ROBUST ROOM EQUALIZATION USING SPARSE SOUND-FIELD RECONSTRUCTION

Radoslaw Mazur, Fabrice Katzberg, 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. There, a prefilter is used to modify the played signal, which renders the echoes inaudible for a given position. When targeting a human listener who is always performing small movement, the method fails due to the spatial mismatch. In such a case, a whole volume needs to be equalized, which requires a huge amount of measurements. In this work, we propose to reduce this burden by measuring only a random subset of the required room impulse responses and reconstruct a grid employing sparse methods. By further interpolation an equalization of the whole target area is achieved.

Index Terms— Room Impulse Response, Reshaping, Equalization, Sparse Reconstruction

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 scalings. 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].

The simplest approach is to design the prefilter in such a way that the GIR becomes a unit pulse or, more generally, a bandpass [2, 3]. When optimizing using a quadratic criterion, the unwanted parts of the GIR are greatly reduced. Unfortunately, with this approach, the signal still contains audible echoes. However, by exploiting the properties of the human auditory system, another approach is more feasible. The idea is not to remove the echoes completely, but rather render them inaudible for a human listener. Typically, the average temporal masking curve [4] is used to describe the perceived reverberation. This relaxed approach has been very successful [5]. Additionally, the authors proposed to use a p-norm based criterion instead of the quadratic term as in [3]. This allows for a better control of the late echoes.

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 (e. g. on a dense grid) [8], the whole listening area becomes equalized to the same extent. For bigger volumes, the use of multiple loudspeakers may become 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. The second group of algorithms uses single RIRs. They model additional errors in the optimization and add regularizers [7]. In [9], 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 [10] regularization was achieved by using short filters.

In this work, we propose a processing strategy for measured sound-field data used for room equalization, in order to improve performance. Given a set of RIRs sampled at arbitrary positions inside the listening area, estimates of RIRs on a dense spatial grid will be obtained by solving a system of linear equations. Unlike the measurements, the estimates are supposed to satisfy the Nyquist-Shannon sampling theorem in each dimension. Thus, in general, the linear system is underdetermined. Nevertheless, as recently shown in [11], robust recovery of the grid RIRs is possible employing the methods of compressed sensing (CS) [12, 13] and solving the system with sparsity constraint, i.e., permitting only a small number of variables to be non-zero. By using the recovered grid and applying further interpolation, equalizers for any target point inside the listening area may be derived. Thus, if the listener position is tracked, appropriate equalizers can be provided for every position.

In the experimental part of this paper, we compare the performance of the equalization algorithm in [7] fed with (1) the measured RIRs, (2) the CS based estimates on the grid, and (3) the CS based estimates reconstructed over the entire listening area. We will see that equalization results may improve significantly when relying on the sparse RIRs reconstructed directly at the target points, even for points being far away from measurement positions.

This paper is organized as follows. In the next section the idea of sparse sampling and reconstruction of RIRs will be given. In Section 3, the p-norm based equalization will be reviewed. In Section 4, experiments for the new approach will be shown. Finally, some conclusions will be given in the last section.

2. MEASURED RIR DATA

The straightforward way is to use measured RIRs directly for room equalization: in order to equalize at target points, sampled RIRs at

This work has been supported by the German Research Foundation under Grant Nos. ME 1170/8-1 and ME 1170/10-2.

the nearest measurement positions could be used as input for the equalizer. Nevertheless, in the following we propose strategies for processing measured RIR data before performing the equalization step.

2.1. Sparse Recovery of Grid RIRs

The idea of this strategy is to obtain RIR estimates at virtual positions $\tilde{\mathbf{r}}_d$ ($d \in \{1, \ldots, D\}$) by using the measured RIRs at arbitrary points $\mathbf{r}_m$ ($m \in \{1, \ldots, M\}$). Unlike the measured points $\mathbf{r}_m$, the virtual positions $\tilde{\mathbf{r}}_d$ are supposed to satisfy the Nyquist-Shannon sampling theorem, thus, in general, the problem is underdetermined with $D > M$ and must be solved using CS.

