(n) = \max\left(\tfrac{1}{w_u(n)}, -60\text{dB}\right) \tag{8}$$

as

$$\text{nPRQ} = \begin{cases} \frac{1}{\|\boldsymbol{g}_E\|_0} \sum_{n=N_0}^{L_g-1} g_E(n), & \|\boldsymbol{g}_E\|_0 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

with

$$g_E(n) = \begin{cases} 20\log_{10}(|g(n)| w_u(n)), & |g(n)| > g_{\text{os}}(n) \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

When there is no reverberation, i.e., when all coefficients are below the compromise temporal masking curve, the nPRQ is equal to zero. Higher values denote audible reverberation.

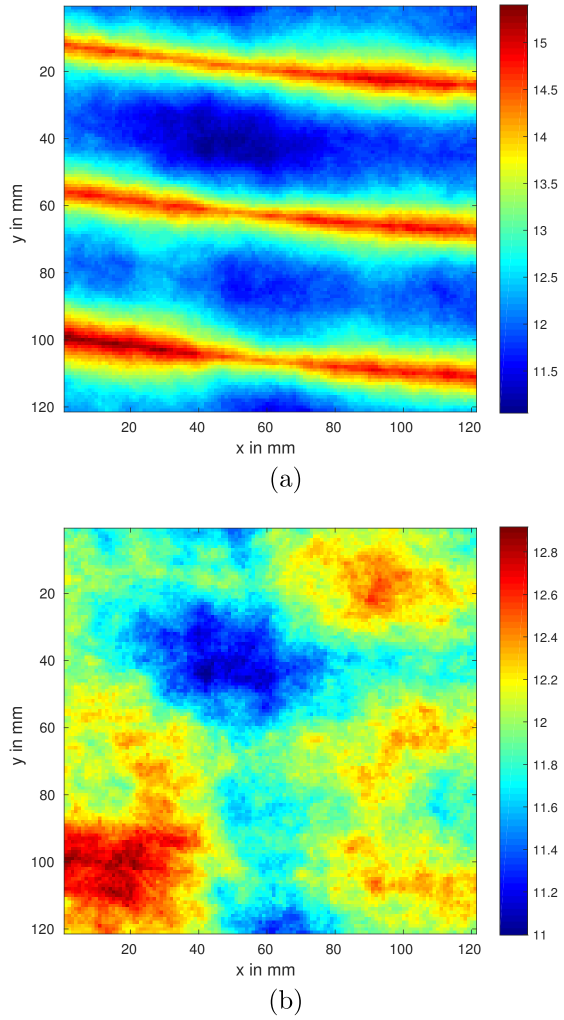
Figure 3 shows the results for the measurement of the nPRQ in a $12 \times 12\,\text{cm}$ sized listening area. In Figure 3(a) the old method is used, which leads to strong artifacts. The periodic repetitions indicate the positions, where the main peak has changed its position by a whole sample. The period is

$$\lambda = \frac{c_0}{f_s} = \frac{340\frac{\text{m}}{\text{s}}}{8000\text{Hz}} = 4.25\,\text{cm} \tag{11}$$

with c_0 the speed of sound. The measurement errors due to the discretization in the process of parameter estimation dominate the results. All fine details are lost and the results at different positions are not comparable.

In Figure 3(b) the new approach with oversampling factor of 16 has been used. Here, the artifacts are gone, and the fine details are becoming visible. Overall, the audible reverberation is only slightly changing in the listening area.

In Figure 4 the results of a reshaping are shown. Here, a single point has been reshaped with an equalizer of length $L_h = 2000$. The color codes the improvement/deterioration of the perceived echoes in terms of ΔnPRQ. Again, using the old formulation the results are dominated by the periodic artifacts as shown in Figure 4(a). In Figure 4(b) the new method has been used. Here, the results are more consistent. At the target point, the reduction of the audible reverberations is the best. With higher distance form the target point, the equalization performance gets worse. The circular shape is consistent with the theoretical results of [6].



