COMPRESSIVE SAMPLING OF SOUND FIELDS USING MOVING MICROPHONES

Fabrice Katzberg, Radoslaw Mazur, Marco Maass, Philipp Koch, and Alfred Mertins

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

ABSTRACT

For conventional sampling of sound-fields, the measurement in space by use of stationary microphones is impractical for high audio frequencies. Satisfying the Nyquist-Shannon sampling theorem requires a huge number of sampling points and entails other difficulties, such as the need for exact calibration and spatial positioning of a large number of microphones. Dynamic sound-field measurements involving tracked microphones may weaken this spatial sampling problem. However, for aliasing-free reconstruction, there is still the need of sampling a huge number of unknown sound-field variables. Thus in real-world applications, the trajectories may be expected to lead to underdetermined sampling problems. In this paper, we present a compressed sensing framework that allows for stable and robust sub-Nyquist sampling of sound fields by use of moving microphones.

Index Terms— Room impulse responses, dynamic sound-field measurement, microphone array, compressed sensing

1. INTRODUCTION

The knowledge of room impulse responses (RIRs) that describe sound fields enables us to improve acoustic rendering by means of listening room compensation [1, 2, 3]. Common stationary approaches for measuring RIRs are the use of perfect sequences [4, 5], maximum-length sequences [6, 7], and exponential sine sweeps [8].

In [9], the spatio-temporal sampling of RIRs has been investigated. The spatial sampling of RIRs by use of equidistantly spaced microphones requires an extremely high effort. In addition to the calibration effort involving the compensation of spatio-temporal deviations and the equalization of the frequency responses of the individual microphones, an array of microphones will most likely never be dense enough to satisfy the Nyquist-Shannon sampling theorem without significant problems for very high audio frequencies. In order to make spatial sampling practical, methods for the dynamic measurement of RIRs have been proposed. In [10], a technique is presented that allows for the reconstruction of RIRs along the trajectory of a moving microphone by exploiting the Doppler effect. A specially designed input signal is needed for this and the speed of the microphone must be constant. An entirely different dynamic approach has been proposed very recently in [11, 12], where measurements taken along tracked microphone trajectories are related to a modeled sampling grid in space via interpolation. Based on the trajectory, excitation, and measured signal, 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 K unknown parameters is nonzero, the theory of compressed sensing (CS) allows for uniquely solving underdetermined systems of linear equations with sparsity constraint [13, 14]. In the context of signal processing, this translates to sub-Nyquist sampling of a signal vector $\boldsymbol{h} \in \mathbb{C}^U$ by merging the steps of sampling and compression. Under the assumption that $\boldsymbol{h}$ lives in a sub-space of dimension $K \ll M < U$ with respect to some unitary basis $\boldsymbol{\Psi} \in \mathbb{C}^{U \times U}$, the recovery of the signal by use of M incoherent [15] samples is possible by solving the regularized problem

$$\underset{\boldsymbol{c} \in \mathbb{C}^U}{\arg\min} \, \|\boldsymbol{x} - \boldsymbol{A}\boldsymbol{\Psi}^H \boldsymbol{c}\|_{\ell_2}^2 \quad \text{s.t.} \quad \|\boldsymbol{c}\|_0 \leq K, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{C}^M$ is the measurement signal, $\boldsymbol{A} \in \mathbb{C}^{M \times U}$ is the linear sampling operator, $\boldsymbol{c} = \boldsymbol{\Psi}\boldsymbol{h}$ is the sparse coefficient vector, and $\|\boldsymbol{c}\|_0$ is a pseudo norm counting the number of nonzero elements in $\boldsymbol{c}$. The problem (1) is NP-hard [16] and is typically solved using either a relaxation into an ℓ_1-minimization problem allowing for convex optimization, such as basis pursuit [17, 18], LASSO [19], or Dantzig selector [20], or a greedy algorithm such as the orthogonal matching pursuit (OMP) [21], compressed sensing matching pursuit (CoSaMP) [22], or iterative hard thresholding (IHT) [23, 24]. In order to guarantee stable and robust recovery of any approximately K-sparse signal, any arbitrary set of K columns of the CS matrix $\boldsymbol{\mathcal{A}} = \boldsymbol{A}\boldsymbol{\Psi}^H$ must build up a nearly orthogonal system, which is formalized by the so-called restricted isometry property (RIP) [25]. Verifying the RIP of a matrix is a combinatorial NP-hard problem. For practical applications, a more tractable property to evaluate the sampling process is the coherence [15]

