A Compressed Sensing Framework for Dynamic Sound-Field Measurements

Fabrice Katzberg, *Student Member, IEEE*, Radoslaw Mazur, *Member, IEEE*, Marco Maass, *Student Member, IEEE*, Philipp Koch, *Student Member, IEEE*, and Alfred Mertins, *Senior Member, IEEE*

***Abstract*—The conventional sampling of sound fields by use of stationary microphones is impractical for large bandwidths. Satisfying the Nyquist–Shannon sampling theorem in three-dimensional space requires a huge number of sampling positions. Dynamic sound-field measurements with moving microphones together with a compressed-sensing recovery allow for weakening the spatial sampling problem. For bandlimited signals, the dynamic samples taken along the microphone trajectory may be related to the room impulse responses on a virtual grid in space via spatial interpolation. The tracking of the microphone positions and the knowledge of the excitation sequence allow for setting up a linear system of equations that can be solved for the room impulse responses on the modeled virtual grid. Nevertheless, there is still the necessity for recovering a huge number of sound-field variables, in order to ensure aliasing-free reconstruction. Thus, for practical applications, random or suboptimally chosen trajectories may be expected to lead to underdetermined sampling problems for a given volume of interest. In this paper, we present a compressed sensing framework that enables us to uniquely solve the dynamic sampling problem despite having underdetermined variables. The spatio-temporal sampling problem is integrated into compressed sensing models that allow for stable and robust sub-Nyquist sampling given incoherent measurements. For a modeled equidistant grid and sparse Fourier representations, the influence of the microphone trajectories on the compressed sensing problem is investigated and a simple expression is derived for evaluating trajectories with regard to compressed-sensing based recovery.**

***Index Terms*—Compressed sensing, room impulse responses, dynamic sound-field measurement, microphone array.**

I. Introduction

In CLOSED-ROOM scenarios, sound propagation usually involves multiple reflections from the walls leading to reverberation. However, acoustic applications such as multichannel sound systems and wavefield-synthesis systems typically assume a free-field or low-reverberation environment and decrease their performance when the sound field is influenced by multiple strong interfering reflection paths. In order to reduce the acoustical scene effects and improve performance in reverberant environments, methods for listening-room compensation are often applied [1]–[6]. For this, the exact knowledge of the sound field within the volume of interest is essential.

Common stationary methods for measuring room impulse responses (RIRs) that describe the sound field are the use of perfect sequences [7], [8], maximum-length sequences (MLS) [9], and exponential sine sweeps for excitation [10].

In [11], the spatio-temporal sampling of RIRs has been investigated. The spatial sampling of RIRs by use of stationary microphones requires an extremely high effort in calibration. In addition to the need for compensating spatio-temporal deviations, the frequency response of each microphone must be equalized. Moreover, an array of microphones will most likely never be dense enough to satisfy the Nyquist-Shannon sampling theorem without significant problems for very large audio bandwidths.

In order to make spatial sampling practical, methods for the dynamic measurement of RIRs have been proposed. In [12], a technique is presented that allows for the reconstruction of RIRs along the trajectory of a moving microphone by exploiting the Doppler effect. At this, a special input signal is required and the velocity of the microphone must be constant. A totally different dynamic approach has been proposed very recently in [13], where measurements taken along tracked microphone trajectories are related to a modeled sampling grid in space via interpolation. Based on this, a linear system of equations is set up, whose solution yields the RIRs on the modeled grid.

For the case that only a small number of unknown parameters are nonzero, the theory of compressed sensing (CS) [14], [15] allows for uniquely solving underdetermined systems of linear equations with sparsity constraint. From a signal processing point of view, this translates to sub-Nyquist sampling of a signal by merging compression and sampling into one step. Given incoherent measurements [16], the reconstruction of an unknown signal is possible, although the Nyquist-Shannon requirements are not directly met.

There exist methods that exploit the sparsity of the early parts of RIRs and the exponential decay of later parts for CS based RIR recordings. In [17], large sets of RIRs for multiple fixed source-microphone configurations have been recovered by using a convex optimization algorithm that considers convex penalties promoting both sparsity and the exponentially decaying envelope of RIRs. Also for the recovery of entire sound-fields with only a few stationary, randomly placed microphones, there are

Manuscript received November 15, 2017; revised March 26, 2018 and June 8, 2018; accepted June 11, 2018. Date of publication June 27, 2018; date of current version August 8, 2018. This work was supported by the German Research Foundation under Grants ME1170/10-1 and ME1170/8-1. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hiroshi Saruwatari. *(Corresponding author: Fabrice Katzberg.)*

The authors are with the Institute for Signal Processing, Universtiy of Lübeck, Lübeck 23562, Germany (e-mail: katzberg@isip.uni-luebeck.de; mazur@isip.uni-luebeck.de; maass@isip.uni-luebeck.de; koch@isip.uni-luebeck.de; mertins@isip.uni-luebeck.de).

Digital Object Identifier 10.1109/TASLP.2018.2851144

CS based techniques available [18], [19]. In [18], the sparsity of RIRs in time domain is exploited to interpolate the early-reflection part within a volume of interest (VOI). In [19], a sparse plane wave approximation is applied to reconstruct the sound field at low frequencies up to 400 Hz. Based on measurements with a spherical microphone array, a CS framework using spherical harmonics for sound-field reconstruction has been proposed in [20]. Also a sparse spherical-harmonics approximation is used in [21], in order to interpolate RIRs given stationary measurements.

In this paper, we present a CS framework for the sound-field recovery from samples taken by moving microphones that move within the volume or area of interest. The method allows us to provide the information required for techniques such as listening-room compensation. The measurement framework is based on the dynamic model proposed in [13]. While in [13] the sampling matrix was required to have full column rank, we here consider the underdetermined case. This is important for practical applications in which the microphone travels along quasi random trajectories (e.g., using hand-held microphones), where the resulting linear system turns out to be ill-posed and underdetermined.

By exploiting that the spectrum of sound fields is ideally restricted to a hypercone along the temporal frequency axis [11], we combine the dynamic sampling procedure with a CS framework that allows for stable and robust sound-field recovery despite of underdetermined variables. The influence of the microphone trajectory on the resulting CS matrix is investigated and a simple trajectory-dependent expression for the coherence of the CS problem is derived for spectrally flat excitation. Furthermore, we embed two extensions of the dynamic sampling model into the CS based recovery. By using perfect sequences for excitation, the time-decoupled problem [13] enables us to recover the sound field at reduced computational demand. The multigrid approach [22] allows us for decomposing the sound-field spectrum into disjoint subbands which may be recovered separately.

The paper is organized as follows. Models of sound-field sampling and reconstruction as well as the dynamic sampling model with its extensions are outlined in Section II. Section III emphasizes spatial sampling difficulties that go along with sound-field sampling, even in case of dynamic measurements. The common theoretical tools that allow for CS based recovery and our CS model for the dynamic sound-field sampling are given in Section IV. The structure of the CS matrix for Fourier representations, including a trajectory and coherence analysis, are presented in Section V. Based on the iterative hard thresholding method (IHT) [23], we provide a simple CS algorithm for the sound-field recovery in Section VI and present the results of experiments with simulated data in Section VII.

II. Dynamic Sampling Model

For a given configuration of a sound source emitting a Dirac pulse, the room impulse response $h(t)$ characterizes the course of the received sound pressure at a specific listener position depending on time t. Assuming an acoustical environment which is a linear time-invariant (LTI) system for a fixed emitter-receiver pair, the temporal relation between the excitation signal $s(t)$ and the observation signal $x(t)$ is

$$x(t) = \int_{-\infty}^{\infty} h(\tau)s(t-\tau)d\tau. \tag{1}$$

Sound fields describe the sound pressure with respect to both time t and receiver position $\boldsymbol{r} = [r_x, r_y, r_z]^T$. For a single sound source at fixed position, the sound pressure field is

$$p(\boldsymbol{r}, t) = \int_{-\infty}^{\infty} h(\boldsymbol{r}, \tau)s(t-\tau)d\tau, \tag{2}$$

where $h(\boldsymbol{r}, t)$ is the spatially varying RIR from the source location to the point $\boldsymbol{r}$. In [24], the sound field according to (2) was termed plenacoustic function. Characteristics and sampling of that function are explicitly described in [11]. For the case in which a single source at a fixed position emits a Dirac pulse at $t = 0$, the sound field is simplified to the spatio-temporal RIR $p(\boldsymbol{r}, t) = h(\boldsymbol{r}, t)$ [11], [24].

A. Models for Sound-Field Reconstruction

For our dynamic sampling scheme, we define fixed virtual sampling locations inside the volume of interest. Moving microphones, in general, sample at varying intermediate positions between these virtual points. The idea is to use the dynamic measurements to set up a linear system of equations with RIRs at the virtual locations as variables. Accordingly, the sampling matrix defined as $\boldsymbol{A}$ must perform the interpolation task of relating the spatio-temporal data acquired at intermediate positions to the sought RIRs at the virtual locations. Solving the system for the unknown variables, in our framework by use of CS principles, provides the desired sound-field information at the virtual points in space. In order to allow for the reconstruction of band-limited sound fields, these virtual points must follow sampling strategies for which it is known that an error-free interpolation is possible, such as, e.g., uniform, quincunx, and spherical patterns.

In comparison to a uniform-grid model, quincunx and spherical models may lead to a smaller number of unknowns in the linear system, however, they clearly raise the complexity in modeling and analyzing the sampling matrix. For quincunx sampling [11], the non-trivial interpolation filters building $\boldsymbol{A}$ are not separable in dimensions. For spherical sampling patterns, physically inspired by the Helmholtz integral theorem [25], $\boldsymbol{A}$ would need to model the dependence of the microphone samples taken within the VOI to the coefficients of a spherical-harmonics expansion. Alternatively, one might also consider sampling with a moving microphone on the surface of the VOI only, leading to a different interpolation problem, but this would require a completely different measurement setup in which sampling with a hand-held, position-tracked microphone is not an option. While the spherical array is specially designed to address the measurement of 3D volumes, a uniform-grid model provides the full scalability to recover along lines, on planes, and within volumes. Further, a virtual equidistant grid allows for a simple and straightforward approach to relate samples acquired during the dynamic sampling process to the unknowns in the linear system. The resulting sampling matrix $\boldsymbol{A}$ only contains the source signal and spatial interpolation coefficients that depend on the microphone trajectory. The effort for setting up and solving the system is extremely low. In general, the spatial dimensions on the uniform grid are separable, which allows for efficiently implementing sparsifying transforms for our CS framework. Also the interpolation coefficients in $\boldsymbol{A}$ become separable, which reduces the effort for calculating and evaluating higher-order interpolation polynomials. Taking all these points into account, we chose a uniform-grid model for the dynamic sampling problem, as also considered in [13], [26].