There are different approaches to model the ensemble of points $\tilde{\mathbf{r}}_d$ in order to allow for appropriate interpolation between them, such as, e.g., uniform, quincunx [14], and spherical patterns [15]. Due to practical reasons (e.g., interpolation coefficients and sparsifying transform may become separable), we model a uniform setup of $\tilde{\mathbf{r}}_d$ (cf. [16, 17])

For virtual points forming the equidistant grid

$$G = \left\{ \mathbf{r_g} \,|\, \mathbf{r_g} = \mathbf{r}_0 + [g_x\Delta, g_y\Delta, g_z\Delta]^T \right\}, \tag{1}$$

with $\mathbf{r}_0$ being the grid origin and $\mathbf{g} = [g_x, g_y, g_z]^T \in \mathbb{Z}^3$ representing the discrete grid variables, aliasing-free reconstruction requires

$$\Delta < \frac{c_0}{2f_c}, \tag{2}$$

where c_0 is the speed of sound and f_c is the temporal cutoff frequency [14].

The recovery is based on solving a linear system of equations that is set up by projecting spatio-temporal RIRs at $\tilde{\mathbf{r}}_d \in G$ onto the measurement space spanned by RIRs acquired at $\mathbf{r}_m \in \mathbb{R}$. Let us define the measurement vector

$$\mathbf{m} = \left[\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots, \mathbf{h}_M^T\right]^T \in \mathbb{R}^{ML} \tag{3}$$

as concatenation of the measured RIRs of length L and the target vector

$$\mathbf{d} = \left[\tilde{\mathbf{h}}_1^T, \tilde{\mathbf{h}}_2^T, \ldots, \tilde{\mathbf{h}}_D^T\right]^T \in \mathbb{R}^{DL} \tag{4}$$

as concatenation of the virtual-grid RIRs satisfying the spatial sampling theorem (2). We can then formulate the overall recovery problem in terms of the linear system of equations

$$\mathbf{m} = \mathbf{A}\mathbf{d} + \boldsymbol{\eta}, \tag{5}$$

where $\boldsymbol{\eta} \in \mathbb{R}^{ML}$ is a perturbation vector and $\mathbf{A} \in \mathbb{R}^{ML \times DL}$ is an interpolation matrix.

Since $D > M$, the linear system (5) provides an infinite number of least-squares solutions for $\mathbf{d}$. Nevertheless, as recently shown in [11], the principle of CS allows for finding a stable and robust solution for the grid RIRs, also in the underdetermined case. The spectrum of sound fields is ideally confined to a hypercone along the temporal frequency axis [14], thus, we can represent the equidistantly computed RIRs collected in $\mathbf{d}$ by a sparse coefficient vector

$$\mathbf{c} = \boldsymbol{\Psi}\mathbf{d},$$

where the unitary matrix $\boldsymbol{\Psi} \in \mathbb{C}^{DL \times DL}$ performs the 4D discrete Fourier transform. The vector $\mathbf{c} \in \mathbb{C}^{LD}$ is assumed to be K-sparse, i.e., at most K coefficients are non-zero. Accordingly, the least-squares problem may be regularized as

$$\underset{\mathbf{c} \in \mathbb{C}^U}{\arg\min} \; \|\mathbf{x} - \mathbf{B}\mathbf{c}\|_{\ell_2}^2 \quad \text{s.t.} \quad \|\mathbf{c}\|_{\ell_0} \le K, \tag{6}$$

with the CS matrix $\mathbf{B} = \mathbf{A}\boldsymbol{\Psi}^H$ and the pseudonorm $\|\mathbf{c}\|_{\ell_0}$ counting the number of non-zero elements in $\mathbf{c}$. The problem (6) is NP-hard [18] and is solved in practice by using greedy algorithms [19, 20, 21] or applying convex optimization tools to corresponding ℓ_1-minimization problems [22, 23, 24, 25].

By using the CS based solution, $\mathbf{d} = \boldsymbol{\Psi}^H\mathbf{c}$ provides recovered RIRs on the modeled grid (1). These RIRs may be directly used for constructing equalizers.