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{1 \leq u \neq v \leq U} \frac{|\langle \boldsymbol{a}_u^c, \boldsymbol{a}_v^c \rangle|}{\|\boldsymbol{a}_u^c\|_{\ell_2} \|\boldsymbol{a}_v^c\|_{\ell_2}}, \tag{2}$$

where $\boldsymbol{a}_u^c$ denotes the u-th column of $\boldsymbol{\mathcal{A}}$. The theoretical guarantees for unique CS based recovery improve for a smaller coherence [26].

There are existing methods that exploit the principle of CS for sound-field recovery with only a few stationary, randomly placed microphones. In [27], a sparse plane wave approximation is applied to reconstruct the sound field at low frequencies up to $400\,\text{Hz}$. In [28], the sparsity of RIRs in time domain is exploited to interpolate the early-reflection part within a volume of interest.

In this paper, we provide a compressed sensing framework for sound-field measurements using moving microphones. In [11, 12], the sampling matrix was required to have full column rank and to formulate a well-posed problem in order to obtain the sound-field estimate by simply calculating the unique least-squares solution of the modeled linear system. However, for practical applications, especially when considering hand-held microphones, we generally expect trajectories that lead to ill-posed and even underdetermined problems. By exploiting that the spatio-temporal spectrum of sound-fields is ideally restricted to a hypercone along the tem-

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

poral frequency axis [9], we combine the dynamic sampling model from [11, 12] with a CS framework allowing for stable and robust sound-field recovery despite underdetermined variables.

2. DYNAMIC SOUND-FIELD SAMPLING

Sound fields describe the sound pressure with respect to both time t and receiver position $\boldsymbol{r} = [r_x, r_y, r_z]^T$. Assuming an acoustical environment that is linear and time-invariant, the sound pressure field is $p(\boldsymbol{r}, t) = \int_{-\infty}^{\infty} h(\boldsymbol{r}, \tau) s(t-\tau) d\tau$, where $s(t)$ is the signal emitted by a fixed source and $h(\boldsymbol{r}, t)$ is the spatially varying RIR from the source location to the point $\boldsymbol{r}$. Characteristics and sampling of $p(\boldsymbol{r}, t)$ are explicitly described in [9]. For the case in which a Dirac pulse is emitted at $t = 0$, the sound field is simplified to the spatio-temporal RIR $p(\boldsymbol{r}, t) = h(\boldsymbol{r}, t)$.

2.1. Uniform Sampling Grids

Measuring the sound field $h(\boldsymbol{r}, t)$ is basically a sampling problem. Assuming a bandlimited signal, it can be reconstructed through equidistant sampling in both time and space dimensions. Let $T = \frac{1}{f_s}$ 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. Accordingly, we have samples at equidistant time points $t_n = nT$ with $n \in \mathbb{Z}$ being the discrete variable of the time signal. For uniform sampling in the spatial dimensions, we consider a Cartesian grid where the equidistant sampling points $\boldsymbol{r}_{\boldsymbol{g}} \in \mathcal{G}$ are given by the set

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

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 $\lambda = \frac{c_0}{f_c}$ without aliasing, the spatial grid requires $\Delta_\xi < \frac{c_0}{2f_c}$, where c_0 is the speed of sound and Δ_ξ denotes the spatial sampling interval for each dimension $\xi \in \{x, y, z\}$.

2.2. Dynamic Sampling Model

The following descriptions consider a single microphone sampling on the trajectory $\boldsymbol{r}(n)$. The extension to multiple microphones is straightforward [12].

