Sound-field measurement with moving microphones

Fabrice Katzberg,[a] Radoslaw Mazur, Marco Maass, Philipp Koch, and Alfred Mertins
Institute for Signal Processing, University of Lübeck, 23562 Lübeck, Germany

(Received 11 November 2016; revised 21 April 2017; accepted 24 April 2017; published online 16 May 2017)

Closed-room scenarios are characterized by reverberation, which decreases the performance of applications such as hands-free teleconferencing and multichannel sound reproduction. However, exact knowledge of the sound field inside a volume of interest enables the compensation of room effects and allows for a performance improvement within a wide range of applications. The sampling of sound fields involves the measurement of spatially dependent room impulse responses, where the Nyquist-Shannon sampling theorem applies in the temporal and spatial domains. The spatial measurement often requires a huge number of sampling points and entails other difficulties, such as the need for exact calibration of a large number of microphones. In this paper, a method for measuring sound fields using moving microphones is presented. The number of microphones is customizable, allowing for a tradeoff between hardware effort and measurement time. The goal is to reconstruct room impulse responses on a regular grid from data acquired with microphones between grid positions, in general. For this, the sound field at equidistant positions is related to the measurements taken along the microphone trajectories via spatial interpolation. The benefits of using perfect sequences for excitation, a multigrid recovery, and the prospects for reconstruction by compressed sensing are presented. © *2017 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4983093]

[DKW] Pages: 3220–3235

I. INTRODUCTION

Closed-room scenarios exhibit undesirable acoustic properties for various audio applications. Acoustic systems, such as communication devices and multichannel audio systems, ordinarily assume a free-field environment and decrease their performance in the presence of reverberation. However, there are existing methods for listening-room compensation[1–3] which aim at removing the acoustical influences of rooms during sound-field reproduction.

There are several stationary methods available for the measurement of room impulse responses (RIRs). Common approaches are the use of perfect sequences[4,5] and maximum-length sequences[6,7] (MLS). Perfect sequences are periodic pseudo-random signals that have an autocorrelation function which is a periodic sequence of Dirac pulses. If the period is longer than the RIR, the measurement reduces to a correlation of the received signal with the sequence. MLS sequences are binary sequences that come close to perfect sequences, and the longer the period becomes, the better the reconstruction gets. Another well established method is the use of exponential sine sweeps.[8] As with the previous methods, the RIR can be obtained by correlating the measured signal with an appropriate second signal, often referred to as the inverse signal. An advantage of sweep-based measurements is that they are less sensitive to system nonlinearities. Finally, also perfect sweeps[9] have been introduced, which combine the properties of perfect sequences and sweeps. For dynamic setups, such as acoustic echo cancellation (AEC),[10] in which the RIRs change over time, the common approach is to demand a minimum mean squared error (MMSE) between the measured signal and the output of an adaptive filter that is excited by the same input as the loudspeaker. Of course, the method can also be used to measure time-invariant RIRs. In this case, an excitation with white noise would be preferable, because it provides maximum convergence speed.[10]

The above mentioned methods are essentially all based on minimizing a mean squared error. However, there also exist methods for RIR estimation that exploit the sparsity of the early parts of RIRs and the exponential decay of later parts. A group of these methods is known as proportionate update schemes.[11–14] In these methods, which are already common practice in echo cancellation, an individual step size is computed for each filter coefficient, resulting in faster convergence for the few large filter taps of a sparse RIR. A method for multichannel estimation that exploits the sparsity of the early parts of RIRs by introducing an $\ell_{p,q}$-norm based regularization term has been also proposed.[15] The term $\ell_{p,q}$ means that the ℓ_q-norm of a set of ℓ_p-norms is computed. The authors report that the convergence rate improves due to the sparsity constraint and that the non-uniqueness problem of multichannel identification,[16] although not yet resolved, is less severe than with least-squares approaches.

In order to describe the spatio-temporal sound field inside a defined volume of interest, the concept of the plenacoustic function[17,18] (PAF) has been introduced. Generally, the PAF encapsulates the information on the entire set of spatio-temporal RIRs for any position in space. The sampling of the PAF by considering equidistant sampling points in space that satisfy the Nyquist-Shannon sampling theorem is not practical for larger bandwidths.

There is an existing method for the dynamic measurement of a set of RIRs using one microphone that moves along a given trajectory.[19] Despite the motion of the

[a]Electronic mail: katzberg@isip.uni-luebeck.de

 0001-4966/2017/141(5)/3220/16/$30.00

microphone, the RIRs are reconstructed for all positions along the trajectory. In particular, motion along a line and a circle have been studied. To make this principle work, a specially designed input signal is needed, and the speed of the microphone must be constant and is restricted to an upper limit. Quite different from RIR measurements with a fixed microphone, the excitation signal must not contain all audio frequencies, but only a certain subset. The omitted frequencies are essentially generated through the Doppler effect. The technique is suited to record relatively short RIRs and has also been proposed to record head related transfer functions with a microphone that moves along a circular trajectory around the head. Further, a setup was considered in which a microphone was moved along a random trajectory.[20] However, this setup was not used to reconstruct all RIRs along the trajectory or even the entire PAF within a volume, but to simulate time-varying channels that are governed by the wave equation and for the computation of frequency-dependent update rules for adaptive algorithms.

In this paper, we propose an approach which allows for the measurement of the PAF using a moving array with a manageable number of microphones. The simplest setup involves only one hand-held microphone. Each recorded sample of a microphone signal is regarded as the result of the projection of the wanted parameters onto an appropriate sampling vector and contributes one equation to a linear system, which finally allows for a reconstruction of the sound field inside the considered volume. This procedure is described in Sec. III. Section IV characterizes potential sources of error and gives the error analysis for spectrally flat excitation.

The proposed method is entirely different from the dynamic measurement procedure proposed by Ajdler *et al.*[19] The generic approach described in this paper places no specific demands on the speed of the microphone array or on the excitation signal. However, a higher velocity of movement may induce a higher noise level caused by air flow. A potential application of the proposed method is the high-precision measurement of sound fields using hand-held microphones whose positions are continuously tracked, e.g., using gyroscopes.

In Sec. V we present three extensions for the general method. We show that, by using perfect sequences, the system of linear equation decomposes into smaller problems with heavily reduced computational demand. Furthermore, we introduce a multiresolution recovery scheme which enables us to reconstruct low-frequency content for the case in which the wideband recovery problem is ill-conditioned. In addition, it improves the performance in the presence of noise. Section V C discusses prospects for incorporating the dynamic sampling procedure into the framework of compressed sensing, in order to allow for the reconstruction by means of arbitrary trajectories which typically lead to an underdetermined system of linear equations. However, experiments discussed in Sec. VI focus on the determined case.

The paper uses the following notations. Vectors and matrices are represented in boldface characters, lower and upper case, respectively. The superscript $\{\cdot\}^T$ means transposition. $\|\cdot\|_{\ell_p}$ returns the ℓ_p-norm of a vector. The operation $\mathrm{diag}\{\cdot\}$ turns a set of scalars or matrices into a diagonal matrix or block diagonal matrix, respectively. The trace of a matrix is indicated by $\mathrm{tr}\{\cdot\}$. The element at the ith row and jth column of a matrix is given by $[\cdot]_{i,j}$. $\boldsymbol{I}_{D\times D}$ denotes the $D\times D$ identity matrix and $\boldsymbol{0}_{D\times D}$ means the $D\times D$ zero matrix. The asterisk signifies the convolution operation. The expectation operation is denoted by $E\{\cdot\}$. The unit pulse is represented by $\delta(n)$, which equals 1 for $n=0$ and 0 otherwise. The modulo operation is denoted by $a \bmod b$, which gives the remainder after the division of a by b. The sets of integer, real, and complex numbers are denoted by $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$, respectively.

II. SOUND FIELDS IN CLOSED ROOMS

In this section, the underlying sound propagation model for the simulations in Sec. VI is described. Furthermore, the plenacoustic function is outlined, which is used in this paper as representation of the spatio-temporal sound field.

A. Sound propagation model

Consider a fixed sound source that emits the continuous-time sound signal $s(t)$ depending on time t. In closed-room scenarios, the received signal $x(t)$ is a filtered version of $s(t)$: it consists of a succession of the first incoming direct sound, the early reflections from the walls, and the diffuse late-field reverberation with exponentially decaying energy. For a certain listener position in space, the room impulse response $h(t)$ characterizes the time dependent sequence of the received sound pressure that would result from an emitted Dirac pulse. Assuming a linear time-invariant (LTI) system for a fixed emitter-receiver pair, the temporal connection between the excitation signal and the observed signal is given by

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

Mathematically, the sound propagation and thus RIRs are described by the solutions of the wave equation. Anyway, assuming that the emitted wavelengths are much smaller than the dimensions of the room, the sound propagation may be approximated in terms of geometrical acoustics. In fact, this model restricts the original problem of wave propagation to the much simpler description of energy propagation. Accordingly, all effects involving phase differences are neglected. However, for sound signals with many spectral components, constructive and destructive phase effects can be assumed to cancel out, i.e., the total energy at the listener position is obtained by simply adding the energies of the components.[21]

The geometrical model considers sound rays as straight lines of different directions along which the sound energy travels with constant velocity. These sound rays can be regarded as small sections of infinitely small solid angles that are cut out of spherical waves.[21] Before reaching the listener position and superposing there, the sound rays pass possibly consecutive reflections and absorptions at the walls,

except for the ray of the direct path. The image method[22] provides an efficient algorithm for the ray-based simulation of RIRs inside rectangular rooms. This method creates mirror images of the sound source to calculate the reflected paths as direct paths from the virtual sources to the receiver. Each of these direct paths contributes a delayed and scaled impulse to the RIR, depending on the distance, the order of reflection, and the reflection coefficients of the walls. In order to achieve a more realistic behavior of zero-mean RIRs, the DC component has been removed by a low-frequency high-pass filter in Ref. 22. The simulations in the experimental part of this paper are based on an enhanced version of the image method including an improved sampling and phase accuracy of the simulated RIRs.[23] This balances the positive and negative filter taps in a reasonable way.

