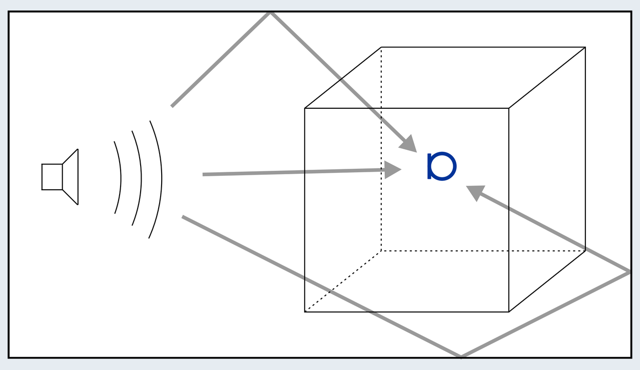
UNIVERSITÄT ZU LÜBECK
INSTITUTE FOR SIGNAL PROCESSING

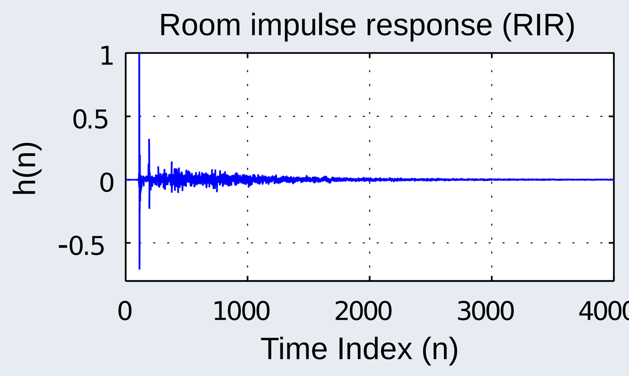
Robust Room Equalization Using Sparse Sound-Field Reconstruction

Radoslaw Mazur, Fabrice Katzberg, and Alfred Mertins

1. Introduction

- Sound transmission within reverberant environments:





Result: Degraded quality in the listening area.

- For enhancing the perceived quality, room impulse response (RIR) equalization can be used.
- For equalization of a volume, knowledge of RIRs in the target area is needed.
- RIRs are very sensitive to spatial variations.
- Problem: Usually, it is not feasible to measure RIRs at all necessary position.
- This work shows that equalization performance improves when using sparse RIR estimates from (sub)sampled sound-field data.

2. Conventional RIR Interpolation

- For equidistant positions $\boldsymbol{r_g} \in \mathcal{G}$ with

$$\mathcal{G} = \left\{ \boldsymbol{r_g} \,|\, \boldsymbol{r_g} = \boldsymbol{r}_0 + \left[g_x\Delta, g_y\Delta, g_z\Delta\right]^T \right\}, \quad (1)$$

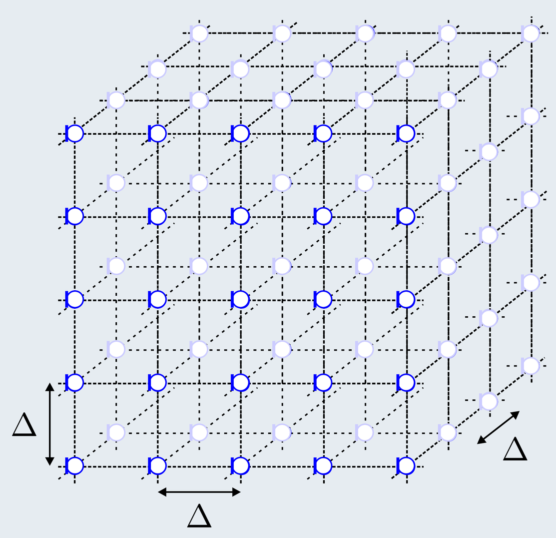
RIR estimates read

$$\hat{h}_{\boldsymbol{r}}(n) \approx \sum_{\boldsymbol{g}\in\mathbb{N}} \varphi_{\boldsymbol{r}}(\boldsymbol{g})\, h(\boldsymbol{g}, n), \quad (2)$$

where $\boldsymbol{g} = \left[g_x, g_y, g_z\right]^T$ spans a uniform grid and the kernel $\varphi_{\boldsymbol{r}}(\boldsymbol{g})$ approximates the sinc function.