Aiming at the recovery of the finite set of $N = XYZ$ RIRs at discrete grid positions $\boldsymbol{g} \in G$, with $G = \{0, \ldots, X-1\} \times \{0, \ldots, Y-1\} \times \{0, \ldots, Z-1\}$ spanning a uniform grid inside the volume of interest, the dynamic measurement process is modeled as

$$x(\boldsymbol{r}(n), n) = \sum_{m=0}^{L-1} \sum_{\boldsymbol{g} \in G} \varphi_n(\boldsymbol{g})\, h(\boldsymbol{g}, m)\, s(n-m) + \eta(n), \tag{4}$$

where $\varphi_n(\boldsymbol{g})$ is an interpolation function for approximating the sound field at continuous positions $\boldsymbol{r}(n)$ for discrete time points n as linear combination of the grid RIRs $h(\boldsymbol{g}, n)$ subject to displacements $\boldsymbol{r}_{\boldsymbol{g}} - \boldsymbol{r}(n)$ (cf. Sec. 3.3). 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 $\boldsymbol{x} = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S} \boldsymbol{h}_u + \boldsymbol{\eta}$, where $\boldsymbol{x} \in \mathbb{R}^M$ contains the microphone samples, $\boldsymbol{\eta} \in \mathbb{R}^M$ comprises the noise, $\boldsymbol{h}_u \in \mathbb{R}^L$ is a sought grid RIR indexed by $u \in \{1, \ldots, N\}$, $\boldsymbol{\Phi}_u = \operatorname{diag}\{\varphi_0(\boldsymbol{g}_u), \ldots, \varphi_{M-1}(\boldsymbol{g}_u)\}$ is the corresponding diagonal matrix for interpolation, and $\boldsymbol{S} \in \mathbb{R}^{M \times L}$ is the convolution matrix of the source signal. The modeling of positions $\boldsymbol{r}_{\boldsymbol{g}_u}$ ($u \in \{1, \ldots, N\}$) 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

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{\eta} \tag{5}$$

with $\boldsymbol{h} \in \mathbb{R}^{NL}$ encapsulating the grid RIRs and the sampling matrix $\boldsymbol{A} \in \mathbb{R}^{M \times NL}$ possessing the block structure

$$\boldsymbol{A} = \left[\boldsymbol{\Phi}_1 \boldsymbol{S}, \boldsymbol{\Phi}_2 \boldsymbol{S}, \ldots, \boldsymbol{\Phi}_N \boldsymbol{S} \right]. \tag{6}$$

The m-th row of $\boldsymbol{A}$ is built up by the sampling vector

$$\boldsymbol{\phi}_m^H = [\varphi_{m-1}(0)\, \boldsymbol{s}(m), \ldots, \varphi_{m-1}(X-1)\, \boldsymbol{s}(m)] \tag{7}$$

with

$$\boldsymbol{s}(m) = [s(m-1), s(m-2), \ldots, s(m-L)] \tag{8}$$

projecting the unknown parameters onto measurement space:

$$\boldsymbol{\phi}_m^H \boldsymbol{h} = \langle \boldsymbol{h}, \boldsymbol{\phi} \rangle = x(m-1). \tag{9}$$

3. COMPRESSED SENSING FRAMEWORK

Considering the sparsely occupied hypercone forming the sound-field spectrum [9], let $\boldsymbol{T}_B \in \mathbb{C}^{B \times B}$ be a unitary matrix performing an orthonormal 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 $\boldsymbol{c} = \boldsymbol{\Psi}\boldsymbol{h}$, 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. The operator $\otimes$ denotes the Kronecker product. To keep the description and analysis simple, in Secs. 3.1, 3.2, and 3.3 the measurement space will be confined to the x dimension only.