B. The plenacoustic function

Ajdler and co-workers[17,18] introduced the concept of the plenacoustic function (PAF), encapsulating all RIRs of a room for a given source configuration. The PAF, denoted in the following as $p(\boldsymbol{r},t)$, describes the sound field in space depending on both time t and receiver location $\boldsymbol{r}=[r_x,r_y,r_z]^T$. In the simplest case, with a single point source emitting a signal $s(t)$ at fixed position, the PAF 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 $\boldsymbol{r}$. Through the LTI system model, the PAF for multiple fixed sound sources consists of a sum of integrals as given in (2). This means, the received sound pressure is a superposition of single source signals, each of them convolved with their specific spatio-temporal RIR. Referring to Ajdler and Vetterli[17] and Ajdler *et al.*[18] we consider the PAF, without loss of generality, only for the case where a single source at fixed position emits a Dirac impulse at $t=0$. With this, the PAF is simplified to the spatio-temporal RIR:

$$p(\boldsymbol{r},t)=h(\boldsymbol{r},t). \tag{3}$$

C. The spatial sampling problem

Measuring the sound field is basically a sampling problem. Under the assumption that the PAF is bandlimited, it can be reconstructed through equidistant sampling in time and space dimensions. According to the Nyquist-Shannon sampling theorem and considering the temporal cutoff frequency f_c, the sampling frequency has to fulfill the condition

$$f_s>2f_c \tag{4}$$

in order to avoid aliasing. Furthermore, the PAF requires sampling in space, so the condition

$$\Delta_\xi<\frac{c_0}{2f_c}\quad\forall\,\xi\in\{x,y,z\} \tag{5}$$

must hold, where c_0 is the speed of sound and Δ_ξ denotes the spatial sampling interval for each dimension x, y, z. This allows for the reconstruction of sound waves in space with the minimal wavelength

$$\lambda_{\min}=\frac{c_0}{f_c}. \tag{6}$$

Because of the relationship in Eq. (5), the uniform sampling of the sound field by use of equidistantly spaced microphones often requires an extremely high effort. An array of microphones will most likely never be dense enough to enable measurements without significant problems for very high audio frequencies. For example, the sampling of the PAF with $f_c=17\,\mathrm{kHz}$ inside a volume of 1 m^3 requires at least 10^6 spatial measuring points. In order to reduce this infeasible effort, sampling schemes with quasi randomly spaced microphones have been proposed, based on the principle of compressed sensing.[24] However, the maximum admissible audio frequency within larger volumes is still significantly reduced. To give an example, a 3D array with 120 microphones randomly positioned within a cube of edge length 2 m is sufficient for a stable reconstruction of frequencies up to 400 Hz using compressed sensing.[24] For this, a sparse plane wave approximation of the PAF in low frequencies is exploited.

A specific problem of microphone arrays is the need of calibration. This includes the compensation of spatio-temporal deviations and the equalization of the frequency responses of the individual microphones. The use of more microphones involves an increasing expense for calibration. As a remedy, methods for automatic calibration have been proposed.[25] In the following, we propose a dynamic sampling procedure using a moving array with a number of microphones that is small enough to allow proper calibration.

III. DYNAMIC SAMPLING PROCEDURE

To determine the sound field within a volume of interest, we consider a scenario in which a single source emits a pre-defined signal and one or more microphones are moved through the volume while their signals are simultaneously recorded together with the microphone-position information. If, for example, a rigid microphone array is moved through the volume, one particular position on the array and its orientation need to be captured in order to recover the positions of all microphones. Within the array, the microphones would need to be calibrated as for the above-mentioned method.[24] Of course, the number of microphones can be traded against the total measurement time, allowing for compromises with a reasonable number of microphones and measurement time. To keep the following description and analysis simple, a single microphone is considered. Extensions to multiple microphones will be discussed in Sec. III C.

While the microphone trajectory may be quite arbitrary during the measurement, the unknown values to be determined by the proposed procedure are the RIRs from the source position to all locations within the volume of interest on a Cartesian grid, as required by the sampling theorem. These grid locations are virtual sampling points from which

the RIRs to any location within the volume of interest can be computed via interpolation, including the locations on the microphone trajectory. In fact, this interpolation is the key to our method. Describing the samples recorded along the trajectory based on the knowledge of the excitation signal and the RIRs on the grid leads to a linear system of equations that can be solved for the unknown grid RIRs.

A. Uniform sampling and reconstruction

First, consider the uniform time sampling of the PAF. As usual, $T = 1/f_s$ denotes the sampling interval in the time dimension with f_s satisfying Eq. (4). This leads to measurements at equidistant sampling points $t_n = nT$, where $n \in \{0, 1, 2, \ldots\}$ denotes the discrete time variable of the causal signal. Beforehand, the temporal bandwidth of the PAF is limited with an analog low-pass filter blocking all frequencies above f_c. The model parameter f_c is determined by the requested application.

For the uniform sampling in the spatial dimensions, consider a Cartesian grid where the equidistant sampling points $\boldsymbol{r}_g \in \mathcal{G}$ are given by the set

$$\mathcal{G} = \{\boldsymbol{r}_g \,|\, \boldsymbol{r}_g = \boldsymbol{r}_0 + \left[g_x\Delta_x, g_y\Delta_y, g_z\Delta_z\right]^T\} \tag{7}$$

with the grid origin $\boldsymbol{r}_0$ and the discrete grid variables in $\boldsymbol{g} = [g_x, g_y, g_z]^T \in \mathbb{Z}^3$. The sampling intervals Δ_x, Δ_y, Δ_z must follow Eq. (5), in order to avoid spatial aliasing.

As already mentioned above, the aim of our method is to determine RIRs at equidistant positions by use of dynamic measurements. For that purpose, a virtual sampling grid in space that fulfills Eq. (5) is modeled, with integer indices in $\boldsymbol{g}$ spanning the virtual grid coordinate system. The RIRs on that grid are denoted as $h(\boldsymbol{r}_g, t_n)$. In practice, the measuring area is finite and the spatial sampling grid is limited to size $X \times Y \times Z$. Overall, the recovery of $h(\boldsymbol{r}_g, t_n)$ inside the finite volume of interest involves $N = XYZ$ RIRs at grid positions $\boldsymbol{g} \in G$ with

$$G = \{0, \ldots, X-1\} \times \{0, \ldots, Y-1\} \times \{0, \ldots, Z-1\}. \tag{8}$$

The amplitudes of these grid RIRs are assumed to vanish into the noise level beyond t_{L-1} for given f_s, so each of the RIRs is limited to length L. Hence, the uniform sampling of the PAF covers $U = NL$ values in total. As from now, the sought grid RIRs are denoted by $h(\boldsymbol{g}, n)$ using the discrete variables.

The ideal reconstruction of the continuous sound field $h(\boldsymbol{r}, t)$ from $h(\boldsymbol{g}, n)$ is accomplished by a 4D sinc filter with infinite support. Due to the exponential energy decay of the RIRs, finite length interpolation filters achieve reasonable approximations for the time dimension. However, the finite number of measurements in the space dimensions still tightens the spatial sampling problem: to improve the spatial reconstruction despite of 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 below the Nyquist rate. For both strategies, the number of spatial measurement points increases.

B. Measurement setup and linear system model

In this section, we interrelate the virtual RIRs, modeled on a uniform grid in space, with the samples of one moving microphone to set up a system of linear equations. In general, the sampling points are located at intermediate positions, so the key of our method is spatial interpolation. The extension to Q microphones is straightforward and allows for the acquisition of Q times more equations during the same measuring time.

Let us consider a single moving microphone and let $x(n)$ denote the measured signal of the microphone which moves along a tracked trajectory through the entire volume of interest. The speed of the microphone is arbitrary. Let $s(n)$ denote the signal emitted by the sound source, and let $\eta(n)$ be the measurement noise, which is assumed to be statistically independent of $s(n)$. The choice of the excitation signal is arbitrary concerning the reconstruction procedure. However, $s(n)$ should cover the entire spectrum of the bandlimited PAF which is to be measured, so white noise and perfect sequences are convenient.

The measurement procedure and a generic virtual sampling grid are outlined in Fig. 1. The position of the dynamic microphone is given by the time dependent vector $\boldsymbol{r}(n) = [r_x(n), r_y(n), r_z(n)]^T$. In theory, any continuous position could be interpolated by means of a spatial grid with intervals Δ_x, Δ_y, Δ_z fulfilling the Nyquist-Shannon requirement. However, the proposed method involves the reverse interpolation problem: given the sound field at continuous positions $\boldsymbol{r}(n)$, the aim is to recover it at regular grid positions $\boldsymbol{r}_g$. So,

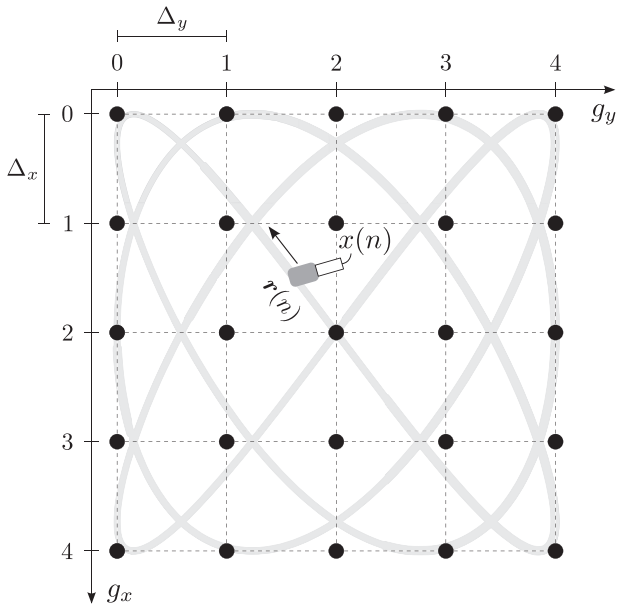

FIG. 1. Arrangement of a virtual 2D sampling grid in space with reference to the grid coordinate system. The spatial sampling intervals Δ_x, Δ_y translate the discrete variables g_x, g_y to the world coordinate system. The dots represent the positions of the virtual grid RIRs. This example sketches one dynamic microphone moving along a Lissajous trajectory in between the grid positions.

the actual measurement signal $x(n)$ already contains the interpolation results, i.e., the weighted linear combinations of pressure values on the spatial grid we are looking for. As long as the moving microphone acquires samples within the interpolation range of one particular grid position, it samples an amount of sound pressure on that grid point. The amount is determined by the corresponding weightings, and, thus, by the trajectory $\boldsymbol{r}(n)$. Consequently, the spatial distance of samples taken by the moving microphone is not restricted to Δ_x, Δ_y, Δ_z.

Each sample recorded by the dynamic microphone, together with the knowledge of $\boldsymbol{r}(n)$, $s(n)$, and the positions $\boldsymbol{r_g}$ of the modeled grid RIRs, contributes one equation to a system of linear equations. The basic concept is to interrelate the measurement signal with the source signal according to Eq. (2), so

$$x(n) = \sum_{k=0}^{L-1} h(\boldsymbol{r}(n), k)\, s(n-k) + \eta(n), \tag{9}$$

where $h(\boldsymbol{r}(n), k)$ is the spatially varying and therefore time dependent RIR at the measuring position at time index n. Its taps are indexed with $k \in \{0, 1, \ldots, L-1\}$. The explicit taps of $h(\boldsymbol{r}(n), k)$ are unknown. They differ significantly for slight variations of $\boldsymbol{r}(n)$, due to the various reflection paths inside the closed room. To interrelate each sample with the same set of unknown variables, the spatially varying RIR is represented as weighted sums of the N modeled grid RIRs, so

$$h(\boldsymbol{r}(n), n) \approx \sum_{\boldsymbol{g}\in G} \varphi(\boldsymbol{r}(n), \boldsymbol{r_g})\, h(\boldsymbol{g}, n). \tag{10}$$

The interpolation function $\varphi(\boldsymbol{r}(n), \boldsymbol{r_g})$ weights the grid RIRs at $\boldsymbol{r_g}$ subject to the displacements $\boldsymbol{r}(n) - \boldsymbol{r_g}$. It can be separated into the product of three one-dimensional functions $\varphi_\xi(r_\xi(n), r_{g_\xi})$ with $\xi \in \{x, y, z\}$, due to the equidistant sampling grid. The interpolation (10) is not necessarily perfect in practice, especially when the spatial sampling grid is modeled with sampling intervals close to the upper limit $c_0/2f_c$. Nevertheless, the interpolation is assumed to be ideal in the following considerations. With adequate oversampling for the spatial grid, an interpolation kernel which is maximally flat in the frequency domain may provide sufficient results. This is shown in the experimental part of this paper.