- The Nyquist-Shannon sampling theorem requires

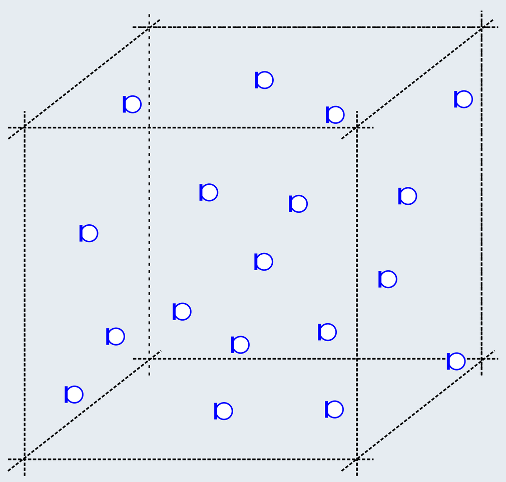
$$\Delta < \frac{c_0}{2 f_{\max}}. \quad (3)$$



3. RIR Estimation Based on Compressed Sensing

- The idea is to use RIR estimates at wanted positions $\tilde{\boldsymbol{r}}_d$ ($d \in \{1, \dots, D\}$) based on spatially subsampled RIRs measured at arbitrary points $\boldsymbol{r}_m$ ($m \in \{1, \dots, M\}$) $\Rightarrow M < D$.
- For $\tilde{\boldsymbol{r}}_d \in \mathcal{G}$, the system of linear equations

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{d} + \boldsymbol{\eta} \quad (4)$$



sets up the reverse interpolation problem to (2).

- $\boldsymbol{m} \in \mathbb{R}^{LM}$ contains the concatenation of M measured RIRs, each of length L.
- $\boldsymbol{d} \in \mathbb{R}^{LD}$ contains the concatenation of D wanted RIRs at positions $\tilde{\boldsymbol{r}}_d \in \mathcal{G}$.
- $\boldsymbol{\eta} \in \mathbb{R}^{LM}$ comprises the measurement noise and the interpolation error.
- $\boldsymbol{A} \in \mathbb{R}^{LM\times LD}$ performs 3D interpolation in line with (2).

I. Estimation of sparse RIRs at virtual-grid points

- Since $M < D$, (4) provides an infinite number of least-squares solutions for $\boldsymbol{d}$.
- Sound-field sparsity in frequency domain [1] allows for using CS methods [2]:

$$\underset{\boldsymbol{c}\in\mathbb{C}^{LD}}{\arg\min} \left\| \boldsymbol{m} - \boldsymbol{A}\boldsymbol{\Psi}^H\boldsymbol{c} \right\|_{\ell_2}^2 \quad \text{s. t.} \quad \|\boldsymbol{c}\|_{\ell_0} \leq K. \quad (5)$$

- $\boldsymbol{c} = \boldsymbol{\Psi}\boldsymbol{d}$ contains a K-sparse frequency representation of grid RIRs $h(\boldsymbol{g}, n)$.
- $\boldsymbol{\Psi} \in \mathbb{C}^{LD\times LD}$ contains a unitary basis for some frequency domain.
- CS based solution $\hat{\boldsymbol{d}} = \boldsymbol{\Psi}^H\hat{\boldsymbol{c}}$ provides estimates of virtual-grid RIRs $\hat{h}(\boldsymbol{g}, n)$.

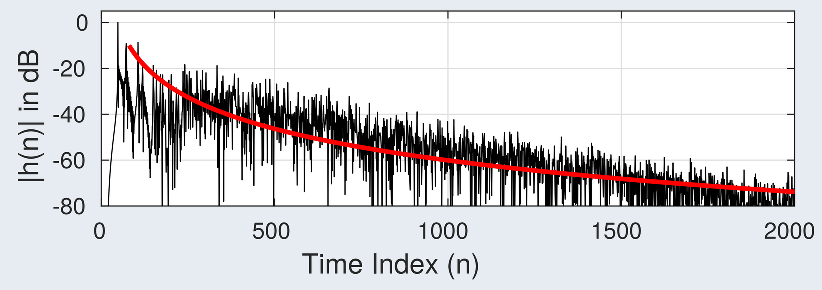
II. Reconstruction by using sparse RIR estimates

- For modeling (4) with virtual-grid points satisfying the Nyquist-Shannon sampling theorem, RIR estimates for any position $\boldsymbol{r} \in \mathbb{R}^3$ are available using

$$\hat{h}_{\boldsymbol{r}}(n) \approx \sum_{\boldsymbol{g}\in\mathbb{N}} \varphi_{\boldsymbol{r}}(\boldsymbol{g})\, \hat{h}(\boldsymbol{g}, n). \quad (6)$$

4. Reshaping of RIRs

- The average temporal masking curve (TMC) is used to describe the audible echoes.