3.1. 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 signal $h(g_x, n)$ contained in $\boldsymbol{h}$. Correspondingly, $\boldsymbol{c} \in \mathbb{C}^{XL}$ comprises the concatenated values of the 2D transformed sound-field spectrum. Following (9), the projection of the unknown parameters onto measurement space may be described by

$$\boldsymbol{\phi}_m^H \boldsymbol{\Psi}^H \boldsymbol{\Psi} \boldsymbol{h} = \langle \boldsymbol{\Psi}\boldsymbol{h}, \boldsymbol{\Psi}\boldsymbol{\phi}_m \rangle, \tag{10}$$

where $(\boldsymbol{\Psi}\boldsymbol{\phi}_m)^H = \boldsymbol{a}_m^r$ is the m-th row of the resulting CS matrix $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{M \times XL}$, i.e., the 2D transformed values of the components in $\boldsymbol{\phi}_m^H$. Following (7) and (8), the components in $\boldsymbol{\phi}_m^H$ are given by the particular sequence of the time reversed and spatially weighted source signal

$$s_m(g_x, n) = \varphi_{m-1}(g_x)\, s(m-1-n). \tag{11}$$

3.2. Fourier Representations

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 2D discrete Fourier transform of (11), given by

$$\mathcal{S}_m(k_x, l) = \frac{1}{\sqrt{XL}} \sum_{g_x=0}^{X-1} \sum_{n=0}^{L-1} s_m(g_x, n)\, e^{-2\pi j \frac{l}{L} n} e^{-2\pi j \frac{k_x}{X} g_x}, \tag{12}$$

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.

3.3. Influence of the Trajectory

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. Regarding (4), 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

$$D_x(n) = \frac{r_x(n) - r_0}{\Delta_x} \tag{13}$$

is the spatial-grid delay. Thus, the ideal frequency response of a FD filter reads

$$\Phi_n^{\text{id}}(e^{j\kappa_x}) = e^{-jD_x(n)\kappa_x}, \tag{14}$$

with the spatial angular frequency κ_x. Let the columns of $\boldsymbol{\mathcal{A}}$ consist of values of the Fourier spectra $\mathcal{S}_m(k_x, l)$ defined in (12): $\boldsymbol{a}^c_{(k_x,l)} = [\mathcal{S}_1(k_x, l), \ldots, \mathcal{S}_M(k_x, l)]^T$. Then, 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

$$\mathcal{S}_{n+m}(k_x, l) = \mathcal{S}_n(k_x, l)\, e^{-2\pi j (D_x(m) - D_x(n))\frac{k_x}{X}}\, e^{-2\pi j m \frac{l}{L}}. \tag{15}$$

Thus, for a spectrally flat excitation sequence and a perfect interpolation kernel, each column in $\boldsymbol{\mathcal{A}}$ can be described by use of structured spatial and temporal phase terms according to

$$\boldsymbol{a}^c_{(k_x,l)} = \begin{bmatrix} a^0_{(k_x,l)} \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(1) - D_x(0))\frac{k_x}{X}}\, e^{-2\pi j 1 \frac{l}{L}} \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(2) - D_x(0))\frac{k_x}{X}}\, e^{-2\pi j 2 \frac{l}{L}} \\ \vdots \\ a^0_{(k_x,l)}\, e^{-2\pi j (D_x(M-1) - D_x(0))\frac{k_x}{X}}\, e^{-2\pi j (M-1) \frac{l}{L}} \end{bmatrix}, \tag{16}$$

where the initial phase state $a^0_{(k_x,l)} = e^{-2\pi j D_x(0)\frac{k_x}{X}}\, \sigma_s\, e^{j\theta_0(l)}$ is determined by the initial grid delay $D_x(0)$ and the initial phase $\theta_0(l)$ of the excitation signal with power σ_s^2.