Substituting Eq. (10) into Eq. (9) and assuming perfect interpolation leads to

$$x(n) = \sum_{k=0}^{L-1} \sum_{\boldsymbol{g}\in G} \varphi(\boldsymbol{r}(n), \boldsymbol{r_g})\, h(\boldsymbol{g}, k)\, s(n-k) + \eta(n). \tag{11}$$

This represents each sample as the weighted sum of the same U unknown variables $h(\boldsymbol{g}, n)$ and constitutes the underlying system of linear equations for the recovery. For arranging this system, we define N vectors

$$\boldsymbol{h}_u = [h(\boldsymbol{g}_u, 0), h(\boldsymbol{g}_u, 1), \ldots, h(\boldsymbol{g}_u, L-1)]^T \tag{12}$$

with $u \in \{1, 2, \ldots, N\}$. Each of them is of length L and contains one RIR on the virtual sampling grid. The solution vector $\boldsymbol{h} \in \mathbb{R}^U$ of the modeled system of linear equations is formed by the concatenation

$$\boldsymbol{h} = \left[\boldsymbol{h}_1^T, \boldsymbol{h}_2^T, \ldots, \boldsymbol{h}_N^T\right]^T. \tag{13}$$

Every value of $h(\boldsymbol{g}, n)$ covering the volume of interest is included in $\boldsymbol{h}$. In Eq. (13) a dimension-wise concatenation along each spatial dimension is reasonable. For example, by concatenating at first along the x dimension, and then along the y and z dimensions in succession, each value of $h(\boldsymbol{g}, n)$ can be assigned to a distinct element of $\boldsymbol{h}$ with index

$$l(\boldsymbol{g}, n) = g_x L + g_y XL + g_z XYL + n + 1. \tag{14}$$

A sampling vector $\phi(n) \in \mathbb{R}^U$ is modeled as

$$\boldsymbol{\phi}(n) = \left[\boldsymbol{w}_1^T(n), \boldsymbol{w}_2^T(n), \ldots, \boldsymbol{w}_N^T(n)\right]^T, \tag{15}$$

where each of the vectors

$$\boldsymbol{w}_u(n) = \varphi(\boldsymbol{r}(n), \boldsymbol{r}_{\boldsymbol{g}_u})\, \boldsymbol{s}(n) \tag{16}$$

corresponds to the virtual grid RIR $\boldsymbol{h}_u$ and consists of the source signal vector

$$\boldsymbol{s}(n) = [s(n), s(n-1), \ldots, s(n-L+1)]^T, \tag{17}$$

which is individually weighted by the interpolation function. Using Eqs. (11), (13), and (15), each sample of the dynamic microphone is coded by the projection

$$x(n) = \boldsymbol{\phi}^T(n)\boldsymbol{h} + \eta(n). \tag{18}$$

Generally, a larger support of the interpolation function allows a sampling vector to cover a larger area in space, and, thus, enables one sample to give information on a larger volume. The definitions of the measurement vector

$$\boldsymbol{x} = [x(0), x(1), \ldots, x(M-1)]^T, \tag{19}$$

the measurement noise vector

$$\boldsymbol{\eta} = [\eta(0), \eta(1), \ldots, \eta(M-1)]^T, \tag{20}$$

and the $M \times U$ sampling matrix

$$\boldsymbol{A} = \left[\boldsymbol{\phi}(0), \boldsymbol{\phi}(1), \ldots, \boldsymbol{\phi}(M-1)\right]^T \tag{21}$$

lead to the system of linear equations

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{\eta}. \tag{22}$$

In order to ensure a unique solution to Eq. (22) in the least-squares sense, the matrix $\boldsymbol{A}$ must have full column rank which means $\mathrm{rank}(\boldsymbol{A}) = U$. This is equivalent to the requirement on $\boldsymbol{A}$ to have U linearly independent sampling vectors in the rows, which, in turn, are specified by the excitation signal, the position of the moving microphone, and the chosen interpolation method. Thus, if $\mathrm{rank}(\boldsymbol{A}) = U$, the linear system (22) is not underdetermined and its unique least-squares solution yields the estimate of $h(\boldsymbol{g}, n)$.

An equivalent and more detailed description for the measurement vector is

$$\boldsymbol{x} = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S} \boldsymbol{h}_u + \boldsymbol{\eta}(n), \tag{23}$$

where $\boldsymbol{\Phi}_u \in \mathbb{R}^{M\times M}$ are diagonal matrices, each of them stacking all M interpolation coefficients for the uth virtual grid RIR,

$$\boldsymbol{\Phi}_u = \mathrm{diag}\{\varphi(\boldsymbol{r}(0), \boldsymbol{r}_{g_u}), \ldots, \varphi(\boldsymbol{r}(M-1), \boldsymbol{r}_{g_u})\}, \tag{24}$$

and $\boldsymbol{S} \in \mathbb{R}^{M\times L}$ is the convolution matrix of the source signal,

$$\boldsymbol{S} = \left[\boldsymbol{s}(0), \boldsymbol{s}(1), \ldots, \boldsymbol{s}(M-1)\right]^T. \tag{25}$$

This compact representation reveals that the sampling matrix can be rewritten as highly structured block matrix

$$\boldsymbol{A} = \left[\boldsymbol{\Phi}_1 \boldsymbol{S}, \boldsymbol{\Phi}_2 \boldsymbol{S}, \ldots, \boldsymbol{\Phi}_N \boldsymbol{S}\right], \tag{26}$$

consisting of N repetitions of $\boldsymbol{S}$ along the columns. The Toeplitz structure of the uth convolution matrix is distorted by the diagonal matrix $\boldsymbol{\Phi}_u$ which scales the rows of $\boldsymbol{S}$ differently. Assuming that the chosen interpolation function satisfies the partition of unity condition

$$\sum_{u=1}^{N} \varphi(\boldsymbol{r}(n), \boldsymbol{r}_{g_u}) = 1 \tag{27}$$

for constant n, the diagonal matrices fulfill

$$\sum_{u=1}^{N} \boldsymbol{\Phi}_u = \boldsymbol{I}_{M\times M}, \tag{28}$$

where $\boldsymbol{I}_{M\times M}$ denotes the $M \times M$ identity matrix.

C. Extension to microphone arrays

The previous considerations regarding one moving microphone can be easily extended to a sampling procedure which involves one moving array with Q mounted microphones. This requires the additional measurement of the array orientation relative to the room axes, e.g., using a gyroscope, and also the tracking of the coordinates $\bar{\boldsymbol{r}}(n) = [\bar{r}_x(n), \bar{r}_y(n), \bar{r}_z(n)]^T$ referring to one particular point on the array which does not change its relative position to the microphone positions $\boldsymbol{r}_q(n)$ $(q \in \{1, \ldots, Q\})$ in case of motion. For the purpose of describing the initial orientation of the array, let

$$\boldsymbol{p}_q = \boldsymbol{r}_q(0) - \bar{\boldsymbol{r}}(0) \tag{29}$$

define the vector originating in $\bar{\boldsymbol{r}}(0)$ and pointing to the position $\boldsymbol{r}_q(0)$ of microphone q on that array. Then, provided that the array is rigid, the positions of the microphones during the measurement process are given by

$$\boldsymbol{r}_q(n) = \bar{\boldsymbol{r}}(n) + \boldsymbol{R}_z(n)\boldsymbol{R}_y(n)\boldsymbol{R}_x(n)\boldsymbol{p}_q, \tag{30}$$

where $\boldsymbol{R}_x(n)$, $\boldsymbol{R}_y(n)$, $\boldsymbol{R}_z(n)$ are rotation matrices along each spatial axis mapping $\boldsymbol{p}_q$ from the initial state to the array orientation at n, in order to cover the rotational degrees of freedom of the microphone array in 3D space.

The choice of Q depends on the application and determines the tradeoff between calibration effort and sampling time. The microphones measuring at positions (30) allow for building up Q sampling vectors $\boldsymbol{\phi}_q(n)$, in total contributing Q equations to the linear system (22) for each instant of time. However, each of these microphones requires additional equalization, synchronization, and specification of $\boldsymbol{p}_q$ subject to $\bar{\boldsymbol{r}}(0)$.

IV. ERROR ANALYSIS

This section is devoted to an error analysis. We first describe some potential error sources which may perturb the modeled linear system (22), built up by the proposed sampling procedure for one moving microphone. Further, we give some considerations about the influence of the Doppler effect in Sec. IV B. When the microphone is moved during recording through the volume of interest, one may expect that, under noisy conditions, also the choice of the trajectory has a significant influence on the accuracy of the estimated RIRs. In Sec. IV C we analyze how variations of the trajectory affect the estimation error. Initially, we show under some idealized conditions how the choice of the trajectory directly influences the estimation error, though the interpolation operation is assumed to be exact. Then we point out that, for a given set of sampling positions in space, the temporal order of taking measurements has no effect on the accuracy.

A. Sources of error

The linear system (22) takes additive measurement noise, caused by the microphone, into account. This is modeled with the additive noise vector $\boldsymbol{\eta}$.

Beyond that, some systematic errors may emerge through inaccuracies during the positional tracking of the microphone. This systematic perturbation leads to multiplicative noise and is taken into account by the formulation

$$\boldsymbol{A} = \left[(\boldsymbol{\Phi}_1 + \tilde{\boldsymbol{\Phi}}_1)\boldsymbol{S}, \ldots, (\boldsymbol{\Phi}_N + \tilde{\boldsymbol{\Phi}}_N)\boldsymbol{S}\right] \tag{31}$$

with the error diagonal matrices $\tilde{\boldsymbol{\Phi}}_u$. Substituting Eq. (31) into Eq. (22) reveals that the additional noise term is given by $\sum_{u=1}^{N} \tilde{\boldsymbol{\Phi}}_u \boldsymbol{S} \boldsymbol{h}_u$. In case the chosen interpolation function constitutes the partition of unity (28), then

$$\sum_{u=1}^{N} \tilde{\boldsymbol{\Phi}}_u = 0_{M\times M}. \tag{32}$$

Due to the finite number of N virtual sampling points to be recovered in space, the spatial interpolation of $h(\boldsymbol{r}(n), n)$ is not ideal in practice, which also leads to systematic errors. To keep the interpolation error small, spatial oversampling is suggested in Sec. III A for the reverse interpolation problem

of reconstructing continuous data $h(\boldsymbol{r}, t)$ from equidistant samples.