Let $T = 1/f_s$ denote the sampling interval in time with sampling frequency f_s fulfilling the Nyquist-Shannon sampling theorem $f_s > 2f_c$, where f_c is the temporal cutoff frequency that, in our case of sound-field sampling, is assumed to be equal to the audio bandwidth involved. Accordingly, we have samples at equidistant points $t_n = nT$ with $n \in \mathbb{N}_0$ being the discrete variable of the causal time signal. For the uniform sampling in the spatial dimensions, we model a Cartesian grid where the equidistant sampling points $\boldsymbol{r}_{\boldsymbol{g}} \in \mathcal{G}$ are given by the set

$$\mathcal{G} = \left\{ \boldsymbol{r}_{\boldsymbol{g}} \mid \boldsymbol{r}_{\boldsymbol{g}} = \boldsymbol{r}_0 + [g_x \Delta_x, g_y \Delta_y, g_z \Delta_z]^T \right\} \tag{3}$$

with the grid origin $\boldsymbol{r}_0$ and the discrete grid variables in $\boldsymbol{g} = [g_x, g_y, g_z]^T \in \mathbb{Z}^3$. In order to reconstruct sound waves in space with minimal wavelength $\lambda = c_0/f_c$ without aliasing, the spatial grid requires $\Delta_\xi < c_0/(2f_c) \,\forall\, \xi \in \{x, y, z\}$, where c_0 is the speed of sound and Δ_ξ denotes the spatial sampling interval for each dimension x, y, z.

By using a 4D sinc filter with infinite support, the continuous sound field $h(\boldsymbol{r}, t)$ can be perfectly reconstructed from the discrete grid RIRs $h(\boldsymbol{r}_{\boldsymbol{g}}, t_n)$. However in practice, the sampling process and thus the number of sampling points are limited. The amplitudes of the grid RIRs decrease exponentially and are assumed to vanish into the noise level beyond t_{L-1} for given f_s. Hence, despite limiting the time samples of the signal $h(\boldsymbol{r}_{\boldsymbol{g}}, t_n)$ to L taps, finite length interpolation filters allow for reasonable approximations in the temporal dimension. The spatial sampling is restricted to a volume of interest in practice, leading to a finite grid in space of size $X \times Y \times Z$. This translates to a rectangular windowing of the sound field along the spatial dimensions, and, thus, to a convolution of the sound-field spectrum with a sinc function along each spatial frequency variable. Nevertheless, the larger the region of sampling is, the faster the spectral decay of the windowed sound field will be [11]. Consequently, to improve the spatial reconstruction despite finite support, either the measuring area has to be chosen larger than the volume of interest, or the spatial sampling grid has to be chosen finer, well above the Nyquist rate. The number of spatial sampling points increases in both cases.

B. Dynamic Sampling Procedure

The sound-field measurement with moving microphones may be regarded as the reverse interpolation problem in space [13]. In general, the sampling points $x(\boldsymbol{r}, n)$ are uniformly taken in the time dimension, but dynamically taken at intermediate positions $\boldsymbol{r}$ of a virtual spatial grid according to (3). The following description considers a single microphone sampling on trajectory $\boldsymbol{r}(n)$, however the extension to multiple microphones is straightforward.

Aiming at the recovery of the finite set of $N = XYZ$ RIRs at discrete grid positions $\boldsymbol{g} \in G$, with

$$G = \{0, \ldots, X-1\} \times \{0, \ldots, Y-1\} \times \{0, \ldots, Z-1\} \tag{4}$$

spanning a regular grid inside the volume of interest, the dynamic measurement process is modeled as

$$x(\boldsymbol{r}(n), n) = \sum_{m=0}^{L-1} \sum_{\boldsymbol{g} \in G} \varphi_n(\boldsymbol{g})\, h(\boldsymbol{g}, m)\, s(n-m) + \eta(n), \tag{5}$$

where $\varphi_n(\boldsymbol{g})$ is an interpolation function for approximating the sound field at continuous position $\boldsymbol{r}(n)$ for discrete time points n as linear combination of the grid RIRs $h(\boldsymbol{g}, n)$ subject to the displacements $\boldsymbol{r}_{\boldsymbol{g}} - \boldsymbol{r}(n)$. The term $\eta(n)$ is a perturbation comprising the measurement noise and the error of the bandlimited interpolation in space. Accordingly, for M samples taken by the moving microphone, the linear measurement model is

$$\boldsymbol{x} = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S} \boldsymbol{h}_u + \boldsymbol{\eta}, \tag{6}$$

where $\boldsymbol{x} \in \mathbb{R}^M$ is the measurement vector,

$$\boldsymbol{x} = [x(\boldsymbol{r}(0), 0), \ldots, x(\boldsymbol{r}(M-1), M-1)]^T, \tag{7}$$

$\boldsymbol{\eta} \in \mathbb{R}^M$ is the noise vector,

$$\boldsymbol{\eta} = [\eta(0), \ldots, \eta(M-1)]^T, \tag{8}$$

$\boldsymbol{h}_u \in \mathbb{R}^L$ contains the u-th RIR on a virtual grid in space,

$$\boldsymbol{h}_u = [h(\boldsymbol{g}_u, 0), \ldots, h(\boldsymbol{g}_u, L-1)]^T, \tag{9}$$

$\boldsymbol{S} \in \mathbb{R}^{M \times L}$ is the convolution matrix of the source signal, and $\boldsymbol{\Phi}_u \in \mathbb{R}^{M \times M}$ is a diagonal matrix stacking all M interpolation coefficients for the u-th virtual-grid RIR,

$$\boldsymbol{\Phi}_u = \operatorname{diag}\{\varphi_0(\boldsymbol{g}_u), \ldots, \varphi_{M-1}(\boldsymbol{g}_u)\}. \tag{10}$$

The modeling of positions $\boldsymbol{r}_{\boldsymbol{g}_u}$ forming a uniform grid in space, the tracking of the trajectory $\boldsymbol{r}(n)$, and the knowledge about the excitation signal $s(n)$ allow for setting up the system of linear equations

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{\eta}, \tag{11}$$

with system matrix $\boldsymbol{A} \in \mathbb{R}^{M \times U}$ having the block structure

$$\boldsymbol{A} = \left[\boldsymbol{\Phi}_1 \boldsymbol{S}, \boldsymbol{\Phi}_2 \boldsymbol{S}, \ldots, \boldsymbol{\Phi}_N \boldsymbol{S} \right] \tag{12}$$

and vector $\boldsymbol{h} = [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_N^T]^T$ consisting of $U = NL$ unknown sound-field variables in total, i.e., the concatenated grid RIRs. In [13], the sampling matrix $\boldsymbol{A}$ was assumed to have full column rank and to formulate a well-posed problem, in order to obtain the estimate of $h(\boldsymbol{g}, n)$ by simply calculating the unique least-squares solution of the linear system (11). Then, for spectrally flat excitation, it could be shown by analyzing $\boldsymbol{A}\boldsymbol{A}^T$ with $M = U$ that the recovery error is highly dependent on the terms $\sum_{u=1}^{N} \varphi_n^2(\boldsymbol{g}_u)$ and, thus, on the trajectory, the modeled spatial grid, and the used interpolation function. Apart from the actual interpolation accuracy, the error is smaller for positions $\boldsymbol{r}(n)$ sampled closer to the grid points $\boldsymbol{r_g}$ and minimal for the set of parameters $\mathcal{P} = \{(\boldsymbol{r}(n), \mathcal{G})\}$ ensuring

$$\underset{\mathcal{P}}{\arg\max} \sum_{n=0}^{M-1} \sum_{u=1}^{N} \varphi_n^2(\boldsymbol{g}_u), \tag{13}$$

which corresponds to a moving microphone sampling exclusively on the grid positions in an arbitrary temporal order [13].

C. Time-Decoupled Problem

By using a repetitive source sequence $s(n)$ of period length L with perfect autocorrelation function leading to the circular convolution matrix $\boldsymbol{S}_o \in \mathbb{R}^{L\times L}$ with $\boldsymbol{S}_o^T \boldsymbol{S}_o = \gamma \boldsymbol{I}_{L\times L}$ for one period of excitation in steady state, the computational complexity for the recovery of $\boldsymbol{h}$ can be substantially reduced [13]. Such a perfect sequence allows for decomposing (11) into L time-decoupled sub-problems, each with a number of unknowns growing only with the size of the spatial grid, i.e., proportional to only f_c^3 instead of f_c^4. For R periods of excitation, the problem (11) simplifies to

$$\boldsymbol{x} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{h}} + \boldsymbol{\eta}, \tag{14}$$

with the highly structured $RL \times NL$ system matrix

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{\Phi}_{1,1} & \dots & \boldsymbol{\Phi}_{N,1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{1,R} & \dots & \boldsymbol{\Phi}_{N,R} \end{bmatrix} \tag{15}$$

consisting of $R \times N$ blocks of diagonal matrices $\boldsymbol{\Phi}_{u,r}$ carrying the L interpolation coefficients of the u-th grid RIR at the r-th period of excitation, and the vector

$$\tilde{\boldsymbol{h}} = \left[(\boldsymbol{S}_o \boldsymbol{h}_1)^T, \dots, (\boldsymbol{S}_o \boldsymbol{h}_N)^T\right]^T = (\boldsymbol{I}_N \otimes \boldsymbol{S}_o)\boldsymbol{h} \tag{16}$$

containing a time-filtered version of $h(\boldsymbol{g}, n)$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{I}_N$ is the identity matrix of size $N \times N$. The filter given by the excitation sequence in $\boldsymbol{S}_o$ only scales the magnitude and changes the phases of the temporal frequency components of the sound-field variables in $\boldsymbol{h}$. According to the block-diagonal structure of $\tilde{\boldsymbol{A}}$, the problem (14) is actually composed of L sub-problems of size $R \times N$ which may be solved separately [13]. Unique least-squares solutions are only possible for $R \geq N$. The sought grid RIRs are obtained by subsequent inverse time-filtering according to $\boldsymbol{h} = \gamma^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{S}_o^T)\tilde{\boldsymbol{h}}$.

D. Multigrid Problem

By exploiting the dispersion relation of propagative sound waves, the dynamic sampling problem (11) may also be decomposed into multiple subband problems by using a multiresolution recovery scheme [22]. According to the LTI model ensuring a constant speed of sound c_0, the dispersion relation

$$\kappa_x^2 + \kappa_y^2 + \kappa_z^2 = \frac{\omega^2}{c_0^2} \tag{17}$$

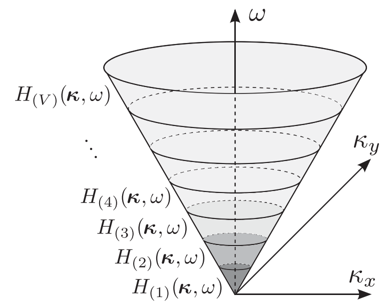


Fig. 1. Spectral cone of sound fields in 2D space. By using the multigrid model, the broadband cone is decomposed into V distinct subbands $H_{(v)}(\boldsymbol{\kappa}, \omega)$ that may be reconstructed separately.

gives a direct relationship between the spatial frequencies κ_x, κ_y, κ_z in rad m^{-1} and the temporal angular frequency $\omega = 2\pi f$ in rad s^{-1} for the continuous case [11]. Consequently, the decomposition of the sound field into V temporal subbands with band limits $\omega_c^{(v-1)} \leq \omega < \omega_c^{(v)}$ ($v \in \{1, \dots, V\}$) inherently induces band limiting in the spatial domain and allows for restricting the corresponding spectra in space to

$$\frac{\omega_c^{(v-1)}}{c_0} \leq \sqrt{\kappa_x^{2^{(v)}} + \kappa_y^{2^{(v)}} + \kappa_z^{2^{(v)}}} < \frac{\omega_c^{(v)}}{c_0}. \tag{18}$$

A subband scheme with constant temporal bandwidth is outlined in Fig. 1 for a plane grid in 2D space.