3.4. Coherence of Measurements

From the column representation (16), the deduction of the coherence (2) belonging to the particular CS matrix of the 4D sound-field sampling problem is straightforward. Let the interpolation function be separable on the spatial grid. By defining the difference between two discrete frequencies for each dimension,

$$\Delta k_x = k'_x - k''_x,\ \Delta k_x \in \{-(X-1), \ldots, X-1\}, \tag{17}$$

$$\Delta k_y = k'_y - k''_y,\ \Delta k_y \in \{-(Y-1), \ldots, Y-1\}, \tag{18}$$

$$\Delta k_z = k'_z - k''_z,\ \Delta k_z \in \{-(Z-1), \ldots, Z-1\}, \tag{19}$$

$$\Delta l = l' - l'',\ \Delta l \in \{-(L-1), \ldots, L-1\}, \tag{20}$$

the vector $\boldsymbol{d} = [\Delta k_x, \Delta k_y, \Delta k_z]^T$, the trajectory

$$\boldsymbol{r}_D(n) = [D_x(n), D_y(n), D_z(n)]^T \tag{21}$$

relative to the modeled grid $\mathcal{G}$, and the trajectory-dependent function

$$\mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d}) = e^{-2\pi j \left(\frac{D_x(n)}{X}\Delta k_x + \frac{D_y(n)}{Y}\Delta k_y + \frac{D_z(n)}{Z}\Delta k_z \right)}, \tag{22}$$

the coherence of the modeled sound-field sampling problem reads

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{(\boldsymbol{d}, \Delta l)} \frac{1}{M} \left| \sum_{n=0}^{M-1} \mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d})\, e^{-2\pi j \frac{n}{L}\Delta l} \right|, \tag{23}$$

with $(\boldsymbol{d}, \Delta l) \neq (\boldsymbol{0}_3, 0)$, where $\boldsymbol{0}_3$ is the zero vector of size three. First, we conclude that arbitrary spectrally white excitation signals lead to the same coherence of the CS problem, independent of their spectral phases. Further, for determining the coherence of $\boldsymbol{\mathcal{A}}$, only varying tupels 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, this knowledge reduces the effort for computing the coherence of the 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 (11) is a real-valued signal in practice, the particular spectra in $\boldsymbol{\mathcal{A}}$ are conjugate symmetric, which may be exploited for saving further computational cost.

The theoretical guarantees for CS based recovery improve for lower coherence, thus, the microphone trajectory is considered optimal for the set of parameters $\mathcal{P} = \{(\boldsymbol{r}_D(n), \mathcal{G})\}$ minimizing (23). The expression (23) may be used for both of the following tasks that arise in practice: finding the optimal trajectory for a sought grid in space and modeling an optimal grid for a given measurement.

Indeed, a perfect interpolator with ideal frequency response is unrealizable in practice. Nevertheless, considering a virtual grid leading to spatial oversampling by factor $\alpha > 2$, the FD filter approximation by an appropriate Lagrange interpolator achieves ideal magnitude and phase response for the relevant frequency range where the bandlimited signal actually lives [29]. Accordingly, all columns of the CS matrix that potentially span the measurement space may be described by (16) and their coherence is given by (23).

4. FAST 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 estimating the solution of (1) and requires low effort in computation and memory [24]. It approaches the optimum by applying the simple update rule

$$\hat{\boldsymbol{c}}^{(i+1)} = T_K \left\{ \hat{\boldsymbol{c}}^{(i)} + \mu\, \boldsymbol{\Psi} \boldsymbol{A}^T (\boldsymbol{x} - \boldsymbol{A}\boldsymbol{\Psi}^H \hat{\boldsymbol{c}}^{(i)}) \right\}, \tag{24}$$

where μ is the step size and T_K is the nonlinear thresholding operator that sets all but the K largest absolute values in the updated signal to zero. Applied to our sound-field sampling model (5), (24) simplifies to a fast update scheme involving the calculation of the residuum according to

$$\boldsymbol{\varepsilon}^{(i)} = \boldsymbol{x} - \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S} \hat{\boldsymbol{h}}_u^{(i)}, \tag{25}$$

the update of the estimated grid RIRs subject to

$$\hat{\boldsymbol{h}}_u^{(i+1)} = \hat{\boldsymbol{h}}_u^{(i)} + \mu \boldsymbol{S}^T \boldsymbol{\Phi}_u \boldsymbol{\varepsilon}^{(i)}, \tag{26}$$

and the transform of the estimated sound-field variables $\hat{\boldsymbol{h}}^{(i)}$ into their sparse frequency representation $\hat{\boldsymbol{c}}^{(i)}$ for hard thresholding. In (25) and (26), the convolutions of the source signal with the estimated RIRs and the residuum, respectively, can be computed very efficiently in Fourier domain. The diagonal matrices $\boldsymbol{\Phi}_u$ only contribute weighting factors subject to the spatial interpolation.