2.2. Reconstruction by Using Sparse Grid RIRs

The virtual-grid RIRs can be used for reconstructing RIRs at any position inside the target area. Since the grid is designed with respect to the spatial Nyquist-Shannon requirement (2), conventional interpolation allows for accurate spatial reconstruction. In practice, RIR estimates for any position $\mathbf{r} \in \mathbb{R}^3$ are available using

$$h(\mathbf{r}, n) \approx \sum_{\mathbf{g}} \varphi_{\mathbf{r}}(\mathbf{g}) h(\mathbf{g}, n), \tag{7}$$

where $n \in \mathbb{N}$ is the discrete time variable and $\varphi_{\mathbf{r}}(\mathbf{g})$ denotes a realizable interpolation kernel that approximates the sinc function (cf. [17]).

3. RIR RESHAPING

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

$$b^{(i)}(n) = a(n) * h^{(i)}(n). \tag{8}$$

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

$$g_d^{(i)}(n) = b^{(i)}(n) w_d(n) \tag{9}$$

and analogously for the unwanted part.

The prefilter is obtained by solving the optimization problem given by

$$\mathrm{MIN}_a : f(\mathbf{a}) = \log\left(\frac{f_u(\mathbf{a})}{f_d(\mathbf{a})}\right) \tag{10}$$

with

$$f_d(\mathbf{a}) = ||\mathbf{b}_d||_{p_d} = \left(\sum_{i=1}^{N_m} \sum_{k=0}^{L_g - 1} |g_d^{(i)}(k)|^{p_d}\right)^{\frac{1}{p_d}} \tag{11}$$

and $f_u(\mathbf{a}) = ||\mathbf{b}_u||_{p_u}$, accordingly. N_m is the number of target points in the listening area. The vectors $\mathbf{g}_d$ and $\mathbf{g}_u$ consists 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 large values for p_d and p_u (typically between 10 and 20), a very smooth shaping, with no outliers, can be achieved.

4. EXPERIMENTS

The experiments have been performed in order to test how the sparse reconstruction and interpolation can improve the reshaping of a whole volume when the actual listener position is known and can be used for the design of reshaping filters. The first experiment uses random positions and performs the equalization based on the nearest neighbor principle. In the second experiment, these known positions are used for a sparse recovery of a grid in the target area. These estimates are again used for a reshaping. Finally, in the third experiment, interpolation between the grid points is performed. This method allows for a design of equalizers for all points independently.

4.1. Experimental Setup

The experiments are based on simulated RIRs of length $L_c = 500$ in an office-sized room with the dimensions of $5.8\,\text{m} \times 4.15\,\text{m} \times 2.55\,\text{m}$ [26]. 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 target area is a plane of size $0.5 \times 0.5\,\text{m}$. On this plane, we acquired RIRs at $M \in \{50, 100, 150, 300\}$ random positions $\mathbf{r}_m \in \mathbb{R}^2$. The random measurement positions for the case $M = 150$ are shown in Fig. 1.

4.2. Random Positions

The first experiment was conducted using the random positions directly for equalization. For every measured RIR $h_i(n)$, an equalizer $a_i(n)$ of length $L_a = 500$ has been estimated using (10). With these equalizers, the best equalization for the whole area has been achieved by choosing the equalizer which is the nearest in terms of spatial distance for all points. In Fig. 2 the improvement/deterioration is shown. The color codes the improvement in terms of ΔnPRQ, 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

$$g_{\text{os}}(n) = \max\left(\frac{1}{w_u(n)}, -60\text{dB}\right) \tag{12}$$

as

$$\text{nPRQ} = \begin{cases} \frac{1}{\|\boldsymbol{g}_E\|_0} \sum_{n=N_0}^{L_g-1} g_E(n), & \|\boldsymbol{g}_E\|_0 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

with

$$g_E(n) = \begin{cases} 20\log_{10}(|g(n)|w_u(n)), & |g(n)| > g_{\text{os}}(n) \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

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 example in Fig. 2 shows that at all of the measurement points a good equalization could be achieved. Due to spatial mismatch and not dense enough sampling, the areas in between are not equalized sufficiently. The first line of Table 1 shows the average improvement for the different cases. Only the case with $M = 300$ could achieve a substantial improvement. In all other cases the amount of measurements is too small for a successful reshaping. In these cases, it would be better to not process at all.