The idea of the multigrid approach is to decompose the measurement signal $x(n)$ into distinct temporal subbands. According to (18), lower temporal frequencies involve lower spatial frequencies, thus, for the recovery of lower sound-field bands, simply a coarser virtual grid in space may be modeled. By designing resolution levels $v \in \{1, \dots, V\}$ of the virtual grid in space, where the highest resolution $\Delta^{(V)}$ is supposed to fulfill the sampling theorem for the global cutoff ω_c, the broadband problem (11) is decomposed into the subband problems

$$\boldsymbol{x}^{(v)} = \boldsymbol{A}^{(v)}\boldsymbol{h}^{(v)} + \boldsymbol{\eta}^{(v)} \tag{19}$$

with $\boldsymbol{h}^{(v)} \in \mathbb{R}^{LN^{(v)}}$ containing RIRs on the downsampled grid of level v and $\boldsymbol{x}^{(v)}$ containing the bandpass filtered measurement signal which guarantees the spatial Nyquist-Shannon condition for the particular downsampling factor. The number of samples M is unchanged for each sub-problem, whereas the number of unknown variables is smaller for coarser grids, i.e., $\boldsymbol{A}^{(v)} \in \mathbb{R}^{M\times LN^{(v)}}$ with $N^{(v)} < N^{(v+1)}$. Consequently, lower grid levels allow for more samples per grid RIR, which makes the recovery of low frequencies more robust against noise. The subproblems (19) may be incorporated into a CS based optimization algorithm for further performance gain. In order to obtain the broadband sound field sampled on the finest modeled grid, each of the distinct bands recovered by (19) is spatially upsampled to $\Delta^{(V)}$ and finally summed up. The multiresolution recovery for CS is tested in the experimental part of this paper.

III. Sub-Nyquist Sampling With Moving Microphones

For practical applications, arbitrary microphone trajectories most likely lead to ill-posed or even underdetermined problems for a given volume of interest. In fact, also microphones that are moved in a controlled way along a pre-defined trajectory while taking a number of M samples that is in the order of U will hardly result in a sampling process that allows for uniquely determining all U sound-field variables on a large grid in space. This is due to the following circumstances:

1) In order to satisfy the Nyquist-Shannon sampling theorem in time and 3D space, the number of unknown sound-field variables grows proportionally to f_c^4, thus, the problem becomes very large in practice.
2) A spectrally flat FIR filter approximation for 3D interpolation generally has fast spatial decay, thus, one sample taken by moving microphones effectively cumulates information of a very narrow segment on the grid to be recovered.
3) Boundary grid-positions are not fully surrounded by samples. Measurements are only available from the inside of the volume of interest. For measuring each grid point with equal amount, the boundary of the grid requires more samples than inner parts.
4) With respect to the omnidirectional support of appropriate interpolation kernels (cf. Section V-B), the boundary of the virtual grid needs to be extended beyond the sampled volume of interest, which increases the number of unknowns and inherently leads to an underdetermined linear system (11).

The combination of all above-mentioned points in fact leads to a trajectory dilemma that makes it impossible to solve (11) with conventional methods unless an elongated measurement and a very dense trajectory are used. The problem requires sub-Nyquist sampling in practice, which is possible by exploiting the principle of CS. Therefore, in this work, we combine the dynamic sampling procedure with a CS framework allowing for stable and robust recovery of $h(\boldsymbol{g}, n)$ despite having underdetermined variables. The framework also incorporates the two extensions of the dynamic sampling model, namely the time-decoupling (14) and the multigrid partitioning (19).

By comparison with (11), the formulation (14) tightens the requirements on the trajectory for obtaining a linear system with fully determined variables. Due to the periodic excitation, the trajectory needs to fulfill

$$r(n) \neq \boldsymbol{r}(n - lL)\ \forall\, l \in \mathbb{Z}\backslash\{0\}, \tag{20}$$

in order to avoid linearly dependent rows in $\tilde{\boldsymbol{A}}$. Furthermore, the time-decoupling demands that the microphone measures each phase of the periodic excitation completely within the spatial interpolation range of any grid RIR. To give an extreme example, consider the spatial split of the virtual-grid positions into disjoint subsets $\mathcal{G}_1 \cup \mathcal{G}_2 = \mathcal{G}$ with $\mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$ and the temporal split of the excitation phases $\rho(n) = (n \bmod L)$ into disjoint subsets $\mathcal{T}_1 \cup \mathcal{T}_2 = \{0, \ldots, L-1\}$ with $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$. In case of a moving microphone with spatial sampling points $\boldsymbol{r}(n)$ covering the entire volume of interest, but being located in the interpolation range of $\mathcal{G}_1$ only at time points n with $\rho(n) \in \mathcal{T}_1$ and in the range of $\mathcal{G}_2$ only for $\rho(n) \in \mathcal{T}_2$, the linear system (14) will always be underdetermined, even for an infinite number of dynamic samples. The reason behind this is the structure of $\tilde{\boldsymbol{A}}$, which inherently turns the temporal order of positions $\boldsymbol{r}(n)$ into an important key factor for obtaining a well-posed problem. The CS based recovery involving additional regularization may weaken this strong dependency on the trajectory.

IV. Compressed Sensing Framework

Room impulse responses in closed rooms are usually sparse in their early parts (due to direct and a few successive sound reflections at the room walls) and dense in their later parts (due to multiple reflections). In addition, following the dispersion relation (17), the spatio-temporal spectrum of $h(\boldsymbol{r}, t)$ is ideally measurable on the 3D surface of a 4D hypercone along the temporal frequency axis ω. In Sections IV-A and IV-B, we outline the common theoretical tools that generally allow for CS based recovery under ideal and non-ideal conditions, respectively. Based on that theory, the CS model for the introduced spatio-temporal measurement procedure is characterized in Section IV-C.

A. Sparse Recovery From Noiseless Data

The basic sampling model (6) allows for the recovery of signal $h(\boldsymbol{g}, n)$ in vector $\boldsymbol{h} \in \mathbb{R}^U$ using M samples of measurement $x(n)$ in vector $\boldsymbol{x} \in \mathbb{R}^M$. The link between both signal vectors is given by the system matrix $\boldsymbol{A} \in \mathbb{R}^{M\times U}$ whose m-th row is the sampling vector $\boldsymbol{\phi}_m^H$ performing the projection of the unknown parameters onto measurement space,

$$x(m-1) = \langle \boldsymbol{h}, \boldsymbol{\phi} \rangle = \boldsymbol{\phi}_m^H \boldsymbol{h}. \tag{21}$$

For the underdetermined case with a set of linearly independent sampling vectors smaller than U, the principle of CS enables us to identify a unique vector $\boldsymbol{h}$ that solves the linear system together with a sparsity constraint.

The general requirement for CS is that $\boldsymbol{h}$ lives in a subspace of dimension $K \ll M$ with respect to an appropriate basis. Considering the sparsely occupied hypercone forming the sound-field spectrum, let $\boldsymbol{T}_B \in \mathbb{C}^{B\times B}$ be a unitary matrix performing an orthogonal transform of a 1D signal into some frequency representation. For a multidimensional regular grid, assume the separability of the transform. Regarding grid RIRs concatenated in $\boldsymbol{h}$ first along the x dimension and then along the y and z dimensions in succession, the sampled sound field is described by the coefficients of the 4D frequency representation encapsulated in

$$\boldsymbol{c} = \boldsymbol{\Psi}\boldsymbol{h}, \tag{22}$$

where $\boldsymbol{\Psi} = \boldsymbol{T}_Z \otimes \boldsymbol{T}_Y \otimes \boldsymbol{T}_X \otimes \boldsymbol{T}_L$ is a unitary $U \times U$ matrix and $\boldsymbol{c} \in \mathbb{C}^U$ is a K-sparse vector, which means that only a small number of, at most, K frequency coefficients in $\boldsymbol{c}$ is non-zero.

In case of no perturbations and perfect sparsity of $\boldsymbol{c}$, the basic sampling problem (11) can be regularized according to

$$\underset{\boldsymbol{c}\in\mathbb{C}^U}{\arg\min}\ \|\boldsymbol{c}\|_{\ell_0} \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{A}\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{h} = \boldsymbol{\mathcal{A}}\boldsymbol{c} \tag{23}$$

with $\boldsymbol{\mathcal{A}} = \boldsymbol{A}\boldsymbol{\Psi}^H$ and the pseudonorm $\|\boldsymbol{c}\|_{\ell_0}$ counting the number of non-zero elements in $\boldsymbol{c}$. Defining $\mathcal{C}_K$ as the set of all K-sparse signal vectors and the spark of matrix $\boldsymbol{\mathcal{A}}$ as the smallest number of columns being linearly dependent,

$$\mathrm{spark}(\boldsymbol{\mathcal{A}}) = \min\{K : \mathrm{kern}(\boldsymbol{\mathcal{A}}) \cap \mathcal{C}_K \neq \{\mathbf{0}\}\}, \tag{24}$$

a unique solution of the regularized problem (23) is guaranteed for $\mathrm{spark}(\boldsymbol{\mathcal{A}}) > 2K$ [27]. This means that vectors living in the null space of $\boldsymbol{\mathcal{A}}$, except the trivial zero vector, must have at least $2K+1$ non-zero elements in order to distinguish between measurements resulting from two different K-sparse signals and to ensure the implication $\boldsymbol{c} \neq \boldsymbol{c}' \Rightarrow \boldsymbol{\mathcal{A}}\boldsymbol{c} \neq \boldsymbol{\mathcal{A}}\boldsymbol{c}'$ for any $\boldsymbol{c}, \boldsymbol{c}' \in \mathcal{C}_K$. Note that the difference $\boldsymbol{c} - \boldsymbol{c}' \in \mathcal{C}_{2K}$ is maximally $2K$-sparse.

To identify a unique signal vector solving (23), an extensive combinatorial search over all possible $\binom{U}{K}$ sparse subsets of $\boldsymbol{c}$ is required, which constitutes an NP-hard problem [28]. However, under stronger conditions on the null space of $\boldsymbol{\mathcal{A}}$, the relaxation into an ℓ_1-norm minimization problem is feasible, such as basis pursuit [29], [30], LASSO [31], or Dantzig selector [32]. For instance, the optimization problem in terms of basis pursuit is set up by simply replacing the pseudonorm in (23) with the ℓ_1-norm according to

$$\underset{\boldsymbol{c}\in\mathbb{C}^U}{\arg\min}\ \|\boldsymbol{c}\|_{\ell_1} \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{\mathcal{A}}\boldsymbol{c}. \tag{25}$$

Such ℓ_1-minimization problems can be easily cast into linear programs or second-order cone programs and thus can be solved via interior point methods [33].