(a)

(b)

Figure 3: Reverberation estimation in the listening area in terms of nPRQ-Measure. (a) Old method; (b) New method.

Summary

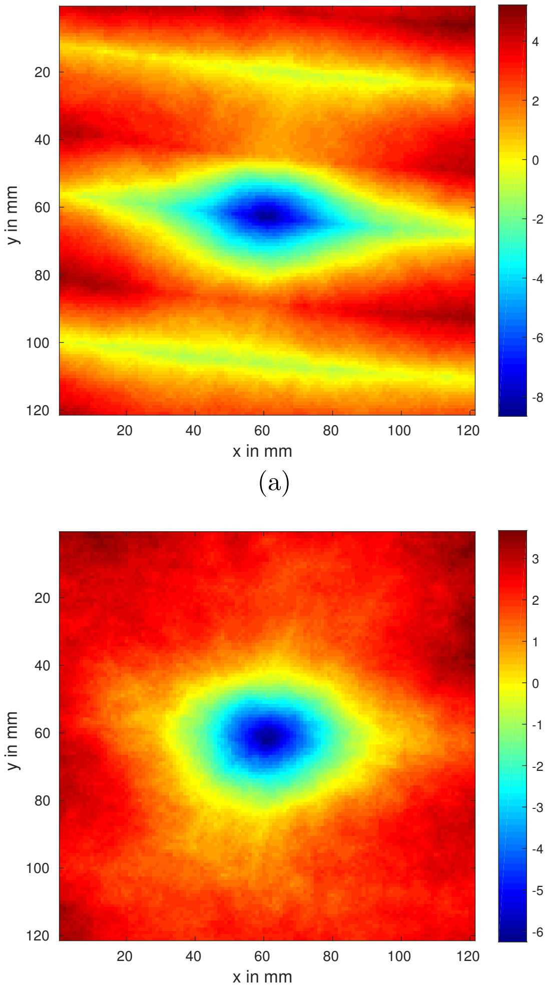
In this work, we have proposed to estimate the parameters of the nPRQ-Measure with sub-sample precision. The new approach allows for a meaningful comparison at different points in the listening area.

References

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, pp. 884–900, Nov. 1994.

[2] 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, pp. 899–907, Nov. 1989.

[3] M. Kallinger and A. Mertins, "Room impulse response shortening by channel shortening concepts," in *Proceedings of the Asilomar Conference on Sig-*



(a)

(b)

Figure 4: Equalization of a single point in the listening area. The color codes the improvement/deterioration of the perceived echoes in terms of ΔnPRQ. (a) Old method; (b) New method.

nals, Systems, and Computers, (Pacific Grove, CA , USA), pp. 898–902, Oct. 2005.

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

[5] 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, pp. 249–259, Feb. 2010.

[6] B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 311–319, May 2000.

[7] 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, pp. 1829–1842, Aug. 2012.

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

[9] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *The Journal of the Acoustical Society of America*, vol. 141, pp. 3220–3235, May 2017.

[10] R. Mazur, F. Katzberg, H. Phan, and A. Mertins, "Room equalization based on measurements with moving microphones," in *Proc. Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, (San Francisco, USA), March 2017.

[11] I. Kodrasi and S. Doclo, "Improving the conditioning of the optimization criterion in acoustic multichannel equalization using shorter reshaping filters," in *EURASIP Journal on Advances in Signal Processing*, vol. 11, Feb 2018.

[12] J. O. Jungmann, R. Mazur, and A. Mertins, "Perturbation of room impulse responses and its application in robust listening room compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Vancouver, BC, Canada), pp. 433–437, May 2013.

[13] P. Vaidyanathan, *Multirate Systems And Filter Banks*. Prentice Hall signal processing series, Pearson, 2004.

[14] E. A. P. Habets, "Room impulse response generator." http://home.tiscali.nl/ehabets/rir_generator.html, Sept. 2010.