SUBBAND PROCESSING FOR SPATIALLY ROBUST ROOM IMPULSE RESPONSE RESHAPING

Radoslaw Mazur, Fabrice Katzberg, Martina Böhme, and Alfred Mertins

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

ABSTRACT

The perceived quality of a sound played in a closed room is often degraded by the added reverberation. In order to combat these effects, the methods of room impulse response equalization may be used. Typically, properties of the human auditory system, such as temporal masking, are used for a better control of the late echoes. This is accomplished by designing prefilters which modify the played signals and render the echoes inaudible at a given position. Small spatial mismatch usually results in reduced performance and bigger offsets even yield added reverberation. This problem can be tackled by simultaneously reshaping multiple positions in the listening area, but it requires a huge amount of impulse responses. In this work we propose to mitigate this burden by equalizing only certain frequency components of the room impulse responses. When equalizing only the low frequencies, the mismatch and the degradation can be reduced. This allows for less measurements in the listening area, where a human listener is able to move freely.

***Index Terms*—** Room Impulse Response, Reshaping, Equalization, Spatial Mismatch, Subband processing

1. INTRODUCTION

Sounds played in closed rooms are reflected multiple times on the walls and other objects. Due to these reflections, a listener receives the signal multiple times with different delays and scaling. This process can be modeled by a convolution with the room impulse response (RIR). Usually, it degrades the perceived quality for a human listener. These distortions can be reduced by applying a prefilter that results in a global impulse response (GIR, the convolution of the RIR and prefilter) which has no audible echoes [1].

Simple approaches try to design a prefilter which results in a GIR being a unit pulse [2, 3]. These methods are minimizing the unwanted parts of the GIR, but are not successful since there remain clearly audible late echos in the signal. In order to combat these, the properties of the human auditory system can be utilized. Instead of trying to remove all the echos, they are rendered inaudible for a human listener. This relaxed requirement has been used in [4]. Here, the authors used the compromise temporal masking curve for describing the perceived reverberation [5]. In order to control the late echos, a p-norm based criterion has been used.

A typical human listener is not able to keep still. Small changes of the position result in changed RIRs and the performance of the system is significantly degraded. For bigger displacements this may even result in added reverberation [6]. For spatially robust designs, different approaches have been proposed. In general, these can be grouped into two classes. The first class of algorithms uses the multi-position method. Here, the prefilters are designed in such a way that multiple points in the listening area are equalized [7, 8]. With enough points that fulfill the time-space sampling theorem [8], the whole listening area becomes equalized. For bigger volumes, multiple loudspeaker are necessary. Overall, this MIMO approach is very demanding, as it requires a huge amount of measurements – from all loudspeakers to all positions on the grid. Recently, this amount could be reduced with the use of moving microphones [9]. This approach has been validated for RIR reshaping in [10]. The second group uses single RIRs, models additional errors in the optimization and adds regularizers [7]. In [11], the authors were able to extend the equalized volume by generating multiple hypothetical RIRs. This extended volume comes at the cost of reduced performance at the target point. In [12] regularization was achieved by using short filters.

In [13] the authors proposed to equalize only the low frequencies of an RIR in order to remove the resonant frequencies of a room. In this work, we modify this approach and equalize a listening area that is sampled not dense enough for the methods from [7, 8]. The spatial arrangement of the microphones translates to a maximum frequency which can be equalized. Using this cut-off frequency a subband decomposition is performed and only the eligible part is processed.

This paper is organized as follows. In the next section we will review the multi-position p-norm based equalization

This work has been supported by the German Research Foundation under Grants No. ME 1170/8-1 and ME 1170/10-1.

978-1-5386-8151-0/18/$31.00

method. In Section 3, the subband method will be combined with the mult-position approach. In Section 4, experiments for the subsampled case will be shown. Finally, some conclusions will be given in the last section.