The equivalence of the solution for the ℓ_1-problem with the unique K-sparse solution for the ℓ_0-problem is guaranteed for a measurement process where $\boldsymbol{\mathcal{A}}$ is built in such a way that the energy of any signal vector $\boldsymbol{v}$ living in the null space of $\boldsymbol{\mathcal{A}}$ is sufficiently spread over multiple coefficients. This requirement is formalized by the so-called null space property (NSP) of order K, defined as

$$\|\boldsymbol{v}\|_{\ell_1} > 2\,\|\boldsymbol{v}_{\mathcal{I}}\|_{\ell_1}, \forall\, \boldsymbol{v} \in \mathrm{kern}(\boldsymbol{\mathcal{A}})\backslash\{0\}, \forall\, |\mathcal{I}| \leq K, \tag{26}$$

where $\mathcal{I}$ is a set of indices restricting $\boldsymbol{v}$ to the corresponding vector elements. If $\boldsymbol{\mathcal{A}}$ satisfies (26), then there exists a K-sparse signal vector $\boldsymbol{c} \in \mathcal{C}_K$ uniquely solving (23) and (25), since, for any other other signal vector $\boldsymbol{c}' \in \mathbb{C}^U$ with $\boldsymbol{\mathcal{A}}\boldsymbol{c}' = \boldsymbol{\mathcal{A}}\boldsymbol{c}$, (26) guarantees that $\|\boldsymbol{c}'\|_{\ell_1} > \|\boldsymbol{c}\|_{\ell_1}$ [34], [35].

Evaluating the spark or the NSP of a measurement matrix $\boldsymbol{\mathcal{A}}$ has combinatorial computational complexity and is as hard as the ℓ_0-problem (23) itself [36]. For practical applications, a more tractable property of $\boldsymbol{\mathcal{A}}$ is the coherence [16]

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{1\leq u\neq v\leq U} \frac{|\langle \boldsymbol{a}_u^c, \boldsymbol{a}_v^c\rangle|}{\|\boldsymbol{a}_u^c\|_{\ell_2}\|\boldsymbol{a}_v^c\|_{\ell_2}} \in \left[\sqrt{\frac{U-M}{M(U-1)}}, 1\right], \tag{27}$$

where $\boldsymbol{a}_u^c$ denotes the u-th column of $\boldsymbol{\mathcal{A}}$. The theoretical guarantees for unique ℓ_0-recovery and ℓ_1-recovery improve for a smaller coherence [27], [37].

B. Stable and Robust Recovery

In practice, the samples taken by moving microphones are biased by certain error sources [13], and the 4D sound-field spectrum is not exactly K-sparse, e.g., due to evanescent waves [11]. Hence, with respect to (25), we are rather interested in solving the ℓ_1-problem with quadratic constraints according to

$$\underset{\boldsymbol{c}\in\mathbb{C}^U}{\arg\min}\ \|\boldsymbol{c}\|_{\ell_1} \quad \text{s.t.} \quad \|\boldsymbol{\mathcal{A}}\boldsymbol{c} - \boldsymbol{x}\|_{\ell_2} \leq \epsilon \tag{28}$$

with ϵ being the upper magnitude bound of the residual.

Signals that are well approximated by K-sparse representations via hard thresholding of the absolute values are called compressible [35]. In order to guarantee stable and robust CS based recovery of compressible signal vectors $\boldsymbol{c}$, a stronger condition than the NSP is necessary. Any subset of $2K$ columns of $\boldsymbol{\mathcal{A}}$ must behave like a nearly orthogonal transform that almost preserves the signal energy. In other words, $\boldsymbol{\mathcal{A}}$ is required to map any $2K$-sparse vector nearly isometrically from signal space into measurement space. The sampling matrix, with ℓ_2-normalized columns, is said to have the restricted isometry property (RIP) of order $2K$ and level $\delta \in (0,1)$ if

$$(1-\delta)\,\|\boldsymbol{c}\|_{\ell_2}^2 \leq \|\boldsymbol{\mathcal{A}}\boldsymbol{c}\|_{\ell_2}^2 \leq (1+\delta)\,\|\boldsymbol{c}\|_{\ell_2}^2,\ \forall\, \boldsymbol{c} \in \mathcal{C}_{2K}. \tag{29}$$

On the one hand, this property guarantees a unique solution via ℓ_0-minimization [38]. On the other hand, a small amount of additive noise in measurement space with energy $\|\boldsymbol{\mathcal{A}}\boldsymbol{c} - \boldsymbol{\mathcal{A}}\boldsymbol{c}'\|_{\ell_2}^2$ for $\boldsymbol{c}, \boldsymbol{c}' \in \mathcal{C}_K$, leads to an error energy $\|\boldsymbol{c} - \boldsymbol{c}'\|_{\ell_2}^2$ on the recovered signal that is not arbitrarily large for a system matrix satisfying (29). Moreover, the RIP ensures robustness in terms of multiplicative noise caused by a mismatch of $\boldsymbol{\mathcal{A}}$ [39], in our case, due to inaccuracies during the positional tracking of the microphones.

The smallest δ for which $\boldsymbol{\mathcal{A}}$ satisfies (29) is called the RIP constant δ_{2K}. A certain upper limit of δ_{2K} guarantees the unique distinction between measurements obtained from two different K-sparse signal vectors when solving the ℓ_1-problem [38], [40], i.e., the RIP implies the NSP. Furthermore, the RIP constant determines upper bounds for the recovery error. It controls the impact of perturbations induced by noise and by the K-sparse signal approximation [38], [41].

Verifying the RIP of a matrix is a combinatorial NP-hard problem [36]. Nevertheless, it was shown in previous works that measurement matrices generated by identical and independent distributions, such as Gaussian, Subgaussian and Bernoulli [42]–[44], random Fourier ensembles [45], [46], and random convolution ensembles [47], [48] satisfy the RIP with exponentially high probability for a wide range of $M < U$. Obviously, a realistic trajectory cannot be totally random, since the current position of the microphone is highly dependent on its previous position: the speed of the microphone is limited, so, usually, it is impossible to reach any location inside the volume of interest instantly. For non-random sampling, the coherence property (27) may be used as an indicator for RIP guarantees. The coherence of $\boldsymbol{\mathcal{A}}$ directly affects the upper bound for its RIP constant [37].

Dropping the idea of uniform recovery guarantees, weaker conditions that cope without the RIP suffice for stable and robust recovery [49]. A uniform recovery guarantee in terms of CS means that, for a fixed instance of $\boldsymbol{\mathcal{A}}$, all (approximately) K-sparse signals can be recovered with high probability. In contrast, for a fixed compressible signal and a random choice of the measurement matrix, guarantees for the so-called non-uniform recovery of the specific signal can be provided. Thus, uniform recovery implies non-uniform recovery [46].

C. Dynamic and Compressive Sound-Field Sampling

The CS based recovery of $h(\boldsymbol{g}, n)$ can be divided into two central problems, since the samples $x(\boldsymbol{r}(n), n)$ taken by the moving microphone are regular in time domain, but generally non-equidistant in space domain.

First, let us simplify the aggregated sampling problem by constraining the dynamic microphones to perform measurements exactly on the points of the modeled spatial grid, i.e., $\boldsymbol{r}(n) \in \mathcal{G}$. Then, it is obvious that random trajectories translate to random time sampling of each grid RIR. Thus, the measurement with moving microphones becomes equivalent to the measurement with equidistant stationary microphones, each of them neglecting particular time points. The overall sound field $h(\boldsymbol{g}, n)$ can be recovered by separately solving the N linear systems of equations

$$\boldsymbol{R}_u \boldsymbol{x} = \boldsymbol{R}_u \boldsymbol{S} \boldsymbol{h}_u, \tag{30}$$

where the binary $M_u \times M$ matrix $\boldsymbol{R}_u$ restricts $\boldsymbol{x}$ and $\boldsymbol{S}$ to the actual M_u rows that were randomly measured by the moving microphones for the particular grid position. Assume that each grid RIR is K_u-sparse (e.g., because it is dominated by a few direct reflections). Provided that the magnitudes of the excitation sequence $s(n)$ are similar and that its spectrum is flat, sampling a convolution product at randomly selected time points as in (30) is an isotropic and incoherent process guaranteeing non-uniform CS recovery by

$$\underset{\boldsymbol{h}_u \in \mathbb{R}^L}{\arg\min} \ \|\boldsymbol{h}_u\|_{\ell_1} \quad \text{s.t.} \quad \|\boldsymbol{R}_u \boldsymbol{S} \boldsymbol{h}_u - \boldsymbol{R}_u \boldsymbol{x}\|_{\ell_2} \leq \epsilon \tag{31}$$

for a wide range of $K_u \ll M_u < L$ [49]. Moreover, for $\boldsymbol{S}$ being a circular convolution matrix, also uniform recovery was shown in [47], [48] by proving low RIP constants with high probability. For circulant matrices constructed by independent, identically distributed variables that follow the Gaussian or Rademacher distribution, even any arbitrary (non-random) subset of M_u rows in (30) suffices for uniform recovery [48]. Therefore, regarding the basic model (11) of the dynamic sampling procedure, the excitation with white Gaussian noise is suitable for the CS based recovery. Note that linear convolution can be extended to the circular convolution case [47]. For the time-decoupled problem (14), which inherently involves circular convolution matrices, we propose the use of pseudo-random MLS for excitation: they are spectrally flat (except for DC) and closely follow a Rademacher process where the values $+1$ and -1 appear with almost equal probability [9]. Thus, the energy of MLS is maximally spread over the time signal. In contrast, tertiary Ipatov sequences [8], which also have a perfect autocorrelation function and allow for time-decoupling, are disadvantageous for the CS based recovery that exploits sparsity in time domain. They contain a certain number of zeros leading to a signal where the energy is concentrated on a smaller set of time points (cf. [47]).

Indeed, in practice, the moving microphone will be mainly located at intermediate positions of the spatial grid, so samples are taken in continuous space. For random sampling, the procedure is isotropic and incoherent and provides stability and robustness for non-uniform recovery [49]. Regarding (28) and considering the sparse spectral cone of sound fields, a randomly moving microphone measures a sparse multivariate trigonometric polynomial at random positions in space. Note that bandlimited signals are represented by trigonometric polynomials with a specific maximum degree. Exact CS recovery of a sparse polynomial from its random samples by basis pursuit was shown in [50]. More generally, in [46], uniform and non-uniform recovery guarantees were presented for signals that have a sparse expansion in a bounded orthonormal system of functions.

The time-decoupled problem (14) generically allows for decomposing the overall issue of spatio-temporal sound-field sampling into multiple spatial problems. By incorporating the sparsifying frequency transform and using (15), (16), and (22), the CS based optimization problem can be formulated as

$$\underset{\boldsymbol{c} \in \mathbb{C}^U}{\arg\min} \ \|\boldsymbol{c}\|_{\ell_1} \quad \text{s.t.} \quad \|\tilde{\boldsymbol{\mathcal{A}}} \boldsymbol{c} - \boldsymbol{x}\|_{\ell_2} \leq \epsilon \tag{32}$$

with $\tilde{\boldsymbol{\mathcal{A}}} = \tilde{\boldsymbol{A}}(\boldsymbol{I}_N \otimes \boldsymbol{S}_o)\boldsymbol{\Psi}^H$. Obviously, this reformulation involving frequency representation no longer allows for decomposing and solving multiple subproblems separately. However, the block diagonal structure of $\tilde{\boldsymbol{A}}$ and the circularity of $\boldsymbol{S}_o$ may heavily reduce the number of calculations performed by CS recovery algorithms.