5. EXPERIMENTS AND RESULTS

For the following experiments, we simulated RIRs and microphone measurements by use of the image source method [30] considering a room of size $5.8\,\mathrm{m} \times 4.15\,\mathrm{m} \times 2.55\,\mathrm{m}$. 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 $1\,\mathrm{m}$. The origin of the spatial grid $\mathcal{G}$ was set to $\boldsymbol{r}_0 = [2.75, 1.4, 0.8]^T$. For excitation, we used white Gaussian noise with power $\sigma_s^2 = 1$. Measurements were taken by one moving microphone performing a Lissajous trajectory with frequency ratio $f_x/f_y = 3/4$ [31] covering the entire region of interest. For the recordings, white Gaussian noise with a signal-to-noise ratio of $30\,\mathrm{dB}$ was added. In order to allow for a baseline of the reconstruction quality and to compare our CS based recovery with the strategy used in [11, 12], where the sound-field estimate was the least-squares (LS) solution obtained by the pseudoinverse of $\boldsymbol{A}$, we consider a sampling scenario ensuring that the system matrix is small enough to allow for computing the pseudoinverse. Thus, we sampled the sound-field on a planar grid of size 5×5 with spacing $\Delta_{x,y} = 0.025\,\mathrm{m}$ at fixed height $0.8\,\mathrm{m}$. The RIRs were limited to length $L = 500$. In total, this problem involves $U = 12500$ sound-field variables. Of course, the CS recovery scheme from Sec. 4 is practical for much larger problems, since for updating the estimate only fast Fourier transforms, point-wise multiplications and vector summations are required.

As evaluation criterion for the quality of the recovered sound-field involving N grid RIRs, the mean normalized system misalignment [11, 12]

$$\mathrm{MNSM} = \frac{1}{N} \sum_{u=1}^{N} \frac{\|\boldsymbol{h}_u - \hat{\boldsymbol{h}}_u\|_{\ell_2}^2}{\|\boldsymbol{h}_u\|_{\ell_2}^2} \tag{27}$$

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. Several numbers of $M \in \{15000, 10000, 8000, 6000\}$ samples taken along the Lissajous trajectory were tested for reconstruction. The sampling process was modeled with $\phi_n(\boldsymbol{g})$ being the Lagrange interpolator and having full support on the grid. For the CS based recovery via IHT, the step size $\mu = 10^{-4}$ was chosen. As sparse frequency representation of the sound-field variables we used their Fourier transforms. The iterative recovery was started with $\hat{\boldsymbol{c}}^{(0)} = \boldsymbol{0}_U$. Beginning with a small K, the sparsity constraint was successively relaxed every 50 iterations, so that $K = M/2$ was reached after 500 iterations.

The MNSMs of the CS based sound-field estimates subject to the number of iterations are presented in Fig. 1(a). The markers on the curves indicate the successive sparsity relaxations. The linear system is overdetermined only for $M = 15000 = 1.2\,U$. The error of its unique LS solution is the dashed baseline in Fig. 1(a) at $-8.23\,\mathrm{dB}$. In comparison with that, for the same number of samples, the IHT algorithm obtains a significant reduction of the error to $-14.02\,\mathrm{dB}$. For a lower number of $M = 10000 = 0.8\,U$ samples, there is no unique LS solution, but the proposed CS method is still competitive and achieves an error of $-9.24\,\mathrm{dB}$. Thus, the IHT solution for the underdetermined case is even better than the LS solution obtained by 40% more samples. In Fig. 1(b), the corresponding energies of the residual signals are presented. The IHT algorithm converges to a minimal residuum. After about 400 iterations the residuum still decreases, whereas the actual accuracy gain for the sound-field estimate stagnates as shown in Fig. 1(a). This is due to the non-perfect measurement model, which contains inconsistencies caused by measurement noise and non-perfect spatial interpolation.