2. RIR RESHAPING

The multichannel RIR reshaping method uses N_m microphones in the listening area and N_s speakers for playback. The RIRs $c_{ij}(n)$ from the j-th loudspeaker to the i-th microphone are modeled with length L_c. With $h_k(n)$ being the k-th prefilter of length L_h, the global impulse response (GIR)

$$g_i(n) = \sum_{k=1}^{N_s} h_k(n) * c_{ik}(n) \tag{1}$$

at the i-th position has the length of $L_g = L_c + L_h - 1$. As proposed in [7], two windows, $w_d(n)$ and $w_u(n)$, are used for defining the desired and unwanted parts of the GIR as $g_{d,i}(n) = w_d(n)g_i(n)$ and $g_{u,i}(n) = w_u(n)g_i(n)$. In this work, we define these weighting windows as in [4, 5] in order to capture the compromise temporal masking limit of the human auditory system.

The prefilters $h_k(n)$ can be estimated by minimizing

$$\mathrm{MIN}_h : f(\mathbf{h}) = \log\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \tag{2}$$

with

$$f_d(\mathbf{h}) = ||\mathbf{g}_d||_{p_d} = \left(\sum_{i=1}^{N_m}\sum_{k=0}^{L_g-1} |g_{d,i}(k)|^{p_d}\right)^{\frac{1}{p_d}} \tag{3}$$

and $f_u(\mathbf{h}) = ||\mathbf{g}_u||_{p_u}$, accordingly. The vectors $\mathbf{g}_d$ and $\mathbf{g}_u$ consists of stacked wanted and unwanted parts of the N_m global RIRs. As there does not exist a closed form solution for (2), the optimization is carried out by a gradient descent method [7].

A smooth shaping with no outliers can be achieved by choosing sufficiently high values for p_d and p_u. Typical values are between 10 and 20, so that the p-norms sufficiently approximate the infinity norm, which would guarantee an outlier-free reshaping but is more costly to optimize than the p-norm.

3. SUBBAND BASED RESHAPING

In the case of spatial mismatch between the listener and the reference positions, the performance of the solution given by (2) degrades significantly. In [6] the average frequency-dependent error in the single channel case has been estimated as

$$F(\omega) = 2 - 2\frac{\sin(\omega\Delta/v)}{\omega\Delta/v} \tag{4}$$

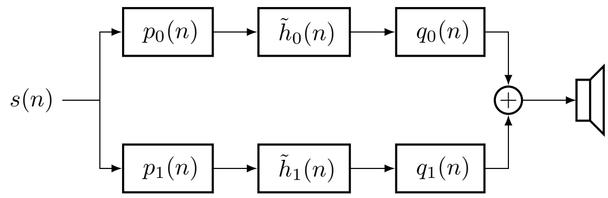

Fig. 1. The subband model of the prefilter using a two channel nondecimating filter bank.

with Δ being the displacement, v the speed of sound, and $\omega = 2\pi f$ the angular frequency. Equation (4) shows, that for higher frequencies and displacements, the result may be even worse than no equalization at all. An example will be given in the experiments section.

In order to fully equalize a volume, a dense sampling of the listening area on a regular grid is required. Let $D = \frac{1}{f_{ss}}$ denote the distance between the microphones, f_{ss} the sampling frequency in space, and f_{ts} the sampling frequency in time. As shown in [8], sampling with

$$f_{ss} \geq \frac{f_{ts}}{v} \tag{5}$$

leads all RIRs inside the listening area to have the same decay behavior as at the measurement points. Unfortunately, this approach requires a very high number of necessary measurements which usually cannot be obtained even for medium-sized volumes.