In the following section, the influence of the trajectory on the CS problem is analyzed. At this, we will see that the sparse representation via $\boldsymbol{\Psi}$ performing discrete Fourier transforms possess many advantages. On the one hand, the structure and the properties of the CS matrices can be easily determined. On the other hand, the choice of the spectrally flat excitation signal becomes irrelevant for the coherence of the resulting CS problem. As mentioned above, when exploiting the sparsity of RIRs in time domain by use of $\boldsymbol{\Psi} = \boldsymbol{I}_U$, the temporal energy distribution of the excitation sequence is crucial for the CS problem.

V. Trajectory Analysis for Spectrally Flat Excitation

In Section V-A the general structure of the CS matrix $\boldsymbol{\mathcal{A}}$ is characterized. Also the time-decoupled problem involving $\tilde{\boldsymbol{\mathcal{A}}}$ is considered, where $s(n)$ is a periodic sequence with perfect autocorrelation function as described in Section II-C. The dependency of the CS matrix on the trajectory is investigated in Section V-B, given a spectrally flat excitation sequence $s(n)$, assuming a perfect interpolator, and applying discrete Fourier transforms for the sparse frequency representation of $\boldsymbol{h}$. Based on this, in Section V-C we deduce an expression that reduces the computational cost for calculating the coherence (27) of the given CS problem from a quadratic problem in the number of sound-field variables to a linear problem. The coherence of the CS matrix only depends on the trajectory relative to the spatial grid to be recovered. Exploiting this, the requirement on the microphone trajectory considered optimal for CS sampling is formalized. This enables us to evaluate given dynamic measurements at low computational cost.

To keep the initial description and analysis simple, in Sections V-A and V-B the measurement space will be confined to the x-dimension only. This means that we consider the recovery of RIRs at uniform points along a line in the x-direction. The extension to spatial 3D sampling is straightforward, since both the interpolation function and the sparsifying transform are assumed to be separable on the virtual grid in space.

A. General Structure of the CS Matrix

Let $\boldsymbol{\Psi} = \boldsymbol{T}_X \otimes \boldsymbol{T}_L$ perform some orthonormal 2D transform along both discrete variables of the sought impulse response $h(g_x, n)$ contained in $\boldsymbol{h}$. Correspondingly, $\boldsymbol{c} \in \mathbb{C}^{XL}$ comprises the concatenated values of the 2D transformed sound-field spectrum. Following (21), the unknown parameters are projected onto measurement space according to

$$\boldsymbol{\phi}_m^H \boldsymbol{\Psi}^H \boldsymbol{\Psi} \boldsymbol{h} = \langle \boldsymbol{\Psi} \boldsymbol{h}, \boldsymbol{\Psi} \boldsymbol{\phi}_m \rangle = \langle \boldsymbol{c}, \check{\boldsymbol{\phi}}_m \rangle. \tag{33}$$

Hence, the new representation by $\boldsymbol{c}$ leads to the CS matrix $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M \times XL}$ whose m-th row $\boldsymbol{a}_m^r = \check{\boldsymbol{\phi}}_m^H$ contains the 2D transformed values of the components in the m-th row of $\boldsymbol{A}$. Corresponding to (10) and (12), the m-th row of $\boldsymbol{A}$ is built up by the sampling vector

$$\boldsymbol{\phi}_m^H = [\varphi_{m-1}(0)\, \boldsymbol{s}(m), \ldots, \varphi_{m-1}(X-1)\, \boldsymbol{s}(m)], \tag{34}$$

with

$$\boldsymbol{s}(m) = [s(m-1), s(m-2), \ldots, s(m-L)]. \tag{35}$$

Thus, one row in $\boldsymbol{A}$ is formed by the corresponding sequence of the time reversed and spatially weighted source signal

$$s_m(g_x, n) = \varphi_{m-1}(g_x)\, s(m-1-n). \tag{36}$$

Accordingly, one row in $\boldsymbol{\mathcal{A}}$ contains the values of the 2D orthonormal transform of (36).

The CS matrix $\tilde{\boldsymbol{\mathcal{A}}}$ of the time-decoupled problem essentially possess the same structure as $\boldsymbol{\mathcal{A}}$. In fact, the repetitive excitation leads to a basic sampling matrix $\boldsymbol{A}$ that is equivalent to $\tilde{\boldsymbol{A}}(\boldsymbol{I}_N \otimes \boldsymbol{S}_o)$. However, due to the periodicity in time, the temporal component in (34) is identical for every L-th row.

For $\boldsymbol{\Psi}$ performing the 2D discrete Fourier transform on the sound field, the m-th row $\boldsymbol{a}_m^r$ building $\boldsymbol{\mathcal{A}}$ is composed of the Fourier representation

$$\mathcal{S}_m(k_x, l) = \frac{1}{\sqrt{XL}} \sum_{g_x=0}^{X-1} \sum_{n=0}^{L-1} s_m(g_x, n)\, e^{-2\pi j \frac{l}{L} n} e^{-2\pi j \frac{k_x}{X} g_x}, \tag{37}$$

where $k_x \in \{-\frac{X-1}{2}, \ldots, \frac{X-1}{2}\}$ and $l \in \{-\frac{L-1}{2}, \ldots, \frac{L-1}{2}\}$ are the sampled frequency variables for the space and time dimension, respectively. For simplicity and without loss of generality, both the grid length X and the RIR length L are assumed to be odd.

B. Influence of the Trajectory on the CS Matrix

From the point of view of signal processing, the coefficients given by the interpolation function act like a digital filter in the spatial domain that depends on the microphone position relative to the modeled grid. Accordingly, the rows of $\boldsymbol{\mathcal{A}}$ and $\tilde{\boldsymbol{\mathcal{A}}}$, are primarily determined by the trajectory $r_x(n)$ for spectrally flat excitation. This issue is described in the following subsection.

Regarding (5), we can consider $\varphi_n(g_x)$ to fulfill a spatial alignment task, i.e., to perform a fractional delay (FD) in space on the sound field $h(g_x, n)$, in order to fit samples $x(r_x(n), n)$ taken in continuous space into the modeled spatial grid. The impulse response of an ideal FD filter is a shifted and sampled sinc function, $\varphi_n^{\text{id}}(g_x) = \text{sinc}(g_x - D_x(n))$, where

$$D_x(n) = \frac{r_x(n) - r_0}{\Delta_x} \tag{38}$$

is the delay consisting of the integer part $\lfloor D_x(n) \rfloor$ and the fractional part $d_x(n) = D_x(n) - \lfloor D_x(n) \rfloor$. Thus, the ideal frequency response of a FD filter reads

$$\Phi_n^{\text{id}}(e^{j\kappa_x}) = e^{-jD_x(n)\kappa_x}, \tag{39}$$

with constant magnitude response

$$|\Phi_n^{\text{id}}(e^{j\kappa_x})| = 1, \tag{40}$$

linear phase response

$$\arg\left\{\Phi_n^{\text{id}}(e^{j\kappa_x})\right\} = \theta_n^{\text{id}}(e^{j\kappa_x}) = -D_x(n)\kappa_x, \tag{41}$$

and constant phase delay $\tau_n^{\text{id}} = D_x(n)$. For $D_x(n) \neq 0$, the ideal FD filter has infinite length, and, thus, is not realizable.

In order to design a realizable FD filter, several finite-length approximations for the sinc function have been proposed. A discussion and comparison of common techniques including non-recursive (FIR) and recursive (IIR) filter approximations can be found in [51]. The maximally flat FD FIR filter approximation of length $P+1$ is equivalent to the coefficients of the classical Lagrange interpolation method, given by the P-th order polynomial

$$\varphi_n(g_x) = \prod_{\substack{p=0 \\ p \neq g_x}}^{P} \frac{D_x(n) - p}{g_x - p}. \tag{42}$$

In [13], [22], an odd-order Lagrange interpolator was used for both spatial measurement interpolation and spatial subband synthesis (cf. Section II-D), respectively, given a well-posed problem. The maximum order of the interpolating polynomial was limited by confining the support of the Lagrange kernel to local grid points. The support was centered around the measurement position $r_x(n)$, since the odd-order Lagrange interpolator yields the best approximation for $\frac{P-1}{2} \leq D_x(n) \leq \frac{P+1}{2}$ [51]. Then, it never overestimates the amplitude of the signal, i.e., the maximum of the magnitude response of the FD filter approximation is one. The approximation error highly depends on the fractional part $d_x(n)$. The worst case occurs with a fractional delay of $d_x(n) = 0.5$, which leads to an excessive magnitude error at high frequencies and even to an exact zero at the Nyquist frequency. However, at low frequencies, the magnitude and phase delay curves coincide with the ideal response for any $d_x(n)$. Thus, for a virtual grid with spacing Δ_x leading to more than a twofold spatial oversampling, the Lagrange interpolation shows nearly optimal performance (cf. [13]). The magnitude and phase delay responses of Lagrange interpolators

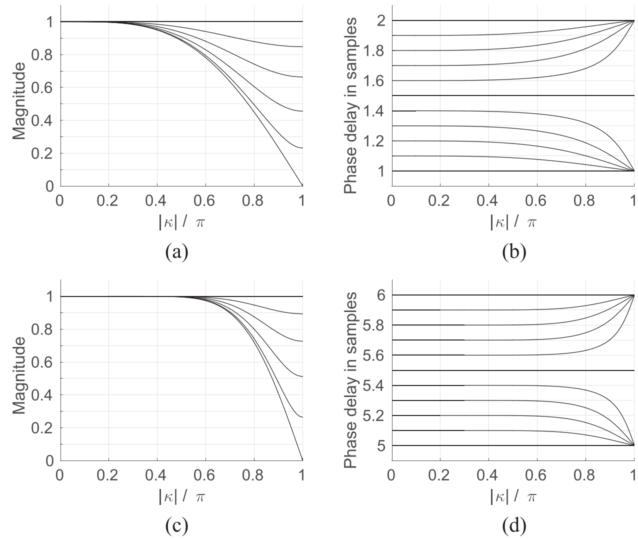


Fig. 2. Magnitude and phase delay responses of Lagrange FD filters $\Phi(e^{j\kappa})$ of order $P = 3$ (upper row) and order $P = 11$ (lower row). The curves are plotted for eleven equidistant fractional delay values D in the interval $[\frac{N-1}{2}, \frac{N+1}{2}]$. Note that in (a), (c) there are only six curves, since the magnitude responses for fractional delays D and $P - D$ are the same.

with $P = 3$ and $P = 11$ are illustrated in Fig. 2 for various fractional delay values.