B. Doppler effect

When considering a moving receiver, the Doppler effect causes a frequency shift on the observed signal. For the emitted frequency ω_s and the speed of movement ν_r of the receiver, the sensed frequency is

$$\omega_x = \omega_s \left(1 + \frac{\nu_r}{c_0}\right). \tag{33}$$

Here, the sign of ν_r is positive when the receiver is moving towards the source and negative when the receiver is moving away from the source. In the following subsection, we explain under which conditions the Doppler effect may be considered negligible for the reconstruction. However, a higher velocity of the moving microphone will induce more noise caused by air flow in practice.

For the dynamic measurement procedure proposed in Ref. 19, the measurement signal of the moving microphone is processed at one stretch. The whole observed signal is transformed into frequency domain, where, taking the Doppler effect into account, an inverse projection yields the spectrum of spatio-temporal RIRs along the trajectory. The key factor for the reconstruction in Ref. 19 is the speed of movement which has to be constant in order to obtain consistent frequency shifts. In contrast, the general reconstruction method proposed in this paper treats each sample of the measurement signal $x(n)$ independently. For a certain sampling time n, the value sampled at tracked position $\boldsymbol{r}(n)$ is assumed to be the same value that would be sampled simultaneously by a stationary microphone at that particular position. The duration of the process in the microphone for taking one sample may be assumed to be so small that the actual effect of integration over space induced by the movement, i.e., spatial averaging of the measured sound pressure, is negligible for velocities performed by a human or robot. This means that each measured value $x(n)$ is considered to be one snapshot in time and space which contributes one equation, Eq. (18), to the linear system (22). Accordingly, the values of the measurement vector $\boldsymbol{x}$ can even be arbitrarily permuted, provided that the same permutation is applied to the sampling vectors in the rows of the system matrix $\boldsymbol{A}$.

Of course, the Doppler effect affects the highest frequencies that fall beyond the cutoff frequency of the anti-aliasing prefilter in case the microphone moves toward the source. Vice versa, when the microphone is receding from the source, frequencies beyond the actual cutoff would pass. For both cases, a little amount of oversampling, which will be present in practice anyway, is the solution. For example, when the microphone is moving away from the source with up to $\nu_r = 17\,\mathrm{m/s}$ and $c_0 = 340\,\mathrm{m/s}$, then, considering Eq. (33), in addition to temporal oversampling by factor 1.05, also spatial oversampling by the same factor is required for the virtual grid in order to avoid aliasing.

When a multiresolution scheme is used for the recovery of the spatial grid as described in Sec. V B, the Doppler effect is also existent. Here, the aim is to recover distinct subbands of the PAF which requires band-pass filtering of the measurement signal $x(n)$. The individual frequency bands are successively recovered by solving one linear system for each. The motion of the microphone toward and away from the sound source leads to frequency displacements between adjoining frequency bands according to Eq. (33). Analogously, in case of appropriate oversampling in space, the subsequent synthesis of the entire PAF spectrum from the distinct subbands obtains the full recovery result respecting all concerned frequencies without aliasing.

C. Error analysis for spectrally flat excitation

Let us consider the following scenario. The source signal $s(n)$ is either white noise with variance σ_s^2 or a perfect sequence with period length L and autocorrelation

$$r_{ss}(m) = \sigma_s^2\, \delta(m \bmod L), \tag{34}$$

where σ_s^2 represents the signal power. For white noise, the signal matrix $\boldsymbol{S}$ satisfies

$$E\{\boldsymbol{S}\boldsymbol{S}^T\} = L\sigma_s^2 \boldsymbol{I}_{M\times M}. \tag{35}$$

For one period of a perfect sequence, we have

$$\boldsymbol{S}\boldsymbol{S}^T = L\sigma_s^2 \boldsymbol{I}_{M\times M}. \tag{36}$$

The additive measurement noise $\boldsymbol{\eta}$ is modeled as independent and identically distributed white noise with covariance matrix $\boldsymbol{R}_{\eta\eta} = E\{\boldsymbol{\eta}\boldsymbol{\eta}^T\} = \sigma_\eta^2 \boldsymbol{I}_{M\times M}$. The noise term is uncorrelated to the parameters in $\boldsymbol{h}$. The parameters in $\boldsymbol{h}$ are assumed to have zero mean (cf. Sec. II A) and variance σ_h^2, so the autocovariance matrix of $\boldsymbol{h}$ is $\boldsymbol{R}_{hh} = E\{\boldsymbol{h}\boldsymbol{h}^T\} = \sigma_h^2 \boldsymbol{I}_{U\times U}$. Regarding this model, the use of the linear MMSE estimator for a system matrix $\boldsymbol{A}$ yields the error covariance matrix[26]

$$\boldsymbol{R}_{ee} = E\left\{[\hat{\boldsymbol{h}} - \boldsymbol{h}][\hat{\boldsymbol{h}} - \boldsymbol{h}]^T\right\} = \sigma_h^2 \left[\boldsymbol{I}_{U\times U} + \frac{\sigma_h^2}{\sigma_\eta^2}\boldsymbol{A}^T\boldsymbol{A}\right]^{-1}. \tag{37}$$

The variances of the estimated parameters $\hat{\boldsymbol{h}}$ are located at the principal diagonal of $\boldsymbol{R}_{ee}$ and the mean squared error (MSE) of the estimation is given by

$$\mathrm{MSE} = \mathrm{tr}\{\boldsymbol{R}_{ee}\}. \tag{38}$$

Substituting Eq. (37) into Eq. (38) and using the relationship

$$\mathrm{tr}\{\left[\boldsymbol{I}_{S\times S} + \boldsymbol{B}\boldsymbol{C}^T\right]^{-1}\} = \mathrm{tr}\{\left[\boldsymbol{I}_{W\times W} + \boldsymbol{C}^T\boldsymbol{B}\right]^{-1}\} - (W - S) \tag{39}$$

for matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ of size $S \times W$, the MMSE for the model with $M = U$ can be described as

$$\mathrm{MMSE} = \sigma_h^2 \mathrm{tr}\left\{\left[\boldsymbol{I}_{M\times M} + \frac{\sigma_h^2}{\sigma_\eta^2}\boldsymbol{A}\boldsymbol{A}^T\right]^{-1}\right\}. \tag{40}$$

Exploiting the block representation (26), the square matrix $\boldsymbol{A}\boldsymbol{A}^T$ is given by

$$\boldsymbol{A}\boldsymbol{A}^T = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S}\boldsymbol{S}^T \boldsymbol{\Phi}_u. \tag{41}$$

Due to Eqs. (35) and (36), respectively, the estimation error becomes

$$\mathrm{MMSE} = \sigma_h^2 \mathrm{tr}\left\{ \left[\boldsymbol{I}_{M\times M} + \frac{\sigma_h^2}{\sigma_\eta^2} L\sigma_s^2 \sum_{u=1}^{N} \boldsymbol{\Phi}_u^2 \right]^{-1} \right\}. \tag{42}$$

With reference to Eq. (24), the estimation error in Eq. (42) corresponds to

$$\mathrm{MMSE} = \sigma_h^2 \sum_{n=0}^{M-1} \frac{1}{1 + \frac{\sigma_h^2}{\sigma_\eta^2} L\sigma_s^2 \sum_{u=1}^{N} \varphi^2\left(\boldsymbol{r}(n), \boldsymbol{r}_{g_u}\right)}. \tag{43}$$

In presence of measurement noise, the following observations can be made by reference to the expression (43), independent from the actual interpolation accuracy.

- The MSE depends on the interpolation coefficients $\varphi(\boldsymbol{r}(n), \boldsymbol{r}_{g_u})$, and, thus, on the trajectory $\boldsymbol{r}(n)$, the modeled grid $\mathcal{G}$, and the chosen interpolation function.
- The larger the sum of the squared coefficients,

$$\sum_{u=1}^{N} \varphi^2(\boldsymbol{r}(n), \boldsymbol{r}_{g_u}), \tag{44}$$

  becomes for any measurement position, the smaller the error will be.
- Given that $\sum_{u=1}^{N} \varphi(\boldsymbol{r}(n), \boldsymbol{r}_{g_u}) = 1$ according to Eq. (27) and assuming nonnegative coefficients as for the linear interpolation, the term (44) will be maximal when the coefficients are either zero or one. This means that it would be optimal to sample only on the virtual grid positions. The worst case occurs when the partition of unity yields consistently equal coefficients, which corresponds to sampling on the middle between the grid positions.
- Consider an interpolation with positive and negative coefficients, e.g., the Lagrange interpolation. Then, a larger support of the interpolation function may increase the amount of Eq. (44), and, thus, reduce the estimation error.
- Using the sinc function with infinite support and assuming $N \to \infty$, then, the estimation error is independent of the trajectory, since Eq. (44) always equals 1. A truncation of the sinc function leads, again, to a larger estimation error for larger distances of the measuring positions to the next virtual grid points.

Moreover, the position-dependend terms (44) may also be used to determine the condition number of the resulting system matrix $\boldsymbol{A}$. By applying the singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, Eq. (41) leads to

$$\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S}\boldsymbol{S}^T \boldsymbol{\Phi}_u \tag{45}$$

and thus

$$\boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^T = L\sigma_s^2 \sum_{u=1}^{N} \boldsymbol{\Phi}_u^2, \tag{46}$$

where $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ are orthogonal matrices with $\boldsymbol{U}^T = \boldsymbol{U}^{-1}$, $\boldsymbol{V}^T = \boldsymbol{V}^{-1}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ is a diagonal matrix containing the singular values of $\boldsymbol{A}$. Obviously, the spectrally flat excitation sequence forms the right-singular vectors of $\boldsymbol{A}$ according to $\boldsymbol{V} = 1/\sqrt{L\sigma_s^2}\boldsymbol{S}$. Following Eq. (46), $\boldsymbol{U}$ is the identity matrix or a permutation matrix, respectively, since it must be orthogonal and must not change the diagonality of $\boldsymbol{\Sigma}^2$ correspondent to the right-hand term of the equation. Further, the set of singular values in $\boldsymbol{\Sigma}$ is

$$\Lambda = \left\{ \zeta \in \mathbb{R}_{>0} \,|\, \zeta = \sqrt{L\sigma_s^2 \sum_{u=1}^{N} \varphi^2(\boldsymbol{r}(n), \boldsymbol{r}_{g_u})} \right\}. \tag{47}$$

Accordingly, the condition number of $\boldsymbol{A}$ is

$$\kappa(\boldsymbol{A}) = \frac{\max\{\Lambda\}}{\min\{\Lambda\}}, \tag{48}$$

which is determined by the trajectory, the modeled grid in space, and the interpolation method. Let us consider the worst case for a simple trilinear interpolation on a 3D grid. This consists of a trajectory including both extreme positions, which are exactly on one grid point, and exactly in the middle between eight grid points. The upper bound for the condition of the problem is then given by

$$\kappa_{\mathrm{lin}}(\boldsymbol{A}) \leq \frac{1}{\sqrt{8(0.125^2)}} = \sqrt{8} \approx 2.83. \tag{49}$$

With Eqs. (15) and (16), the sampling vector of each sample covers a certain section in space depending on the microphone position. For an excitation sequence $s(n)$ with uncorrelated samples, the temporal order of measurement positions $\boldsymbol{r}(n)$ could be arbitrarily permuted without affecting the sound-field reconstruction. This will be shown in the following paragraph.