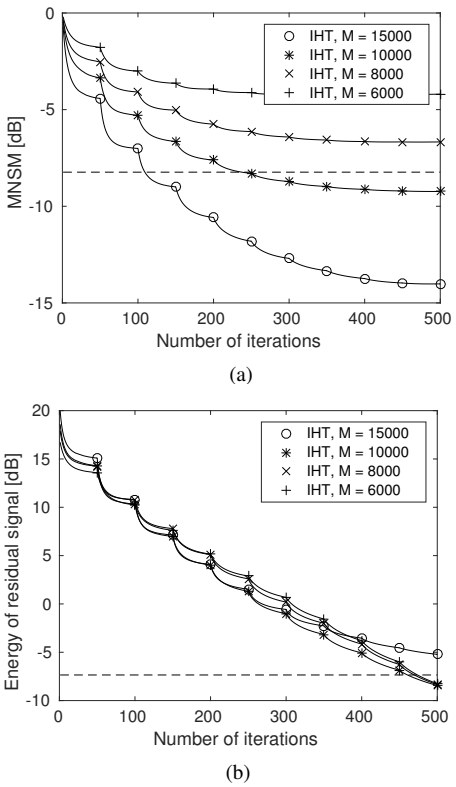

Fig. 1. Results for CS based sound-field recovery using the IHT algorithm, with (a) the errors of the sound-field estimates and (b) the energies of the residual signals for different numbers of samples M, and $U = 12500$ unknown variables. The dashed lines indicate the results of the least-squares solution for $M = 15000$.

6. CONCLUSION

In this paper, we presented a CS framework for the sound-field recovery from samples taken by moving microphones. By knowing the source signal and the microphone positions, a linear system of equations can be set up that leads to the estimation of RIRs on a uniform grid. In order to ensure a stable and robust sound-field recovery even in the underdetermined case, a CS solution involving additional regularization has been derived. It exploits the sparsity of sound fields in frequency domain. The structure of the resulting CS matrix and its dependency on the microphone trajectory have been shown. Based on that, a simple trajectory-dependent expression for the coherence of measurements given spectrally flat excitation has been derived. It enables us to efficiently evaluate microphone trajectories in terms of CS reconstruction. Finally, a simple iterative scheme according to the IHT algorithm has been provided allowing for CS based sound-field recovery at low cost in computation and memory. Further extensions exploiting also the sparsity of the early-reflection part of RIRs are currently under investigation.

7. REFERENCES

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

[2] D. Talagala, W. Zhang, and T. Abhayapala, "Efficient multichannel 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, 2014.

[3] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012, pp. 13–16.

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

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

[6] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc.*, vol. 31, no. 7/8, pp. 478–488, 1983.

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

[8] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Audio Engineering Society Convention*, 2007, pp. 1–21.

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

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

[11] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Measurement of sound fields using moving microphones," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2017, pp. 3231–3235.

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

[13] 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, 2006.

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

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

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

[17] 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, 1998.

[18] 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, 2006.

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

[20] 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.

[21] 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, 2007.

[22] 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, 2008.

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

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

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

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

[27] 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, 2014.

[28] 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, 2013.

[29] V. Välimäki and T. Laakso, "Principles of fractional delay filters," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 3870–3873.

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

[31] T. Knopp, S. Biederer, T. Sattel, J. Weizenecker, B. Gleich, J. Borgert, and T. M. Buzug, "Trajectory analysis for magnetic particle imaging," *Phys. Med. Biol.*, vol. 54, no. 2, pp. 385–392, 2009.