In case of spatial sampling not fulfilling (5), we propose to equalize only the lower-frequency content of the RIR that satisfies the sampling theorem. Additionally, this approach also combats the resonant frequencies as in [13]. This can be achieved, for example, by replacing the prefilters h_k by filter banks shown in Fig. 1. This two-channel design is motivated by the highly nonlinear cost function in (2). In order to achieve a reshaping in single subbands, the analysis filters are required to conserve the general shape of the RIR when being applied. Furthermore, since the subbands are equalized independently, aliasing should be minimized. This can be achieved by not using subsampling and by reducing the number of subbands to a minimum – in this case to two. The complexity reduction by subsampling is not an issue and we are interested in only reshaping the lower band of the RIR up to the cutoff frequency

$$f_c = \frac{v}{2Df_{ts}}. \tag{6}$$

The non-decimating subband decomposition as shown in Fig. 1 puts only little constraints on the analysis and synthesis filters [14]. For a nearly perfect reconstruction the analysis filters

$$p_0(n) = \mathrm{sinc}(n\, f_c)w(n) \tag{7}$$
$$p_1(n) = \delta(n) - \mathrm{sinc}(n\, f_c)w(n) \tag{8}$$

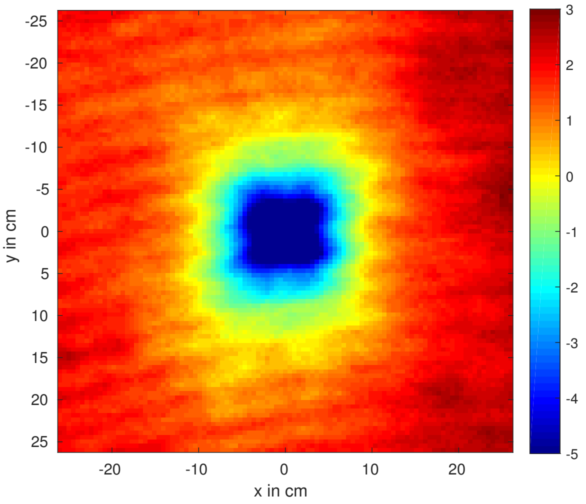

Fig. 2. Equalization using fullband method with three loudspeakers and four microphones. The color codes the improvement/deterioration of the perceived echoes in terms of ΔnPRQ. Blue and green mean improvement, yellow indicates no change and red colors show added reverberation.

with $\delta(n)$ being the unit pulse and $w(n)$ a smooth window (i. e. Hamming window), can be used. The reconstruction filters then reduce to $q_0(n) = q_1(n) = \delta(n)$. With only the lower band being equalized, the global RIR becomes

$$g_i(n) = \sum_{k=1}^{N_s} (\tilde{h}_{k0}(n) * p_0(n) + p_1(n)) * c_{ik}(n). \quad (9)$$

The estimation of $\tilde{h}_{k0}(n)$ is carried out using (2) on the filtered versions of c_{ik}.

4. EXPERIMENTS

The experiments are based on simulated RIRs [15] 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 $t_{60} = 400\,\text{ms}$, which leads to clearly audible echoes. The sampling frequency was chosen as $f_{ts} = 8\,\text{kHz}$, which corresponds to the maximum distance of the microphones of around $D = 4\,\text{cm}$ for the dense sampling which is needed for the fullband reshaping. In Fig. 2 the results for a three-loudspeakers and four-microphones setup are shown. The color codes the improvement in terms of nPRQ, compared to the nonequalized case. The nPRQ from [7] is a measure for quantifying the perceived reverberation. It calculates the overshot above the temporal masking curve, with a lower bound of -60dB of the main peak

$$g_{\text{os}}(n) = \max\left(\frac{1}{w_u(n)}, -60\text{dB}\right) \quad (10)$$

Table 1. Comparison of the performance in the target area for a setup with three loudspeakers and four microphones located on a square with size $S \times S$ in dependence of the microphone distance D. For reference, 3×9 and 3×16 setups are shown in the first part.