Table 1. Comparison of the performance in the target area for different amount of random measurement points. The improvement is calculated in terms of average ΔnPRQ-measure in the target area.

M	50	100	150	300
Random	0.76	-0.16	-0.70	-1.95
Grid 18×18	0.41	-0.91	-1.55	-1.97
Grid 171×171	-0.03	-2.05	-3.23	-4.25

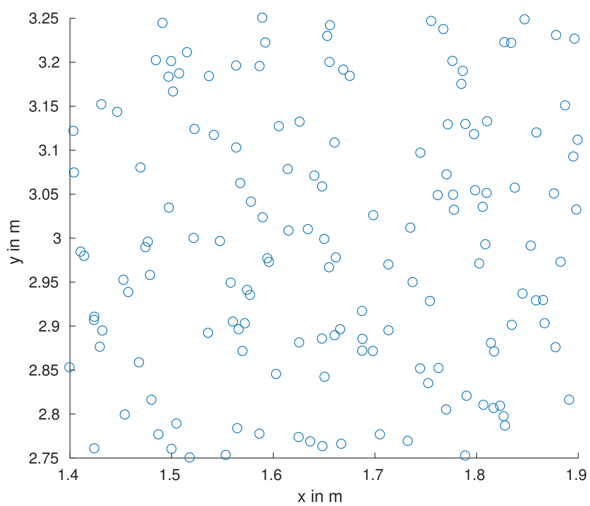

Fig. 1. The random positions in the target area used for equalization. The case with $M = 150$ is shown.

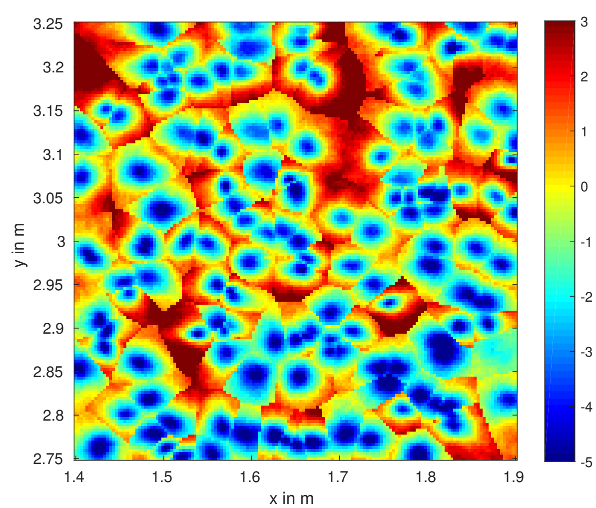

Fig. 2. Equalization using the random positions from Fig. 1. The color codes the improvement/deterioration of the perceived echoes in terms of nPRQ. Blue and green mean improvement, yellow indicates no change and red colors show added reverberation. The plain method is not successful.

4.3. Grid Recovery

The second experiment uses the virtual-grid RIRs recovered from the measured RIRs. For setting up the linear system, we designed a

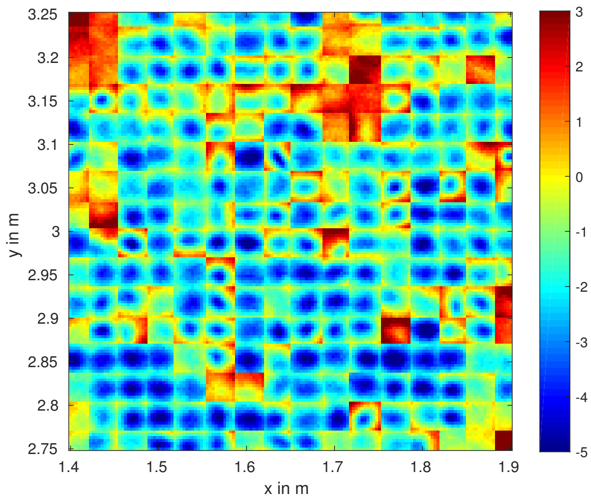

Fig. 3. Equalization on grid RIRs recovered by CS based reconstruction. The regular square structure makes the grid positions visible.