A permutation $P : \{0, \ldots, M-1\} \to \{0, \ldots, M-1\}$ of $\boldsymbol{r}(n)$ leads to a shift of the interpolation function in Eq. (16) and modifies the weighting vectors to

$$\boldsymbol{w}_u(n) = \varphi(\boldsymbol{r}(P(n)), \boldsymbol{r}_u)\, s(n). \tag{50}$$

The order of the diagonal elements in $\boldsymbol{\Phi}_u$ is changed according to P, consistently for all $u \in \{1, \ldots, N\}$. Both the MMSE according to Eq. (42) and the condition number (48) are unaffected by this permutation of diagonal elements. Consequently, the order of spatial measuring points is arbitrary concerning the quality of reconstruction. However, several permutations P might lead to numerically induced differences in solving the linear system (22).

As an example, consider for the same statistical model the simplest case with no interpolation, i.e., the microphone samples exactly on the grid positions for all n, so $\boldsymbol{r}(n) \in \mathcal{G}$. According to this, the weighting vectors are

$$\boldsymbol{w}_u(n) = \delta(\boldsymbol{r}(P(n)) - \boldsymbol{r}_u)\, \boldsymbol{s}(n). \tag{51}$$

Independent of permutation P, each sampling vector $\boldsymbol{\phi}(n)$ contains $(U - L)$ zeros and only one instance of the excitation vector $\boldsymbol{s}(n)$. Only the one virtual position, where the microphone is located at n, is weighted by 1, all the others are zero-weighted. A variation of P simply induces a position shift of $\boldsymbol{s}(n)$ in $\boldsymbol{\phi}(n)$ by blocks of length pL with $p \in \mathbb{Z}$. So the MMSE is constant for a varying P, since $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\phi}^T(n)\boldsymbol{\phi}(n)$ is always a diagonal matrix where the i-th diagonal element is given by

$$[\boldsymbol{A}\boldsymbol{A}^T]_{i,i} = \boldsymbol{s}^T(i-1)\boldsymbol{s}(i-1). \tag{52}$$

V. RECOVERY ON LARGE SPATIAL GRIDS

When applying Eq. (22) for recovery on large spatial grids with high bandwidth, the linear system might become too large for practical applications due to limitations in computational power and memory. According to Eqs. (4) and (5), the number of unknowns grows proportionally to f_c^4 for the 4D case where the time dimension and each of the three space dimensions are involved. For example, to recover the sound field up to 8 kHz with RIR length $L = 1000$ inside a cube of 0.2 m edge length, the system to be solved involves at least 10^6 unknowns.

In Sec. V A, we propose a strategy to decouple the time dimension from the problem (22) by using perfect sequences and solving for a time filtered version of the PAF. This substantially reduces the computational complexity for recovery. In Sec. V B, we present a multiresolution recovery scheme which exploits the inherent dependence of the spatial spectrum on the spectrum of the time dimension. Finally, in Sec. V C we briefly demonstrate that our dynamic approach for sound-field recovery is particularly suited to the compressed-sensing framework.

A. Excitation using perfect sequences

For sound-field recovery on large spatial grids, we propose the periodic excitation by pseudo-random noise $s(n)$ with perfect autocorrelation according to Eq. (34). This corresponds to $s(n)$ having ideally flat magnitude spectrum. Let $\tilde{s}(n)$ be one period of the perfect sequence with length L. In order to fit in with the period length of the chosen pseudo-random sequence, the assumed length of the grid RIRs is also set to L. Anyway, the RIRs can be truncated to arbitrary size after recovery.

Corresponding to Eq. (17), we define L excitation vectors

$$\tilde{\boldsymbol{s}}(n) = [\tilde{s}(n \bmod L), \ldots, \tilde{s}((n - L + 1) \bmod L)]^T \tag{53}$$

for $n \in \{0, \ldots, L-1\}$ for a steady-state situation. The vectors (53) cover all different circularly shifted versions of the chosen pseudo-random noise and build up the circular convolution matrix $\boldsymbol{S} \in \mathbb{R}^{L \times L}$ for one period of excitation as

$$\boldsymbol{S} = \left[\tilde{\boldsymbol{s}}(0), \tilde{\boldsymbol{s}}(1), \ldots, \tilde{\boldsymbol{s}}(L-1)\right]^T. \tag{54}$$

The excitation by R periods of $\tilde{s}(n)$ allows for measuring $M = RL$ samples and leads to the $RL \times L$ convolution matrix

$$\boldsymbol{S}_R = \underbrace{\left[\boldsymbol{S}^T, \ldots, \boldsymbol{S}^T\right]}_{R \text{ times}}^T. \tag{55}$$

Corresponding to Eq. (23), the resulting measurement process is formulated as

$$\boldsymbol{x} = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S}_R \boldsymbol{h}_u + \boldsymbol{\eta}(n). \tag{56}$$

In order to decouple the time dimension, a diagonal structure in the system matrix is needed. Unfortunately, using the system matrix $\breve{\boldsymbol{A}} = [\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_N]$ and solving for $\breve{\boldsymbol{h}} = [[\boldsymbol{S}_R\boldsymbol{h}_1]^T, \ldots, [\boldsymbol{S}_R\boldsymbol{h}_N]^T]^T$ is not sufficient, but this representation may give useful insights for the following derivation. Indeed, this way of partitioning decomposes the large problem into smaller ones. The vector $\breve{\boldsymbol{h}} \in \mathbb{R}^{MN}$ then contains

$$x_u(n) = s(n) * h(\boldsymbol{g}_u, n) \tag{57}$$

for all $u \in \{1, \ldots, N\}$, i.e., the ideally observed signals at the positions on the modeled grid in space. According to Eq. (34), the circularly shifted versions of one period $\tilde{s}(n)$ are uncorrelated to each other. Equivalently,

$$s(n) * \tilde{s}(-n) = \gamma \sum_{k=0}^{R-1} \delta(n - kL), \tag{58}$$

where γ is the energy of one period with

$$\gamma = L\sigma_s^2 = \sum_{n=0}^{L-1} |\tilde{s}(n)|^2. \tag{59}$$

Using Eqs. (57) and (58), R repetitions of the sought grid RIRs could be obtained by calculating

$$\boldsymbol{h}_R = \gamma^{-1} \mathrm{diag}\{\boldsymbol{S}^T, \ldots, \boldsymbol{S}^T\}\breve{\boldsymbol{h}}, \tag{60}$$

since

$$\begin{aligned} \tilde{s}(-n) * x_u(n) &= \tilde{s}(-n) * s(n) * h(\boldsymbol{g}_u, n) \\ &= \gamma \sum_{k=0}^{R-1} \delta(n - kL) * h(\boldsymbol{g}_u, n). \end{aligned} \tag{61}$$

The correlation step involving $x_u(n)$ and $\tilde{s}(n)$ and resulting in Eq. (61) is the same as for the stationary measurement of $h(\boldsymbol{g}_u, n)$ using perfect sequences: in practice, the subsequent averaging over all obtained repetitions would lead to an estimate of the particular RIR under noisy conditions. However, for the proposed dynamic procedure, the samples $x_u(n)$ are available not by themselves, but only indirectly by the projection results (18): the dynamic samples $x(n)$ are considered to be composed of weighted sums of uniformly taken samples $x_1(n), \ldots, x_N(n)$ that ideally would be measured on the virtual grid in space. Therefore, the block diagonal structure

of $\breve{\boldsymbol{A}} \in \mathbb{R}^{M\times MN}$ inherently leads to an underdetermined problem. Nevertheless, the periodicity of $s(n)$ allows for solving the system for a solution respecting only one period of excitation instead of R. Then, due to sub-block shifts inside the diagonal matrices $\boldsymbol{\Phi}_u$, the averaging step over R periods, as mentioned above for the stationary measurement of one RIR, may be embedded into the problem of the dynamic measurement of N RIRs. A determined problem for the dynamic procedure is possible for $R \geq N$. An illustration of the considered shifts is given in Fig. 2(a).

The mathematical derivation of the time-decoupling is based on the orthogonality of the excitation vectors (53) forming the matrix $\boldsymbol{S}$. Following Eq. (58),

$$\boldsymbol{S}^T\boldsymbol{S} = \gamma \boldsymbol{I}_{L\times L}. \tag{62}$$

Relating to the periodic excitation in $\boldsymbol{S}_R$, this ensures the central property

$$\boldsymbol{S}_R\boldsymbol{S}^T = \gamma \underbrace{\left[\boldsymbol{I}_{L\times L}, \ldots, \boldsymbol{I}_{L\times L}\right]^T}_{R\,\text{times}}, \tag{63}$$

which is the key to the decoupling of the time dimension. Note that $\boldsymbol{S}_R$ must be set up for the steady state avoiding zero-padding for the beginning and the end of the sampling procedure. Using Eqs. (56) and (62), the measurement process can be reformulated as

$$\boldsymbol{x} = \sum_{u=1}^{N} \boldsymbol{\Phi}_u \boldsymbol{S}_R \boldsymbol{S}^T \gamma^{-1} \boldsymbol{S}\boldsymbol{h}_u + \boldsymbol{\eta}(n). \tag{64}$$

By defining the modified $RL \times NL$ system matrix

$$\tilde{\boldsymbol{A}} = \left[\boldsymbol{\Phi}_1 \boldsymbol{S}_R \boldsymbol{S}^T \gamma^{-1}, \ldots, \boldsymbol{\Phi}_N \boldsymbol{S}_R \boldsymbol{S}^T \gamma^{-1}\right] \tag{65}$$

$$= \left[\boldsymbol{\Phi}_1 \underbrace{\left[\boldsymbol{I}_{L\times L}, \ldots, \boldsymbol{I}_{L\times L}\right]^T}_{R\,\text{times}}, \ldots, \boldsymbol{\Phi}_N \underbrace{\left[\boldsymbol{I}_{L\times L}, \ldots, \boldsymbol{I}_{L\times L}\right]^T}_{R\,\text{times}}\right] \tag{66}$$

and concatenating the transformed grid RIRs $\boldsymbol{S}\boldsymbol{h}_u$ to the solution vector

$$\tilde{\boldsymbol{h}} = \left[\left[\boldsymbol{S}\boldsymbol{h}_1\right]^T, \ldots, \left[\boldsymbol{S}\boldsymbol{h}_N\right]^T\right]^T, \tag{67}$$