In addition to the actual interpolation accuracy, $\varphi_n(g_x)$ and $r_x(n)$ directly influence the appearance of the CS matrix, and, therefore, the guarantees for the CS based recovery. Let the columns in $\boldsymbol{\mathcal{A}}$ consist of values of the Fourier spectra $\mathcal{S}_m(k_x, l)$ defined in (37):

$$\boldsymbol{a}^c_{(k_x,l)} = [\mathcal{S}_1(k_x, l), \ldots, \mathcal{S}_M(k_x, l)]^T . \tag{43}$$

Using (36), (37), (38), and (39), it can be seen that the change of the microphone position from measuring point $r_x(n)$ to point $r_x(n+m)$ ideally corresponds to recursive phase shifts in the discrete Fourier spectrum given by

$$\mathcal{S}_{n+m}(k_x, l) = \mathcal{S}_n(k_x, l)\, e^{-2\pi j (D_x(m) - D_x(n)) \frac{k_x}{X}}\, e^{-2\pi j m \frac{l}{L}} . \tag{44}$$

Thus, for a spectrally flat excitation sequence and a perfect interpolation kernel fulfilling (39) for any frequency, each of the XL columns of $\boldsymbol{\mathcal{A}}$ can be described by use of structured spatial and temporal phase terms according to

$$\boldsymbol{a}^c_{(k_x,l)} = \begin{bmatrix} a^0_{(k_x,l)} \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(1) - D_x(0)) \frac{k_x}{X}}\, e^{-2\pi j 1 \frac{l}{L}} \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(2) - D_x(0)) \frac{k_x}{X}}\, e^{-2\pi j 2 \frac{l}{L}} \\ \vdots \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(M-1) - D_x(0)) \frac{k_x}{X}}\, e^{-2\pi j (M-1) \frac{l}{L}} \end{bmatrix}, \tag{45}$$

where the initial phase state

$$a^0_{(k_x,l)} = e^{-2\pi j D_x(0) \frac{k_x}{X}}\, \sigma_s\, e^{j\theta_0(l)} \tag{46}$$

is determined by the initial grid delay $D_x(0)$ and by the initial phase $\theta_0(l)$ of the excitation signal with power σ_s^2 leading to the first microphone sample at time $n = 0$. Consequently, all columns of $\boldsymbol{\mathcal{A}}$ possess consistent ℓ_2-norm, $\|\boldsymbol{a}^c_{(k_x,l)}\|_{\ell_2} = \sqrt{M\sigma_s^2}$. The rows of $\boldsymbol{\mathcal{A}}$ only differ in the phases: the temporal domain fulfills a uniform phase delay, and in the spatial domain, in general, a fractional phase shift is performed depending on the trajectory $r_x(n)$ in accordance with (38). Note that (45) also holds for the time-decoupled model with CS matrix $\tilde{\boldsymbol{\mathcal{A}}}$. Here, due to the L-periodic excitation sequence, the temporal phase terms appear repetitively at rows $n + mL$ for fixed n and varying m.

C. Coherence of Measurements

From the column representation (45), the deduction of the coherence (27) belonging to the particular CS matrix is straightforward. Defining the differences of discrete frequency variables $k'_x, k''_x \in \{-\frac{X-1}{2}, \ldots, \frac{X-1}{2}\}$ and $l', l'' \in \{-\frac{L-1}{2}, \ldots, \frac{L-1}{2}\}$ as

$$\Delta k_x = k'_x - k''_x,\ \Delta k_x \in \{-(X-1), \ldots, X-1\}, \tag{47}$$

$$\Delta l = l' - l'',\ \Delta l \in \{-(L-1), \ldots, L-1\}, \tag{48}$$

the coherence of $\boldsymbol{\mathcal{A}}$ is

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{(k'_x,l') \neq (k''_x,l'')} \frac{\left| \left\langle \boldsymbol{a}^c_{(k'_x,l')}, \boldsymbol{a}^c_{(k''_x,l'')} \right\rangle \right|}{\left\| \boldsymbol{a}^c_{(k'_x,l')} \right\|_{\ell_2} \left\| \boldsymbol{a}^c_{(k''_x,l'')} \right\|_{\ell_2}} \tag{49}$$

$$= \max_{(\Delta k_x, \Delta l) \neq (0,0)} \frac{1}{M} \left| \sum_{n=0}^{M-1} e^{-2\pi j \frac{D_x(n)}{X} \Delta k_x}\, e^{-2\pi j \frac{n}{L} \Delta l} \right|, \tag{50}$$

where the initializing temporal phase $e^{j(\theta_0(l') - \theta_0(l''))}$ resulting from the scalar product in (49) is independent of the sum over n, and, thus, dissolves in (50). Considering a 3D grid in space, let us define, by analogy with (38) and (47), the trajectory

$$\boldsymbol{r}_D(n) = [D_x(n), D_y(n), D_z(n)]^T \tag{51}$$

relative to the modeled grid coordinate system, and the vector

$$\boldsymbol{d} = [\Delta k_x, \Delta k_y, \Delta k_z]^T \tag{52}$$

containing the frequency difference for each spatial dimension. Using

$$\mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d}) = e^{-2\pi j \left(\frac{D_x(n)}{X} \Delta k_x + \frac{D_y(n)}{Y} \Delta k_y + \frac{D_z(n)}{Z} \Delta k_z \right)}, \tag{53}$$

the coherence of the modeled sound-field sampling problem reads

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{(\boldsymbol{d}, \Delta l)} \frac{1}{M} \left| \sum_{n=0}^{M-1} \mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d})\, e^{-2\pi j \frac{n}{L} \Delta l} \right| \tag{54}$$

with $(\boldsymbol{d}, \Delta l) \neq (\boldsymbol{0}_{3\times 1}, 0)$. The following observations can be made by reference to the expression (54) for an ideal interpolation kernel:

- Any arbitrary spectrally flat excitation signal $s(n)$ leads to one and the same coherence of the CS problem, independent of its spectral phases.

- The coherence of the CS problem only depends on $\boldsymbol{r}_D(n)$, which is the microphone trajectory relative to the virtual grid in space to be recovered.
- The periodic excitation considered in the time-decoupled model promotes a high correlation of samples taken for equal phases of excitation at similar positions (cf. Section III), and, thus, may easily lead to a higher coherence of $\tilde{\boldsymbol{\mathcal{A}}}$ than given by (54).
- For determining the coherence of $\boldsymbol{\mathcal{A}}$, only varying tuples of frequency differences $(\boldsymbol{d}, \Delta l)$ have to be regarded. Compared to the naive approach of testing any scalar product between two different columns in $\boldsymbol{\mathcal{A}}$, this knowledge reduces the effort for computing the coherence of a given CS matrix from a quadratic problem in $\mathcal{O}(U^2)$ to a linear problem in $\mathcal{O}(U)$, where $U = XYZL$ is the number of unknown parameters.
- Since the rows of $\boldsymbol{A}$ are represented by real-valued signals in practice, the particular spectra in $\boldsymbol{\mathcal{A}}$ are conjugate symmetric. This may be exploited for saving further computational cost. For example, the maximum in (54) could be found by regarding only parameter combinations $(\boldsymbol{d}, \Delta l)$ with $\Delta l \in \{0, \ldots, L-1\}$, $\Delta k_x \in \{0, \ldots, X-1\}$, $\Delta k_y \in \{-(Y-1), \ldots, Y-1\}$, and $\Delta k_z \in \{-(Z-1), \ldots, Z-1\}$.

As mentioned in Sections IV-A and IV-B, the theoretical guarantees for CS based recovery improve for lower coherence. The RIP requires that the columns of $\boldsymbol{\mathcal{A}}$ consisting of measured phases according to (45) must build up a nearly orthogonal system for any set of $2K$ different frequency tuples (k_x, k_y, k_z, l). If the coherence is small, then all columns of $\boldsymbol{\mathcal{A}}$ are almost mutually orthogonal, thus, (54) provides an upper bound for the RIP constant of $\boldsymbol{\mathcal{A}}$. An even sharper bound may be achieved by exploiting (54) for calculating the ℓ_1-coherence function [52], which measures the maximum total coherence between a fixed column and a set of $2K$ other columns. In this context, the CS trajectory may be considered optimal for the set of parameters $\mathcal{P} = \{(\boldsymbol{r}_D(n), \mathcal{G})\}$ ensuring

$$\underset{\mathcal{P}}{\arg\min} \left(\max_{(\boldsymbol{d}, \Delta l)} \left| \sum_{n=0}^{M-1} \mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d})\, e^{-2\pi j \frac{n \Delta l}{L}} \right| \right). \tag{55}$$

The expression (55) may be used for several tasks that arise in practice. In addition to finding optimal trajectories for sought grids or modeling optimal grids for given measurements, the formula may be exploited for an efficient reconstruction strategy where the bandwidth is spatially adapted according to the number of neighbouring samples and their coherence.

Indeed, a perfect interpolator with ideal frequency response (39) is unrealizable in practice. Nevertheless, considering a virtual grid leading to spatial oversampling by a factor $\alpha > 2$, the FD filter approximation by an appropriate Lagrange interpolator achieves ideal magnitude and phase response for the relevant range of these normalized frequencies $|\kappa_x| < \frac{\pi}{\alpha}$ at which the bandlimited signal actually lives (cf. Fig. 2(c) and (d)). Accordingly, the non-ideal columns in $\boldsymbol{\mathcal{A}}$, that refer to sampled spatial frequencies where the FD filter is not perfect according to (39), correspond to entries in $\boldsymbol{c}$ being for sure (approximately) zero.

Thus, all columns of the CS matrix that potentially span the measurement space may be described by (45) and their coherence is given by (54).

VI. Efficient CS Algorithm for RIR Recovery

In the experimental part of this paper, we adapt the IHT algorithm for CS based sound-field recovery. The IHT method provides near-optimal and robust error guarantees for solving CS problems and requires low effort in computation and memory [23]. It approaches the K-sparse solution by applying the simple update rule

$$\hat{\boldsymbol{c}}^{(i+1)} = T_K \left\{ \hat{\boldsymbol{c}}^{(i)} + \mu\, \boldsymbol{\Psi} \boldsymbol{A}^T \left(\boldsymbol{x} - \boldsymbol{A} \boldsymbol{\Psi}^H \hat{\boldsymbol{c}}^{(i)} \right) \right\}, \tag{56}$$

where T_K is the nonlinear thresholding operator that sets all but the K largest absolute values in the updated signal to zero. One iteration according to (56) involves a simple gradient descent step into the direction of the least-squares solution with step size μ followed by a hard projection of the signal estimate onto the subspace of its K-sparse representation. Applied to our sound-field sampling model (11), (56) simplifies to a fast update scheme involving the calculation of the residual according to

$$\boldsymbol{\varepsilon}^{(i)} = \boldsymbol{x} - \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S} \hat{\boldsymbol{h}}_u^{(i)}, \tag{57}$$

the update of the estimated grid RIRs subject to

$$\hat{\boldsymbol{h}}_u^{(i+1)} = \hat{\boldsymbol{h}}_u^{(i)} + \mu\, \boldsymbol{S}^T \boldsymbol{\Phi}_u \boldsymbol{\varepsilon}^{(i)} \tag{58}$$

and the transform of the updated sound-field variables $\hat{\boldsymbol{h}}^{(i+1)}$ into their sparse frequency representation $\hat{\boldsymbol{c}}^{(i+1)} = \boldsymbol{\Psi} \hat{\boldsymbol{h}}^{(i+1)}$ for hard thresholding. In (57) and (58), the convolutions of the source signal with the estimated RIRs and the residual can be computed very efficiently in Fourier domain. The diagonal matrices $\boldsymbol{\Phi}_u$ only contribute weighting factors subject to the spatial interpolation.