- The reshaping method from [3] uses known RIRs $h_i(n)$ of length L_h from a loudspeaker to the i-th position in space.
- With a prefilter $a(n)$ of length L_a, the overall impulse responses are given by

$$g_i(n) = a(n) * h_i(n). \quad (7)$$

- RIR reshaping is carried out according to the desired and unwanted parts:

$$g_{d,i}(n) = w_d(n)\, g_i(n), \qquad g_{u,i}(n) = w_u(n)\, g_i(n). \quad (8)$$

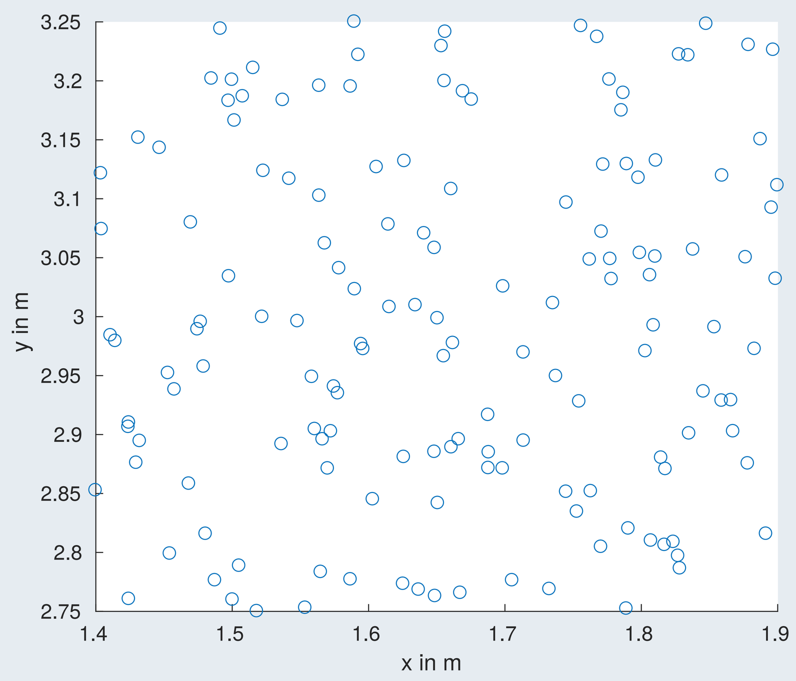
- The prefilter is obtained by solving the optimization problem given by

$$\underset{\boldsymbol{a}\in\mathbb{R}^{L_a}}{\arg\min} \log\left(\frac{f_u(\mathbf{a})}{f_d(\mathbf{a})}\right) \quad \text{with} \quad f_d(\mathbf{a}) = \|\mathbf{g}_d\|_{p_d} = \left(\sum_{i=1}^{N}\sum_{n=0}^{L_g-1} |g_{d,i}(k)|^{p_d}\right)^{\frac{1}{p_d}}, \quad (9)$$

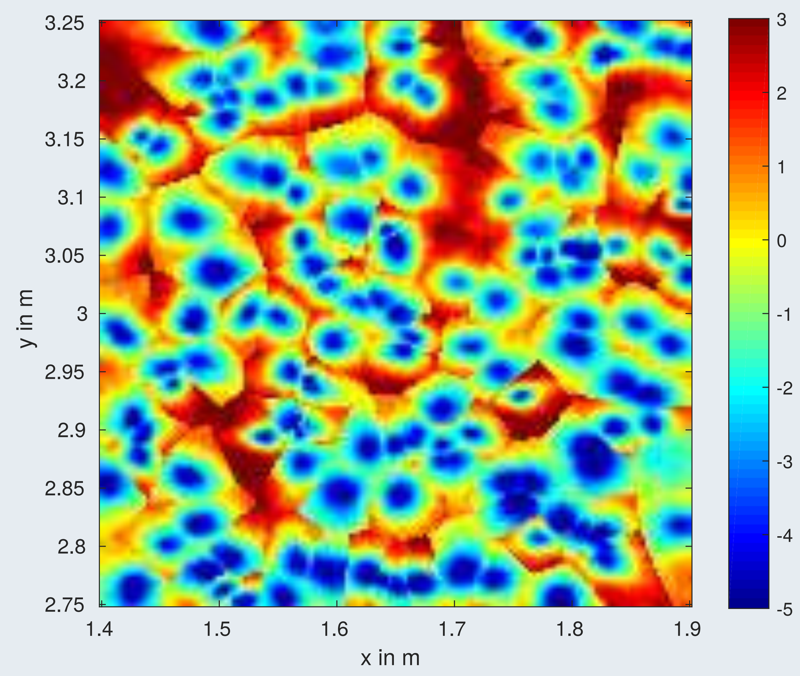
and $f_u(\mathbf{a}) = \|\mathbf{g}_u\|_{p_u}$, accordingly. The vectors $\mathbf{g}_d$ and $\mathbf{g}_u$ consist of stacked desired and unwanted parts of the N global RIRs.