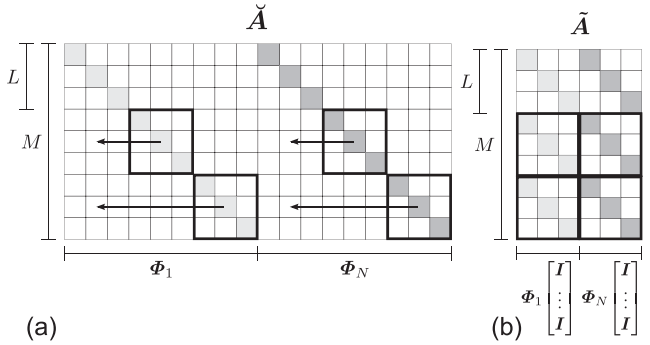

FIG. 2. Illustration of system matrices. (a) The system matrix $\breve{\boldsymbol{A}}$ consisting of the column-wise concatenation of interpolation matrices $\boldsymbol{\Phi}_u$ and (b) the system matrix $\tilde{\boldsymbol{A}}$ consisting of shifted segments of $\boldsymbol{\Phi}_u$ that allow for the decoupling of the time dimension.

the new system of linear equations

$$\boldsymbol{x} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{h}} + \boldsymbol{\eta} \tag{68}$$

is obtained. This way of proceeding inherently requires of the trajectory to fulfill

$$\boldsymbol{r}(n) \notin \{\boldsymbol{r}(n-L), \boldsymbol{r}(n-2L), \boldsymbol{r}(n-3L), \ldots\}, \tag{69}$$

in order to avoid linearly dependent rows in $\tilde{\boldsymbol{A}}$. According to Eq. (67), the solution of the new problem is the PAF filtered by $\boldsymbol{S}$ along the time dimension, i.e., the ideally observed signals on the spatial grid for one period of excitation. As sketched in Fig. 2, the system matrix $\tilde{\boldsymbol{A}}$ is the intended block shifted and clipped version of $\breve{\boldsymbol{A}}$ consisting of $R \times N$ blocks of $L \times L$ diagonal matrices. Each of these diagonal matrices carries the interpolation coefficients of one grid RIR for one of the R periods of excitation. In relation to $\tilde{\boldsymbol{h}}$, the dimensionality of the solution vector $\tilde{\boldsymbol{h}} \in \mathbb{R}^{NL}$ is reduced by factor R.

In total, $\tilde{\boldsymbol{A}}$ possess only up to MN nonzero elements, which means a saving of the entire time dimension. By comparison, the system matrix $\boldsymbol{A}$ of the original problem formulation (22) contains up to MNL nonzero elements. The sparse block structure of $\tilde{\boldsymbol{A}}$ allows to decouple the time dimension by decomposing the large system of size $RL \times NL$ into L smaller problems

$$\boldsymbol{x}_\ell = \tilde{\boldsymbol{A}}_\ell \tilde{\boldsymbol{h}}_\ell + \boldsymbol{\eta}_\ell, \quad \ell \in \{1, \ldots L\}, \tag{70}$$

with the partial vectors

$$\boldsymbol{x}_\ell = [x(\ell-1), x(\ell-1+L), \ldots, x(\ell-1+RL)]^T, \tag{71}$$

$$\tilde{\boldsymbol{h}}_\ell = \left[\tilde{\boldsymbol{s}}^T(\ell-1)\boldsymbol{h}_1, \ldots, \tilde{\boldsymbol{s}}^T(\ell-1)\boldsymbol{h}_N\right]^T, \tag{72}$$

$$\boldsymbol{\eta}_\ell = [\eta(\ell-1), \eta(\ell-1+L), \ldots, \eta(\ell-1+RL)]^T, \tag{73}$$

each of length R and containing every Lth value of the corresponding complete vectors, and the sub-system matrix $\tilde{\boldsymbol{A}}_\ell \in \mathbb{R}^{R\times N}$ whose element in row i and column j is given by

$$\left[\tilde{\boldsymbol{A}}_\ell\right]_{i,j} = \left[\tilde{\boldsymbol{A}}\right]_{(i-1)L+\ell,(j-1)L+\ell} \tag{74}$$

$$= \varphi(\boldsymbol{r}((i-1)L+\ell), \boldsymbol{r}_{\boldsymbol{g}_j}). \tag{75}$$

The size of the sub-problems (70) is independent of L. The length L only determines the number of sub-problems to be solved. Hence, the number of unknowns in each partial system grows proportionally to only f_c^3. The saving of the one dimension reduces the computational cost for recovery significantly.

The solving of all L problems (70) determines the vector $\tilde{\boldsymbol{h}}$ that consists of the representation of each grid RIR $\boldsymbol{h}_u$ by means of the orthogonal basis in $\boldsymbol{S}$ which comprises each circularly shifted version of $\tilde{s}(n)$. Subsequently, according to Eq. (62) the virtual grid RIRs in $\boldsymbol{h}$ are obtained by calculating the inverse transformations,

$$\boldsymbol{h} = \gamma^{-1}\text{diag}\{\boldsymbol{S}^T, \ldots, \boldsymbol{S}^T\}\tilde{\boldsymbol{h}}. \tag{76}$$

This two-step recovery of the PAF does not suffer from coloring of observation noise when solving for least-squares, since, for each RIR, $\boldsymbol{S}$ only changes the phases in the spectrum and scales its magnitude by constant factor $\sqrt{\gamma^{-1}}$.

B. Multiresolution recovery

From a physical point of view, propagative sound waves follow the dispersion relation

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c_0^2}, \tag{77}$$

which gives a relationship between the spatial frequencies k_x, k_y, k_z in rad m^{-1} and the temporal angular frequency $\omega = 2\pi f$ in rad s^{-1} for the continuous case.[18] The velocity of the waves is $c_0 = \omega/\breve{k}$ with the angular wavenumber $\breve{k} = |\boldsymbol{k}|$ and the wave vector $\boldsymbol{k} = [k_x, k_y, k_z]^T$. Since air is a non-dispersive medium for frequencies within the human hearing range, c_0 is independent of ω. So, the speed of sound is only a function of atmospheric conditions inside the closed room, e.g., temperature and pressure, which are assumed to be constant during the measurement process according to the LTI model. In consequence, Eq. (77) reveals a direct connection between the temporal and spatial frequencies. Moreover, when the PAF is bandlimited in time domain to ω_c, then it is also bandlimited in the spatial domain to $\breve{k}_c = \omega_c/c_0$. This may be exploited for a multiresolution recovery scheme of the virtual grid in space.

Consider a very coarse virtual grid with large spatial intervals allowing for the reconstruction of very large wavelengths. This grid only involves a little number of RIRs inside a bounded area, so the recovery of the sound field at low spatial frequencies is already possible, using just a few samples acquired after a short measurement time. Or conversely, by fixing the sampling time, at the cost of being restricted to lower frequencies a coarser virtual grid allows for more samples per grid RIR than a finer one, which makes the recovery more robust against noise. Therefore, we propose a multigrid recovery by solving for subbands of the 4D spectrum of the PAF.

For each of V grid stages $v \in \{1, \ldots, V\}$, a separate linear system of equations

$$\boldsymbol{x}^{(v)} = \boldsymbol{A}^{(v)}\boldsymbol{h}^{(v)} + \boldsymbol{\eta}^{(v)} \tag{78}$$

is built up, where $\boldsymbol{h}^{(v)} \in \mathbb{R}^{LN^{(v)}}$ contains successively more RIRs on a successively finer spatial grid $\boldsymbol{g}^{(v)} \in G^{(v)}$ inside a constant volume of interest with origin $\boldsymbol{r}_0$. Thus, it follows $|G^{(v-1)}| < |G^{(v)}|$ and $N^{(v-1)} < N^{(v)}$. The PAF is still assumed to be bandlimited to ω_c and sampled with $\omega_s > 2\omega_c$ in the time dimension. Some positions on the finest grid respecting the entire bandwidth up to ω_c might be disregarded by the chosen trajectory leading to an underdetermined system of linear equations. Due to the proposed multiresolution technique, at virtual grid positions which are underdetermined for the full bandwidth due to insufficient sampling, the recovery of the RIRs is possible at least for some stages at lower frequencies.

For simplification and without loss of generality, in the following, we set $\Delta = \Delta_x = \Delta_y = \Delta_z$. Let $\Delta^{(1)}$ denote the initial spatial sampling interval for the coarsest grid at stage $v = 1$, with $N^{(1)} = 2^d$ RIRs for the d-dimensional case. These initial grid positions frame the entire volume of interest. The virtual grid for each subsequent stage $v > 1$ is modeled with halving intervals

$$\Delta^{(v)} = \frac{1}{2}\Delta^{(v-1)}, \tag{79}$$

so $N^{(v)} = (2^{v-1} + 1)^d$. At the final stage, $\Delta^{(V)}$ is supposed to fulfill the spatial Nyquist-Shannon condition for ω_c. The grids in all other stages are spatially undersampled for ω_c. According to Eq. (77), shrinking the temporal bandwidth involves inherently the reduction of the spatial bandwidth. Thus, we propose to recover the PAF $h(\boldsymbol{g}^{(v)}, n)$ on the spatial grid $\boldsymbol{g}^{(v)}$ at temporal frequencies ω in the subband

$$\omega_c^{(v-1)} \leq \omega < \omega_c^{(v)}, \tag{80}$$

where

$$\omega_c^{(v)} = \frac{\pi c_o}{\Delta^{(v)}} \tag{81}$$

denotes the maximum temporal frequency according to Eq. (77). With Eq. (79), it follows that $\omega_c^{(v)} = 2\omega_c^{(v-1)}$. In order to solve for these consecutive subbands, the measurement signal in $\boldsymbol{x}^{(v)}$ is chosen to be the filtered version of $x(n)$ fulfilling Eq. (80) with initial lower limit $\omega^{(0)} = 0$.

Due to the Doppler effect discussed in Sec. IV B, the cutoff frequencies $\omega_c^{(v)}$ must be smaller than required by Eq. (81), in order to avoid spatial aliasing. Anyway, as mentioned in Sec. IV A, spatial oversampling is advisable in practice, which is equivalent to choosing additionally smaller limits $\omega_c^{(v)}$ for the band-pass filtering of $x(n)$.

Solving for $\boldsymbol{h}^{(v)}$ in Eq. (78) yields the PAF at temporal frequencies (80) and at spatial frequencies k_x, k_y, k_z resulting in wavenumbers $\breve{k}^{(v)}$ of the subband

$$\frac{\omega_c^{(v-1)}}{c_0} \leq \breve{k}^{(v)} < \frac{\omega_c^{(v)}}{c_0}. \tag{82}$$

The solutions of the V stages represent distinct temporal subbands of the PAF on spatial grids with different resolutions. The fusion of these subbands yields the broadband PAF spectrum up to ω_c and is achieved by applying the recursion

$$h(\boldsymbol{g}^{(v)}, n) = h(\boldsymbol{g}^{(v)}, n) + h(\boldsymbol{g}_{\uparrow 2}^{(v-1)}, n) * \varphi(\boldsymbol{g}^{(v)}), \tag{83}$$

where $\boldsymbol{g}_{\uparrow 2}^{(v)}$ denotes the upsampled virtual grid of stage v by factor two along each spatial dimension and $\varphi(\boldsymbol{g}^{(v)})$ is a d-dimensional low-pass filter for interpolating the zero-padded intermediate positions on the finer grid.