Considering the time-decoupled problem (14), the IHT update rule becomes

$$\hat{\boldsymbol{c}}^{(i+1)} = T_K \left\{ \hat{\boldsymbol{c}}^{(i)} + \mu\, \boldsymbol{\Psi} \boldsymbol{S}_o^T \tilde{\boldsymbol{A}}^T \left(\boldsymbol{x} - \tilde{\boldsymbol{A}} \boldsymbol{S}_o \boldsymbol{\Psi}^H \hat{\boldsymbol{c}}^{(i)} \right) \right\}. \tag{59}$$

Accordingly, the computation of the residual reads

$$\boldsymbol{\varepsilon}^{(i)} = \begin{bmatrix} \boldsymbol{\varepsilon}_1^{(i)} \\ \vdots \\ \boldsymbol{\varepsilon}_R^{(i)} \end{bmatrix} = \boldsymbol{x} - \sum_{u=1}^{N} \begin{bmatrix} \boldsymbol{\Phi}_{u,1} \\ \vdots \\ \boldsymbol{\Phi}_{u,R} \end{bmatrix} \boldsymbol{S}_o\, \hat{\boldsymbol{h}}_u^{(i)}, \tag{60}$$

and the RIR correction step is

$$\hat{\boldsymbol{h}}_u^{(i+1)} = \hat{\boldsymbol{h}}_u^{(i)} + \mu\, \boldsymbol{S}_o^T \sum_{r=1}^{R} \boldsymbol{\Phi}_{u,r}\, \boldsymbol{\varepsilon}_r^{(i)}, \tag{61}$$

with the partial residual signal in $\boldsymbol{\varepsilon}_r^{(i)}$ corresponding to the r-th period of excitation. By comparison with $\boldsymbol{S}$, the matrix $\boldsymbol{S}_o$ is circular and only of size $L \times L$ instead of $M \times L$, which reduces the effort for computing the convolution products.

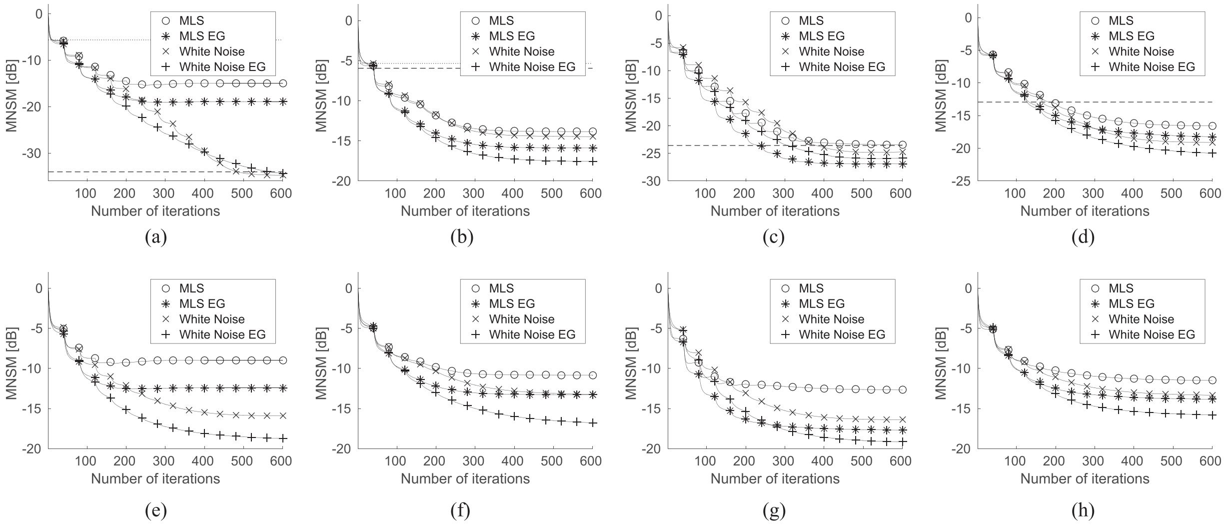


Fig. 3. Error of the CS based sound-field recovery as a function of the number of IHT iterations for the cases of (a)–(d) $M = 1.28\,\mathrm{U}$ samples (i.e., slight oversampling) and (e)–(h) $M = 0.8\,\mathrm{U}$ samples (i.e., slight undersampling), taken along four different microphone trajectories: (a), (e) sampling at random positions exactly on grid points, (b), (f) sampling at random positions exactly in the middle between grid points, (c), (g) sampling along a Lissajous trajectory, and (d), (h) sampling at random positions inside the region of interest. In the upper row ($M = 1.28\,\mathrm{U}$), the dashed lines indicate the quality given by the LS solutions of the basic model. The dotted lines in (a) and (b) show the results for the LS solutions and the time-decoupled model. The LS solution for the time-decoupled model and Lissajous and random sampling lies above 0 dB NSM and is not shown.

By decomposing the measurement signal of the moving microphone into V distinct subband signals $\boldsymbol{x}^{(v)}$, both update rules (56) and (59) may be easily incorporated into a multiresolution recovery scheme according to (19).

VII. Experiments and Results

For the following experiments, we simulated RIRs and microphone measurements by use of the image source method [53] considering a room of size 5.8 m × 4.15 m × 2.55 m. The reverberation time of the room was chosen as $RT_{60} = 0.3\,\mathrm{s}$. The cutoff frequency of the RIRs was $f_c = 4\,\mathrm{kHz}$. The position of the sound source was set to $[1.4, 1.6, 1.0]^T$ in a world coordinate system with unit 1 m. The origin of the spatial grid $\mathcal{G}$ was set to $\boldsymbol{r}_0 = [2.75, 1.4, 0.8]^T$. Measurements were taken by one moving microphone. For the recordings, white Gaussian measurement noise was added.

As error measure for the overall sound-field recovery involving N grid RIRs, the mean normalized system misalignment [13]

$$\mathrm{MNSM} = \frac{1}{N}\sum_{u=1}^{N} \frac{\|\boldsymbol{h}_u - \hat{\boldsymbol{h}}_u\|_{\ell_2}^2}{\|\boldsymbol{h}_u\|_{\ell_2}^2} \tag{62}$$

is used, with $\boldsymbol{h}_u \in \mathbb{R}^L$ containing the true RIR and $\hat{\boldsymbol{h}}_u \in \mathbb{R}^L$ being the reconstructed RIR at grid index u. In order to provide a frequency-dependent error measure for each estimated grid RIR, we define the mean energy spectral density of the error [22] as

$$\text{ME-ESD}(f) = \frac{1}{N}\sum_{u=1}^{N} |H_u(f) - \hat{H}_u(f)|^2, \tag{63}$$

where $H_u(f)$ and $\hat{H}_u(f)$ are the Fourier transforms of the true RIR and the corresponding reconstructed RIR, respectively.

The first experiment gives a comparison between the CS based sound-field recovery proposed in this paper and the strategy used in [13], where the sound-field estimate was the least-squares (LS) solution obtained by the pseudoinverse of $\boldsymbol{A}$. Accordingly, we consider a sampling scenario ensuring that the system matrix is small enough to allow for computing the pseudoinverse. We sampled the sound field on a planar grid of size 5×5 with spacings $\Delta_{x,y} = 0.02\,\mathrm{m}$ at fixed height 0.8 m. The RIRs were limited to length $L = 511$. In total, this problem involves $U = 12{,}775$ sound-field variables. We tested four different types of trajectories covering the entire modeled grid: $\boldsymbol{r}_{\mathrm{G}}(n)$ taking samples exactly on random grid points, $\boldsymbol{r}_{\mathrm{M}}(n)$ taking samples at random positions exactly in the middle between grid points, $\boldsymbol{r}_{\mathrm{R}}(n)$ taking samples at random positions, and $\boldsymbol{r}_{\mathrm{L}}(n)$ taking samples along a Lissajous trajectory with frequency ratio $f_x/f_y = 3/4$ [54]. Of course, the trajectories $\boldsymbol{r}_{\mathrm{G}}(n)$ and $\boldsymbol{r}_{\mathrm{M}}(n)$ are completely unrealistic in practice. They are just used to identify fundamental differences between individual recovery algorithms. For excitation, we used white Gaussian noise for the basic sampling model and a scaled MLS sequence of length L with zero DC offset for the time-decoupled model. The signal-to-noise ratio (SNR) for the measurement signal was set to 40 dB.

The sampling process was modeled with $\phi_n(\boldsymbol{g})$ being the Lagrange interpolator of order three in each spatial dimension. For the CS based recovery via IHT, the step size $\mu = 5{\cdot}10^{-4}$ was chosen. As sparse frequency representation of the sound-field variables we used their Fourier transforms. The iterative recovery was started with the zero vector. Beginning with $K_0 = 1000$, the sparsity constraint was successively relaxed every 40 iterations.

The recovery results for $M = 16{,}352 = 1.28\,\mathrm{U}$ samples and $M = 10{,}220 = 0.8\,\mathrm{U}$ samples, taken along the considered trajectories, are presented in Fig. 3. The markers on the curves indicate the iterations where the sparsity parameter K was increased by 1000. Regarding the potentially overdetermined case $M = 1.28\,\mathrm{U}$, the quality of the LS solutions is indicated with a dashed line for the basic sampling model, and with a dotted line for the time-decoupled model with periodic excitation. For the CS based recovery, we further tested the case in which the modeled grid is extended beyond the actual region of interest, denoted with EG, in order to avoid impairments of the interpolator for samples taken at the boundary of the considered plane. According to the order of the used interpolation filters, this strategy increases the size of the modeled grid to 7×7, which certainly leads to underdetermined variables even for $M = 16{,}352$. Also for the grid-extended case, the MNSM measure was calculated regarding the inner 5×5 sector that spans the actual region of interest.

Concerning the setup with $M = 1.28\,\mathrm{U}$, Fig. 3(a)–(d) show the following results:

- The basic model with random white Gaussian excitation yields overdetermined problems with proper LS solutions, whereas the time-decoupled model leads to underdetermined problems due to the temporal correlations of the periodic MLS excitation as described in Section III.
- The CS based recovery substantially reduces the drawback of periodic excitation.
- The trajectory $\boldsymbol{r}_{\mathrm{G}}(n)$ (Fig. 3(a)) performs best, and the trajectory $\boldsymbol{r}_{\mathrm{M}}(n)$ (Fig. 3(b)) performs worst for sound-field recovery, which corresponds to the analysis in [13].
- For non-optimal trajectories (Fig. 3(b)–(d)), the proposed CS approach significantly improves the sound-field estimate, even for the overdetermined case.

Reducing the number of sampling points by 38% leads to sub-Nyquist sampling with $M = 0.8\,\mathrm{U}$. The results for this setup are depicted in Fig. 3(e)–(h). Except for the optimal trajectory $\boldsymbol{r}_{\mathrm{G}}(n)$ involving no interpolation, the quality of the sound-field estimate obtained by CS reconstruction remains stable despite significantly reduced data.

The strategy of extending the modeled grid beyond the region of interest improves the CS based recovery inside the targeted area in most cases. Despite increasing the overall number of underdetermined sound-field variables, this strategy weakens the boundary problem that occurs when interpolating at the edges of the finite grid, and, thus, leads to a more consistent sampling model.

For the extended-grid scenario, the used interpolator of order three for determining the CS matrix $\boldsymbol{\mathcal{A}}$ possesses magnitude and phase delay responses according to Fig. 2(a) and (b). Obviously, the interpolation kernels are non-ideal for a wide range of higher frequencies. Nevertheless, we tested the expression (54) derived for the efficient calculation of the coherence under the assumption of ideal interpolators. For $M = 0.8\,\mathrm{U}$, we ran $10{,}000$ experiments where the sound field was recovered by the IHT algorithm using samples from randomly generated hybrids of the four trajectories described above. The samples were corrupted with Gaussian noise at 40 dB SNR. In Fig. 4, the resulting recovery errors and the corresponding values for the coherence estimated by the formula (54) are presented. We observe a high correlation between both quantities, even for the low-order FD filter considered in this scenario. In all of the experiments, the coherence according to (54) was found at $\Delta l = 0$, which reveals the trajectory-dependent function (53) as the only important factor for determining the coherence.