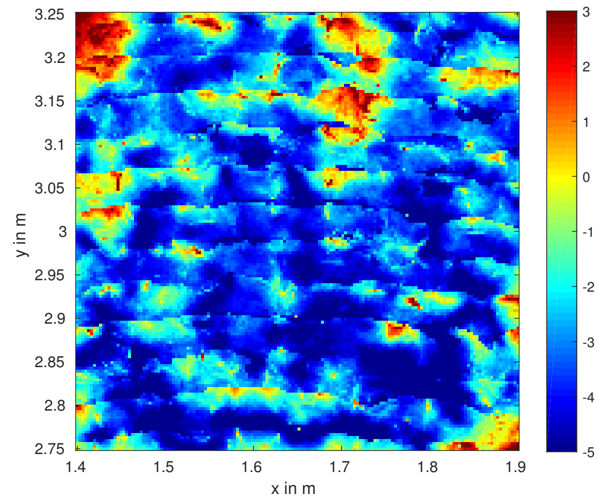

Fig. 4. Equalization after CS based reconstruction and interpolation of the RIRs for all positions. The proposed method allows for a good equalization at almost all points in the target area.

18×18 grid with spacing $\Delta = 3.33$ cm covering the measurement area. The outer grid points were located outside of the measured plane, in order to avoid a truncation of interpolation kernels. We used separable Lagrange interpolators of order three for each dimension (cf. [17]). For the CS based recovery, we used the iterative hard-thresholding algorithm [21] with step size $\mu = 5 \cdot 10^{-2}$ and 3000 iterations. The iterative recovery was started with the initial estimate being the zero vector. Beginning with a small $K_0 = 16M$, the sparsity constraint was successively relaxed every 50 iterations by $K_{i+1} = K_i + K_0$ (cf. [17]).

For the recovered grid RIRs, equalizers have been calculated the same way as in the first experiment. The higher amount of available RIRs allows for, on average, smaller distances from reference positions when using the nearest neighbor approach. The results for the $M = 150$ case are shown in Fig. 3. Here, the grid structure is clearly visible. Most of the recovered RIRs allow for good equalization at their positions and near by. The second line of Table 1 shows that in all cases the average nPRQ has improved. For the case of $M = 300$ the improvement is negligible, since the amount of recovered RIRs, 324, is almost the same.

In general, the results indicate, that the recovery error has smaller influence on the reshaping performance than the displacement error.

4.4. Sparse Reconstruction of Listening Area

In the third experiment, interpolation for all points has been used with the virtual-grid data from the second experiment. The same Lagrange interpolator was employed to obtain RIRs at any target point (i.e., listener position). This allows for an independent design of equalizers for all of these points. In Fig. 4 the results are shown. Almost all points could be successfully processed. The third line of Table 1 shows that, in all cases, the equalization performance could be improved.

Again, the results show that for RIR reshaping it is advantageous to estimate the RIRs at the desired target point. The estimation errors during the sparse reconstruction and interpolation stages have less impact than the errors due to spatial mismatch.

5. CONCLUSION

The traditional approach of reshaping of RIRs based on nearest neighbor approach shows poor performance in the case of spatial mismatch. The proposed reconstruction method employing compressed sensing and interpolation methods allows for a better estimation of RIRs in the target area. Using these RIRs for reshaping leads to an overall better equalization of the target area.

6. REFERENCES

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, 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, no. 11, 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 Signals, Systems, and Computers*, Pacific Grove, CA , USA, Oct. 2005, pp. 898–902.

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

[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, no. 2, 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, no. 3, 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, no. 6, 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] 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, May 2013, pp. 433–437.

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

[11] F. Katzberg, R. Mazur, M. Maass, M. Böhme, and A. Mertins, "Spatial interpolation of room impulse responses using compressed sensing," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Tokyo, Japan, September 2018.

[12] E. Candès and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Febr. 2006.

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[14] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Sept. 2006.

[15] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer, 2015.

[16] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Multigrid reconstruction of sound fields using moving microphones," in *Proc. Workshop on Hands-free Speech Communication and Microphone Arrays*, March 2017, pp. 191–195.

[17] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 1962–1975, November 2018.

[18] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, April 1995.

[19] J. A Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[20] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[21] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, Dec. 2008.

[22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[23] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, March 2006.

[24] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Stat. Soc., Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[25] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.

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