Note that the proposed subband recovery scheme on multigrids is similar to the dyadic tree structure of filter banks with two-band frequency split: the stage-wise filtering of $x(n)$ is for subband analysis and $\varphi(\boldsymbol{g}^{(v)})$ is equivalent to a subband synthesis filter. In Fig. 3, an illustration of the

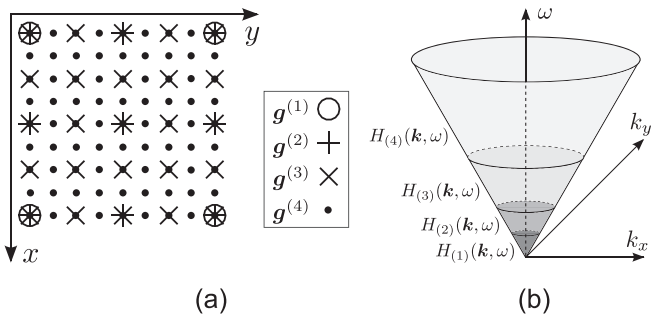

FIG. 3. Outline of multiresolution recovery. (a) The virtual grids $\boldsymbol{g}^{(v)}$ in two-dimensional space and (b) the corresponding PAF subbands $H_{(v)}(\boldsymbol{k},\omega)$ to be recovered.

spatial multigrids and the corresponding PAF subbands is given. Due to the halving grid intervals, the temporal bandwidth is exponentially increasing for subsequent stages.

C. Incorporation into the compressed-sensing framework

Due to the consecutive reflections at the walls, the sound fields in closed rooms inherently exhibit only a few nonzero values for the early part of the time dimension. Furthermore, according to the dispersion relation (77), the sound fields hold highly structured sparsity in frequency domain. Since c_0 is assumed to be constant, Eq. (77) reveals that the entire spectrum of the PAF in 3D space is ideally restricted to the 3D surface of a 4D hypercone along the temporal frequency axis ω.[18] This property is an excellent prerequisite to combine our proposed dynamic measurement procedure with compressed sensing based recovery techniques.

The original system of linear equations (22) is devoted to the problem of recovering the signal $\boldsymbol{h}(\boldsymbol{g},n)$ in vector $\boldsymbol{h}\in\mathbb{R}^U$ from M samples $x(n)$ in measurement vector $\boldsymbol{x}\in\mathbb{R}^M$. The link between both signals is given by the sampling matrix $\boldsymbol{A}\in\mathbb{R}^{M\times U}$. The principle of compressed sensing[27,28] (CS) allows us to find a unique solution for the underdetermined case $M<U$ by merging sampling and compression into one step. From a point of view of signal processing, CS enables the reconstruction of an unknown signal even in the case where the Shannon-Nyquist requirements are not directly met. The general assumption is that $\boldsymbol{h}$ lives in a subspace of dimension $K \ll M$ with respect to an appropriate basis. For instance, a frequency representation of the sound field in $\boldsymbol{h}$ allows one to explicitly reconstruct frequency quartets on the spectral cone according to Eq. (77) which defines the subspace in which the signal actually lives.

Let $\boldsymbol{\Psi}\in\mathbb{C}^{U\times U}$ denote the transformation matrix leading to

$$\boldsymbol{h}=\boldsymbol{\Psi}\boldsymbol{c} \tag{84}$$

with vector $\boldsymbol{c}\in\mathbb{C}^U$ being K-sparse, which means that only a small number of, at most, K coefficients in $\boldsymbol{c}$ is nonzero. Then, the regularized problem

$$\underset{\boldsymbol{c}\in\mathbb{C}^U}{\operatorname{argmin}}\ \|\boldsymbol{x}-\boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{c}\|_{\ell_2}\quad \text{s.t.}\quad \|\boldsymbol{c}\|_0\le \mathrm{K} \tag{85}$$

gives the K-sparse problem, with the pseudo norm $\|\boldsymbol{c}\|_0$ counting the number of nonzero elements in $\boldsymbol{c}$. Solving Eq. (85) requires an extensive combinatorial search over all subsets of $\boldsymbol{c}$ and is NP-hard.[29] However, provided that the measurements are incoherent,[30] the relaxation into an equivalent ℓ_1-norm minimization problem[31–33] enables the solution by convex optimization tools.

VI. EXPERIMENTS AND RESULTS

For the following experiments, we simulated RIRs and microphone measurements by use of the image source method,[22] considering a room of size 5.8 m × 4.15 m × 2.55 m. The reverberation time of the room was chosen as $RT_{60}=0.3$ s. The temporal sampling frequency was $f_s=8$ kHz, correspondingly, the cutoff frequency of the simulated RIRs was $f_c=4$ kHz. According to the Nyquist-Shannon sampling theorem (5), spatial intervals $\Delta\le 0.04$ m are required for the virtual grid. 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 virtual grid $\mathcal{G}$, as in Eq. (7), was set to $\boldsymbol{r}_0=[2.75, 1.4, 0.8]^T$. Binary MLSs with power $\sigma_s^2=1$ and period lengths of $L_p\in\{511, 1023\}$, depending on the lengths of the sought RIRs, were designed, and the positive and negative coefficients were scaled to yield zero DC offset and to obtain a perfect autocorrelation function according to Eq. (34). In real world experiments, however, this adjustment is not necessary, since audio recordings are DC free by themselves. The measurements were simulated in a steady-state room, assuming periodic excitation starting at $n=-L_p$. We sampled the PAF on a plane at height 0.8 m by setting $Z=1$ in Eq. (8). The 3D problem can be seen as a stack of multiple plane grids.

Similar to the definition in Ref. 34 but allowing no scaling of the estimate, we used the *normalized system misalignment*

$$\mathrm{NSM}=\frac{\|\boldsymbol{h}_u-\hat{\boldsymbol{h}}_u\|_{\ell_2}^2}{\|\boldsymbol{h}_u\|_{\ell_2}^2} \tag{86}$$

as evaluation criterion for the quality of a recovered RIRs, 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. For a quantitative description of the recovered PAF involving N RIRs, we use the mean NSM,

$$\mathrm{MNSM}_N=\frac{1}{N}\sum_{u=1}^{N}\frac{\|\boldsymbol{h}_u-\hat{\boldsymbol{h}}_u\|_{\ell_2}^2}{\|\boldsymbol{h}_u\|_{\ell_2}^2}. \tag{87}$$

A. Experiment 1

The first experiment is devoted to a proof of concept of the general validity of the new approach. For this, we compare the best possible dynamic sampling with the conventional static technique. This is achieved by constraining the sampling positions of the dynamic procedure to the points of the spatial grid. In doing so, we avoid all interpolation errors and additionally obtain optimal interpolation coefficients according to Eq. (43).

The spatial grid was modeled with sampling interval $\Delta = 0.02\,\mathrm{m}$ and $X = Y = 5$. The lengths of the RIRs were set to $L = 500$. An MLS with $R = 10$ periods and period length $L_p = 511$ was used.

For the static method, we measured the 25 RIRs on the spatial grid by correlating each measurement signal with one period of the MLS and averaging over all R periods.[6,7] For estimation of the baseline of the recovery quality, different noise conditions were considered. The MNSMs for different signal-to-noise ratios

$$\mathrm{SNR} = \frac{\sigma_s^2}{\sigma_\eta^2} \tag{88}$$

can be seen in the first line of Table I. For the dynamic method, five different setups were considered. The first one used also $Q_{\max} = 25$ microphones. The dynamic sampling was achieved by random rotations of the microphone array around its center. After taking a sample, the array was rotated by a multiple of $\pi/2$. This method fulfills the requirement $\boldsymbol{r}_q(n) \in \mathcal{G}$ for all n and all microphones q. The other setups used a smaller number $Q \in \{20, 15, 10, 5\}$ of microphones which allowed additional translations with a multiple of Δ.

For the dynamic recovery, we tested both, the least-squares solution of the large system (22), and the least-squares solutions of the 511 time-decoupled systems (70). The least-squares solutions were obtained by using the Moore-Penrose pseudoinverses. Both strategies led to the same recovery results, which are also given in Table I. The comparison with the first line shows that the dynamic sampling with 25 microphones is as good as the static one for all noise levels. Slight variations are attributed to the randomness of the measurement noise.

The reduction of the number of microphones leads to a smaller number of samples and therefore a smaller number of equations describing the sought RIRs. The reduced quality of reconstruction follows closely

$$\Delta_{\mathrm{MNSM}} = 10 \cdot \log_{10}\left(\frac{Q_{\max}}{Q}\right) \tag{89}$$

for all noise levels.

TABLE I. MNSM_{25} [dB] of the RIRs on a 5×5 grid, depending on levels of measurement noise. Results for the static method with 25 microphones and the new proposed dynamic method using $Q \in \{25, 20, 15, 10, 5\}$ microphones. Measurements are performed exactly on grid positions.

	SNR [dB]						
Sampling	10	20	30	40	50	60	70
Static	0.43	−9.55	−19.53	−29.57	−39.45	−49.45	−59.51
Dyn-25	0.49	−9.60	−19.61	−29.49	−39.47	−49.58	−59.52
Dyn-20	1.71	−8.28	−18.28	−28.37	−38.29	−48.40	−58.26
Dyn-15	3.10	−6.94	−16.98	−26.98	−37.00	−46.97	−56.97
Dyn-10	4.72	−5.31	−15.26	−25.31	−35.27	−45.33	−55.27
Dyn-5	7.51	−2.49	−12.59	−22.47	−32.47	−42.50	−52.58

B. Experiment 2

In the second experiment, we investigated the effect of spatial oversampling on the recovery results for both the linear and the Lagrange interpolation. The linear interpolation formula provides the weighting coefficients (24) for the grid RIRs at very low computational cost, considering a neighborhood of only two in each dimension. For the uniform grid in space, the Lagrange interpolator is equivalent to a maximally flat fractional delay filter with finite impulse response, as both obtain the same coefficients.[35] The maximally flat filter approximates the ideal interpolation at low frequencies. In order to reduce oscillations as effect of Runge's phenomenon, the degree of the Lagrange polynomials was limited by restricting the maximum distance between a measuring position and the corresponding grid points for interpolation to 5Δ.

We tested the dynamic recovery technique on a plane of size $0.12\,\mathrm{m} \times 0.12\,\mathrm{m}$ considering different sampling intervals $\Delta \in \{0.01\,\mathrm{m}, 0.015\,\mathrm{m}, 0.02\,\mathrm{m}, 0.03\,\mathrm{m}\}$ for the virtual grid in space. Accordingly, the number of grid positions to be recovered varied with $N \in \{169, 81, 49, 25\}$. The lengths of the RIRs were limited to $L = 1000$ taps. An MLS with $L_p = 1023$ was used for excitation. Solving the system (22) for recovery would involve 1.69×10^5 unknowns for the finest grid. Hence, we decoupled the time dimension and solved the linear systems (70) for least-squares, each comprising only 169 unknowns for the finest grid.