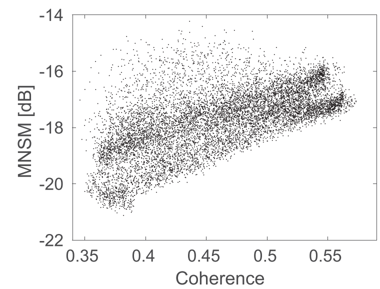


Fig. 4. Error of the CS based sound-field recovery versus the coherence calculated by use of the formula (54) for $10{,}000$ experiments.

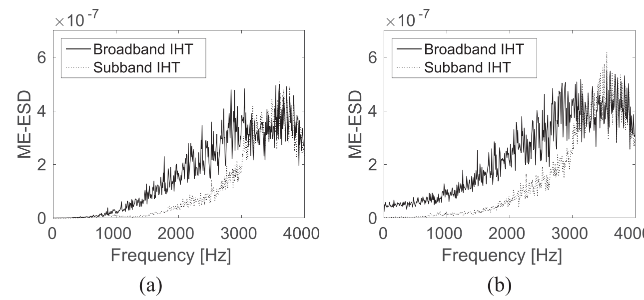


Fig. 5. Frequency-dependent error with and without involving a multiresolution scheme into CS recovery for (a) SNR $= 40$ dB and (b) SNR $= 20$ dB.

Further, both the time-decoupling and the multiresolution recovery scheme were incorporated into the IHT algorithm. For the same room scenario, a volume of interest covered by a $5 \times 5 \times 5$ grid with spacings $\Delta_{x,y,z} = 0.02\,\mathrm{m}$ was considered. The lengths of the RIRs and the MLS was set to $L = 1023$. The scenario involves $U = 1.27 \cdot 10^5$ sound-field variables to be recovered. The microphone took $M = 10^5$ samples along a 3D Lissajous trajectory with frequency ratios $f_x/f_y = 9/10$, $f_x/f_z = 9/8$.

Following the CBM scheme from [22], we modeled $V = 4$ resolution levels, each with temporal bandwidth of 1000 Hz. Hamming windowed bandpass filters of order 50 were used. The step size of the IHT algorithm optimizing on subbands was $\mu = 8 \cdot 10^{-5}$ and the initial sparsity parameters were $K_0 = 2000/(5 - v)$, constantly relaxed after each 50 iterations.

Compared to the wideband recovery, the CS multigrid approach significantly improves the sound-field recovery at low frequency bands for several SNRs. The frequency-dependent error after 600 iterations is presented in Fig. 5 for SNR $= 40$ dB and SNR $= 20$ dB.

VIII. Conclusion

In this paper, we have presented a CS framework for dynamic sound-field recovery. The method allows for uniquely determining RIRs on a virtual array in space by use of moving microphones performing sub-Nyquist sampling. Compared to conventional RIR measurements, the method not only saves time due to sub-Nyquist sampling, but also because no transient times are involved at all. The signals of the microphones, the tracking of their positions, and the knowledge of the source signal enable us to model the dynamic sampling process by means of a linear system of equations. However, in practice, especially for hand-guided microphone trajectories, the linear system will be underdetermined with high probability unless an excessive number of samples is acquired. In order to ensure stable and robust sound-field recovery even in the underdetermined case, a CS solution has been derived. The presented framework exploits the sparsity of sound fields in the frequency domain. By modeling an equidistant grid and using Fourier representations, the structure of the CS matrix and its dependency on the microphone trajectory have been shown. Based on that, a computationally efficient trajectory-dependent expression for the coherence of measurements given spectrally flat excitation has been derived. It allows for efficiently evaluating trajectories in terms of CS reconstruction. Following the IHT algorithm, we provided simple iterative update rules that incorporate two extensions of the dynamic sampling model and enable us to recover sound fields at low cost in computation and memory. Future works will be directed toward using the fast coherence computation to locally adapt the reconstruction bandwidth to the trajectory and to include sparsity measures for the early-reflection part.

References

[1] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[2] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[3] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, 1994.

[4] S. Spors, D. Seuberth, and R. Rabenstein, "Multiactuator panels for wave field synthesis: Evolution and present developments," *J. Audio Eng. Soc.*, vol. 58, no. 12, pp. 1045–1063, Dec. 2010.

[5] D. Talagala, W. Zhang, and T. Abhayapala, "Efficient multi-channel adaptive room compensation for spatial soundfield reproduction using a modal decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1522–1532, Oct. 2014.

[6] J. O. Jungmann, R. Mazur, and A. Mertins, "Joint time- and frequency-domain reshaping of room impulse responses," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 733–737.

[7] H.-D. Lüke, "Sequences and arrays with perfect periodic correlation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 3, pp. 287–294, May 1988.

[8] V. P. Ipatov, "Ternary sequences with ideal periodic autocorrelation properties," *Radio Eng. Electron. Phys.*, vol. 24, pp. 75–99, Oct. 1979.

[9] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, Jun. 1989.

[10] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Audio Eng. Soc. Conv.*, May 2007, pp. 1–21.

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

[12] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *J. Acoust. Soc. Amer.*, vol. 122, no. 3, pp. 1636–1645, Jul. 2007.

[13] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *J. Acoust. Soc. Amer.*, vol. 141, no. 5, pp. 3220–3235, May 2017.

[14] 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, Mar. 2006.

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

[16] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[17] A. Benichoux, L. Simon, E. Vincent, and R. Gribonval, "Convex regularizations for the simultaneous recording of room impulse responses," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1976–1986, Apr. 2014.

[18] R. Mignot, L. Daudet, and F. Ollivier, "Room reverberation reconstruction: Interpolation of the early part using compressed sensing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 11, pp. 2301–2312, Jul. 2013.

[19] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 205–216, Jan. 2014.

[20] E. Fernandez-Grande and A. Xenaki, "Compressive sensing with a spherical microphone array," *J. Acoust. Soc. Amer.*, vol. 139, no. 2, pp. EL45–EL49, Feb. 2016.

[21] N. Antonello, E. D. Sena, M. Moonen, P. A. Naylor, and T. von Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, Oct. 2017.

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

[23] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[24] T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2003, pp. 616–619.

[25] B. Rafaely, *Fundamentals of Spherical Array Processing*, J. Benesty and W. Kellermann, Eds. Berlin, Germany: Springer, 2015.

[26] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Measurement of sound fields using moving microphones," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 3231–3235.

[27] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," in *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, Mar. 2003.

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

[29] 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.

[30] 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, Feb. 2006.

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

[32] 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.

[33] S. Becker, E. Candès, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comput.*, vol. 3, no. 3, pp. 165–218, Jul. 2011.

[34] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[35] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, Jan. 2009.

[36] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014.

[37] T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388–3397, Jul. 2009.

[38] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, May 2008.

[39] M. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[40] M. Davies and R. Gribonval, "Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009.

[41] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing - Theory and Applications*, Y. Eldar and G. Kutyniok, Eds. New York, NY, USA: Cambridge Univ. Press, 2012, pp. 1–64.

[42] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[43] S. Mendelson, A. Pajor, and N. Jaegermann, "Uniform uncertainty principle for Bernoulli and Subgaussian ensembles," *Constructive Approx.*, vol. 28, no. 3, pp. 277–289, Dec. 2008.

[44] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[45] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008.

[46] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery* (Radon Series Comp. Appl. Math. 9), M. Fornasier, Ed. Berlin, Germany: de Gruyter, 2010, pp. 1–94.

[47] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imag. Sci.*, vol. 2, no. 4, pp. 1098–1128, Oct. 2009.

[48] H. Rauhut, J. Romberg, and J. Tropp, "Restricted isometries for partial random circulant matrices," *Appl. Comput. Harmon. Anal.*, vol. 32, no. 2, pp. 242–254, Mar. 2012.

[49] E. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, Nov. 2011.

[50] H. Rauhut, "Random sampling of sparse trigonometric polynomials," *Appl. Comput. Harmon. Anal.*, vol. 22, no. 1, pp. 16–42, Jan. 2007.

[51] V. Välimäki and T. Laakso, "Principles of fractional delay filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Jun. 2000, pp. 3870–3873.

[52] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

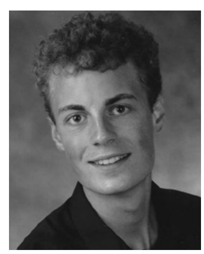
[53] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[54] T. Knopp *et al.*, "Trajectory analysis for magnetic particle imaging," *Phys. Med. Biol.*, vol. 54, no. 2, pp. 385–392, Jan. 2009.

Fabrice Katzberg (S'17) received the B.Sc. and M.Sc. degrees in medical engineering science from the University of Lübeck, Lübeck, Germany, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree at the Institute for Signal Processing, University of Lübeck. His current research interests include machine learning and digital signal and audio processing with special focuses on compressed sensing, and sound-field reconstruction.

Radoslaw Mazur (S'09–M'11) received the Diplominformatiker degree from the University of Oldenburg, Oldenburg, Germany, in 2004 and the Dr. Ing. degree in computer science from the University of Lübeck, Lübeck, Germany, in 2010. He was an Assistant Researcher with the Department of Physics, University of Oldenburg, from 2004 to 2006, and the then joined the University of Lübeck. His current research interests are digital signal and audio processing with special focuses on blind source separation, and listening room compensation.

Marco Maass (S'13) received the B.Sc. and M.Sc. degrees in computer science from the University of Lübeck, Lübeck, Germany, in 2010 and 2012, respectively. He is currently working toward the Ph.D. degree at the Graduate School for Computing in Medicine and Life Sciences, University of Lübeck. He is a Research Associate with the Institute for Signal Processing, University of Lübeck. His research interests include machine learning and image processing with special focuses on filter bank designs, MRI reconstruction, and MPI reconstruction.



Philipp Koch (S'17) received the B.Sc. and M.Sc. degrees in medical engineering science from the University of Lübeck, Lübeck, Germany, in 2013 and 2015, respectively. He is currently working toward the Ph.D. degree at the Institute for Signal Processing, University of Lübeck. His current research interests include image processing, digital signal and audio processing, and biosignal analysis with special focuses on hand movement recognition, and upper limb prosthesis control.

Alfred Mertins (M'96–SM'03) received the Dipl.Ing. degree from the University of Paderborn, Paderborn, Germany, in 1984, the Dr. Ing. degree in electrical engineering, and the Dr. Ing. habil. degree in telecommunications from the Hamburg University of Technology, Hamburg, Germany, in 1991 and 1994, respectively. From 1996 to 1997, he was with the University of Kiel, Germany, and from 1997 to 1998, with the University of Western Australia. In 1998, he joined the University of Wollongong, Australia, where he was an Associate Professor of electrical engineering. From 2003 to 2006, he was a Professor with the University of Oldenburg, Germany. In November 2006, he joined the University of Lübeck, Germany, where he is a Professor and the Director of the Institute for Signal Processing. His research interests include speech, audio, and image processing, multirate systems, pattern recognition, and digital communications.