S	D	f_c/f_s	$\mu_{\Delta\text{nPRQ}}$	$\sigma_{\Delta\text{nPRQ}}$	$\max_{\Delta\text{nPRQ}}$
4 cm	4 cm	1	-5.99	1.30	-3.11
8 cm	4 cm	1	-4.54	1.57	-1.35
16 cm	4 cm	1	-2.83	1.87	-1.37
8 cm	8 cm	1	-2.79	1.35	-0.30
8 cm	8 cm	0.5	-3.19	0.52	-1.78
16 cm	16 cm	1	-0.19	1.20	1.96
16 cm	16 cm	0.5	-2.08	0.72	0.01
16 cm	16 cm	0.25	-2.24	0.32	-1.00

as

$$\text{nPRQ} = \begin{cases} \frac{1}{\|\boldsymbol{g}_E\|_0} \sum_{n=N_0}^{L_g-1} g_E(n), & \|\boldsymbol{g}_E\|_0 > 0 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

with

$$g_E(n) = \begin{cases} 20\log_{10}(|g(n)|w_u(n)), & |g(n)| > g_{\text{os}}(n) \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

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.

The results In Fig. 2 show an improvement of $\Delta\text{nPRQ} = -5.99$ in the target area as indicated in the first line of Table 1. This is a very successful equalization, but it comes at the cost of an heavily deterioration outside of the target area as represented by the red colors.

In Fig. 3 the results for the subsampled case are shown. Here, the distance of the microphones is $D = 8\,\text{cm}$ which results in reduction of factor four (two in each dimension) in terms of necessary measurements. The fullband method, as shown in Fig. 3 (a), equalizes mainly the measurement points while the area in between is only poorly processed. The results for the new subband method in Fig. 3 (b) indicate the whole area being evenly equalized. All used measures, the average ($\mu_{\Delta\text{nPRQ}}$), the smoothness ($\sigma_{\Delta\text{nPRQ}}$), and worst case ($\max_{\Delta\text{nPRQ}}$) are improved, as shown by the second section of of Table 1.

In Fig. 4 the distance of the microphones has been increased to $D = 16\,\text{cm}$. The fullband method is not able to equalize the target area and most points even exhibit an deterioration. The subband method is successful. Again, all measures are improved, as shown in the third part of Table 1. The fully sampled case (line three in Table 1), shows only a slight

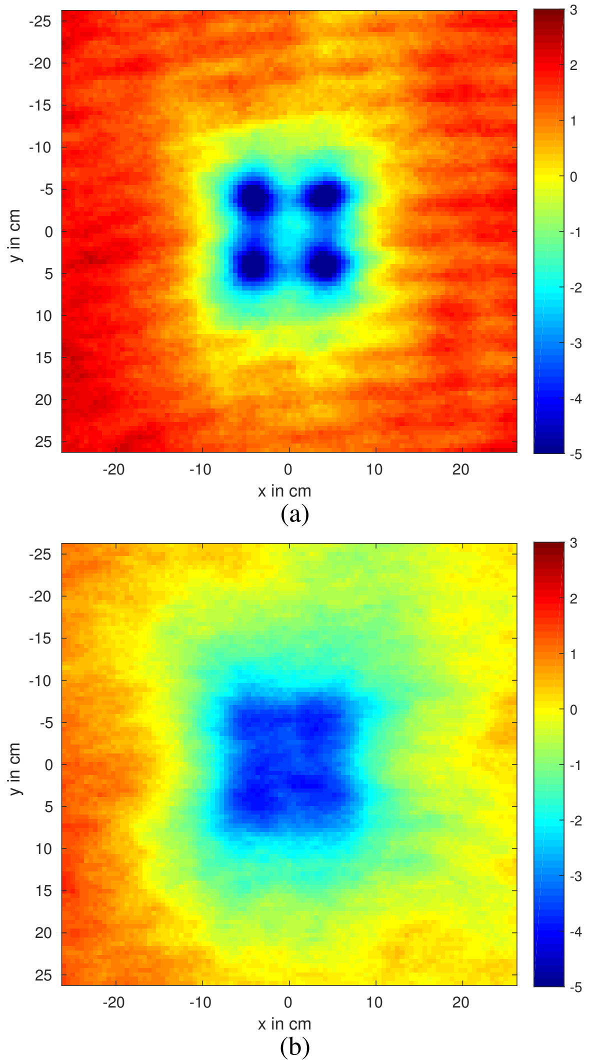

Fig. 3. Equalization in the case of $D = 8cm$. The fullband method in (a) results in the target area being only partially equalized. The new subband method in (b) results in a more even and better equalization.