5. Experiments and Results

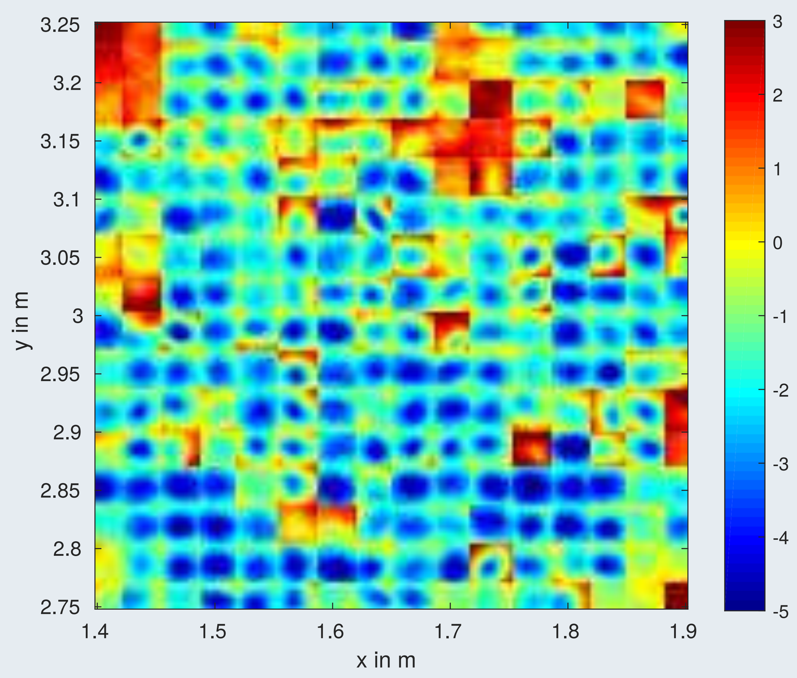
- Inside an office-sized room, RIRs at $M = 150$ random positions $\boldsymbol{r}_m \in \mathbb{R}^2$ were acquired on a plane of size 0.5×0.5 m.
- For every measured RIR $h_i(n)$, an equalizer $a_i(n)$ was estimated using (9).
- The target plane was equalized at any point using the prefilter refering to the closest sampled position.
- Equalization was evaluated by improvement/deterioration of the perceived echoes in terms of nPRQ [3] (measures the overshot above the TMC).



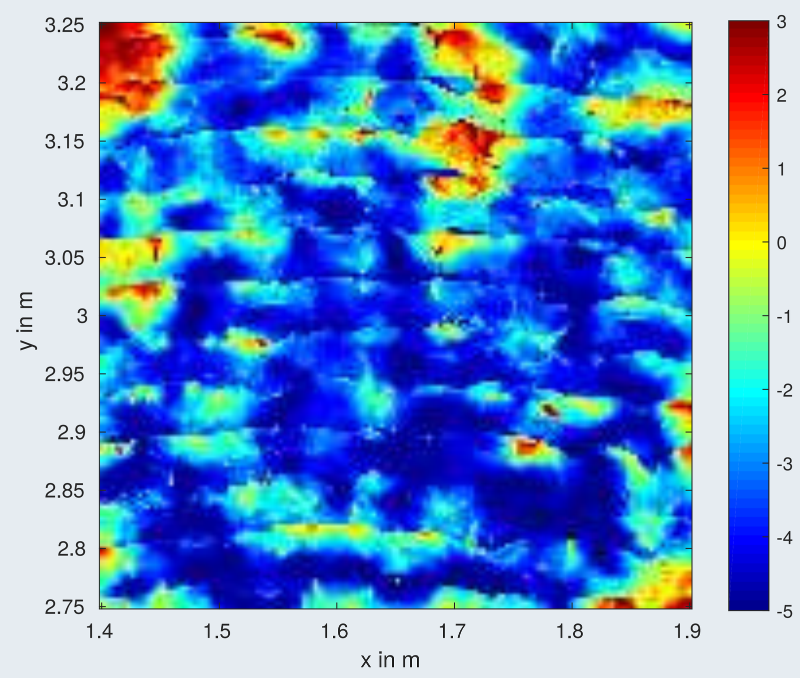
Positions of measured RIRs



ΔnPRQ for nearest-neighbor reshaping



ΔnPRQ based on estimates of sparse RIRs on 18×18 grid (I.)



ΔnPRQ based on reconstruction through sparse RIR estimates (II.)

6. Conclusions

- The traditional approach of RIR reshaping based on the nearest measuring point yields poor performance in case of a large spatial mismatch.
- Sparse estimates of non-available sound-field data improve the equalization.
- Estimation errors of sparse recovery have less impact than errors due to spatial mismatch.

References

[1] T. Ajdler, L. Sbaiz, and M. Vetterli. The plenacoustic function and its sampling. *IEEE Trans. on Signal Processing*, 54(10):3790–3804, Oct. 2006.
[2] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.
[3] 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. on Signal Processing*, 20(6):1829–1842, Aug. 2012.

ICASSP 2019, Brighton, UK