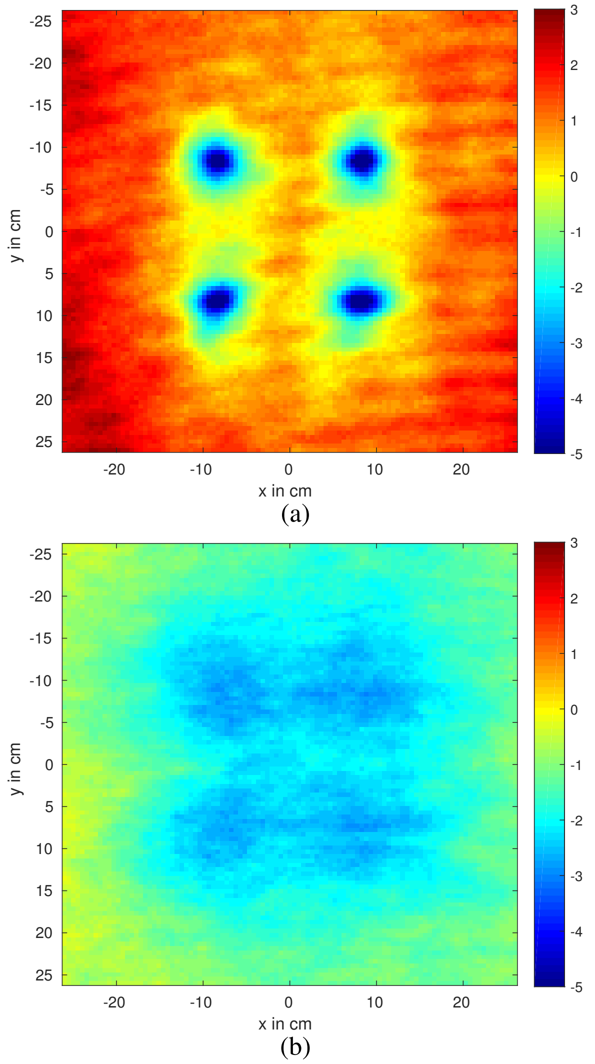

Fig. 4. Equalization in the case of $D = 16cm$. The fullband method in (a) results in only single points being equalized. The new subband method in (b) is able to equalize the whole target area.

improvement, but at the cost of 16 times more measurement effort (factor four in each dimension).

Overall, the new subband method is able to equalize the target area in a smooth way. By mainly equalizing the low frequencies, this method is also able to remove the resonant frequencies of a room.

5. CONCLUSION

In this work we proposed a subband approach for room impulse response reshaping. The traditional methods require a huge amount of measurement which are usually not feasible for a real world situation. By equalizing only certain frequency components of the RIRs, a less dense sampling of the listening area is possible. With significantly less measurement effort, a smooth equalization of the listening area has been achieved.

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 *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA , USA, Oct. 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, Feb. 2010.

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

[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 Transactions on 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 *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, no. 5, pp. 3220–3235, May 2017.

[10] Radoslaw Mazur, Fabrice Katzberg, Huy Phan, and Alfred 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] Jan Ole Jungmann, Radoslaw Mazur, and Alfred 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, May 2013, pp. 433–437.

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

[13] Mihailo Kolundžija, Christof Faller, and Martin Vetterli, "Multi-channel low frequency room equalization using perceptually motivated constrained optimization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, Mar. 2012, pp. 533–536.

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

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