The dynamic measurements were taken over $R = 1000$ periods by only one moving microphone. For the microphone trajectory we used the Lissajous figure

$$\boldsymbol{r}(n) = \boldsymbol{r}_0 + \boldsymbol{l} + \left[A \sin\left(\frac{2\pi n f_x}{R L_p}\right), A \cos\left(\frac{2\pi n f_y}{R L_p}\right), 0\right]^T \tag{90}$$

with $\boldsymbol{l} = [0.06, 0.06, 0]^T$, $A = 0.06$, and frequency ratio $f_x/f_y = 32/33$. Lissajous trajectories for measurement are typically used in magnetic particle imaging.[36] They provide a high density of sampling points in curves at the edges of the sampling area. This is promising to improve the sound-field recovery at outer grid positions, which suffer the most from the interpolation on the finite grid.

For evaluation we use RIRs at common points in space which appear on all grids for any Δ considered. The results are presented in Fig. 4 for SNRs of the measurements by 60 and 40 dB, respectively. On the one hand, Fig. 4(a) shows the MNSM_4 values for the RIRs on the four grid positions at the corners of the spatial grids. Here, the Lissajous trajectory achieved a high density of sampling points. On the other hand, Fig. 4(b) gives the quality of recovery for the central position of each spatial grid, where a significantly lower number of samples was provided. In the middle area, the density of sampling points is so low that virtual grids with $\Delta < 0.01\,\mathrm{m}$ would lead to underdetermined RIRs.

Although a finer spatial grid leads to more unknowns in the linear systems of equations, the quality of sound-field recovery improves with spatial oversampling up to factor

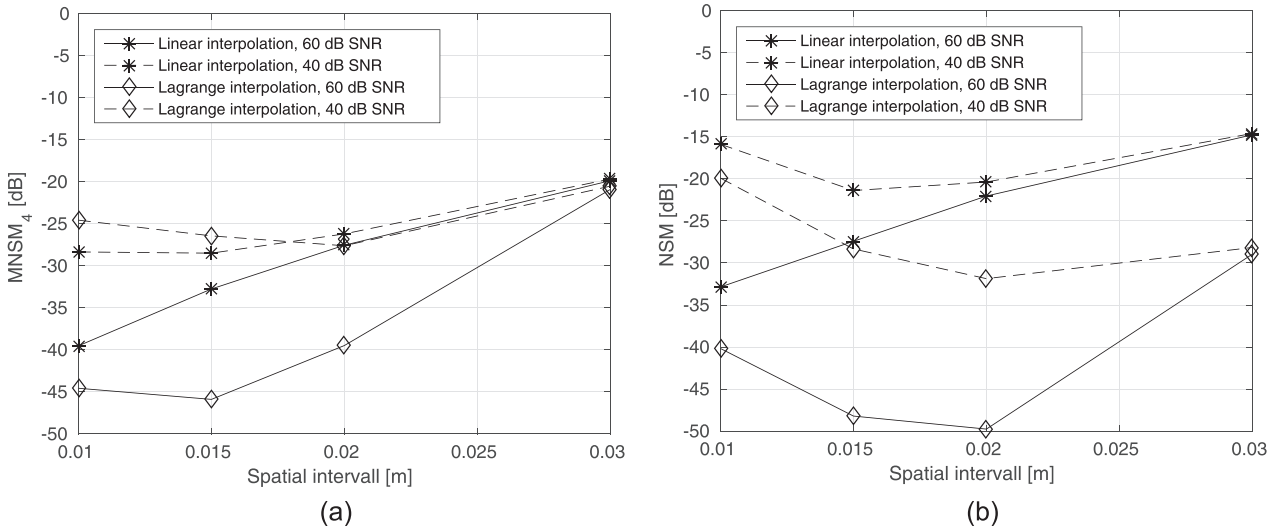

FIG. 4. Dynamic recovery on different virtual grids inside a constant plane using a Lissajous trajectory. Recovery quality of (a) the RIRs at the corners on the virtual grids and (b) the RIR at the center position on the virtual grids.

two ($\Delta = 0.02\,\mathrm{m}$) for both interpolation methods and noise levels in our scenario. As can be seen in Fig. 4, additional oversampling might lead to a diverging quality of the recovered RIRs, depending on the interpolation method, the noise level, and the number of samples for the grid positions concerned. Especially the Lagrange interpolation raises performance on adequately fine grids. However, for too fine grids the performance of the interpolation does not improve significantly, but the ratio between number of unknowns and number of samples still grows, which impairs the recovery using noisy measurements.

C. Experiment 3

In the third experiment, we tested the multigrid recovery scheme presented in Sec. V B and compared it to the direct recovery on one single grid without band-pass filtering of the measurement signal. The lengths of the RIRs were set to $L = 1000$ and an MLS with $L_p = 1023$ and $R = 1000$ was used for excitation.

The volume of interest was a plane of size $0.3\,\mathrm{m} \times 0.3\,\mathrm{m}$. For the dynamic measurement, a microphone array with $Q = 4$ microphones arranged on a quadratic grid with spacing 0.15 m was applied. Referring to the central position $\bar{\boldsymbol{r}}(n)$ of the microphone array, the Lissajous figure according to Eq. (90) with $\boldsymbol{l} = [0.15, 0.15, 0]^T$, $A = 0.075$, and frequency ratio $f_x/f_y = 16/17$ was chosen as trajectory, involving no rotations of the array.

For the multiresolution approach we considered six grid stages with spatial intervals ranging from $\Delta^{(1)} = 0.3\,\mathrm{m}$ for the coarsest grid with $N^{(1)} = 4$ to $\Delta^{(6)} = 0.0094\,\mathrm{m}$ for the finest grid with $N^{(6)} = 1089$. Hamming windowed band-pass filters of order 1000 were used for the filtering of $x(n)$. The cutoff frequencies of the subbands were chosen to be lower than required by Eq. (81), in order to achieve spatial oversampling by factor 4.5. Correspondingly, the limiting frequencies were about $f_c^{(1)} = 125\,\mathrm{Hz}$, $f_c^{(2)} = 250\,\mathrm{Hz}$, $f_c^{(3)} = 500\,\mathrm{Hz}$, $f_c^{(4)} = 1000\,\mathrm{Hz}$, $f_c^{(5)} = 2000\,\mathrm{Hz}$, and $f_c^{(6)} = 4000\,\mathrm{Hz}$. For synthesis of the subbands we tested the linear interpolation and the Lagrange interpolation with maximum sample distance $5\Delta^{(v)}$, i.e., the same methods as used for interpolating the dynamic measurements.

The quality of the recovered PAF is presented in Fig. 5 for various levels on measurement noise. By contrast with the direct recovery on the finest grid with $N = 1089$, the multigrid approach improves the performance by about 3 dB for $\mathrm{SNR} \leq 50\,\mathrm{dB}$.

For very low noise levels, the error of the non-perfect subband decomposition and synthesis is relatively higher than the actual benefit of the multiresolution approach. This is shown by frequency analysis of the error. Therefore, we define the mean energy spectral density of the error, denoted with ME-ESD and calculated as

$$\mathrm{ME\text{–}ESD}(f) = \frac{1}{N}\sum_{u=1}^{N} |H_u(f) - \hat{H}_u(f)|^2, \tag{91}$$

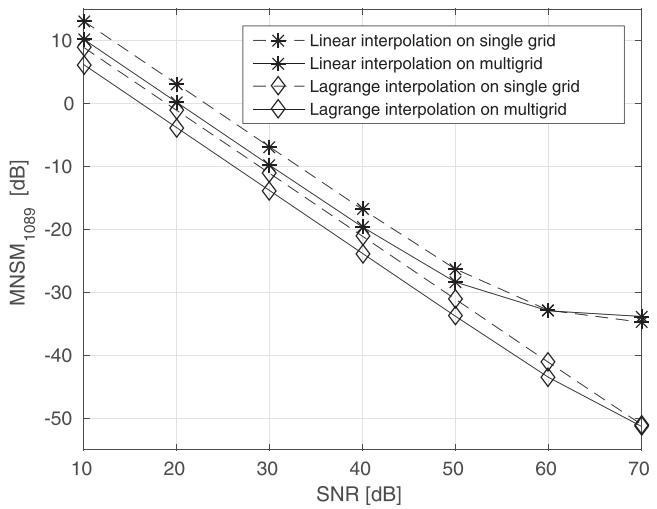

FIG. 5. Comparison of recovery quality with and without multigrid approach for different interpolation methods and levels of measurement noise.

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

In Fig. 6, the ME-ESD values are presented for the recovery with and without the multigrid technique, regarding three noise levels and both interpolation methods. The left-hand sided plots in Fig. 6 show the resulting recovery errors by use of the linear interpolation, at the right-hand side are the results for the Lagrange interpolation. Figures 6(a) and 6(b) present the frequency-dependent error for measurement noise with SNR = 40 dB, Figs. 6(c) and 6(d) for SNR = 50 dB, and Figs. 6(e) and 6(f) for SNR = 60 dB.

For SNR $\leq$ 50 dB, the multigrid recovery scheme obtains much better results in lower frequency bands up to 2000 Hz than the direct recovery on one single grid that involves the whole set of unknowns for all frequencies. In the highest band from 2000 to 4000 Hz, the number of unknowns in $\boldsymbol{h}^{(v)}$ equals the number of unknowns in $\boldsymbol{h}$, thus both strategies obtain nearly the same errors.

We observe two types of frequency errors induced by the multiresolution technique: errors due to the spatial synthesis step and errors caused by the non-perfect subband decomposition of the chosen filters. However, as mentioned above, both of them are clearly below the error level of the direct recovery on one single grid for SNR $\leq$ 50 dB.

Using the linear interpolation for spatial synthesis of the subband RIRs effects that each subband recovery suffers from the typical decrease of performance in high frequencies. Due to the close connection between the temporal and the spatial frequencies, the accuracy of the linear interpolation in space directly affects the recovery quality of the corresponding temporal frequencies. This can be explicitly seen in Fig. 6(e).

In Figs. 6(b), 6(d), and 6(f), the spectral flatness property of the Lagrange interpolation can be observed. Here, the error for the recovery on the single grid is constant over all frequencies, and the error for each multigrid stage is constant inside the considered frequency band beyond the transition frequencies. Instead, the synthesis step via Lagrange interpolation obtains two significant peaks of transition errors, one at the lowest transition frequency $f_c^{(1)} = 125$ Hz with ME-ESD $= 6.8 \times 10^{-10}$, and one at the highest transition frequency $f_c^{(6)} = 2000$ Hz with ME-ESD $= 4.3 \times 10^{-10}$. These error peaks caused by subband filtering exceed the actual

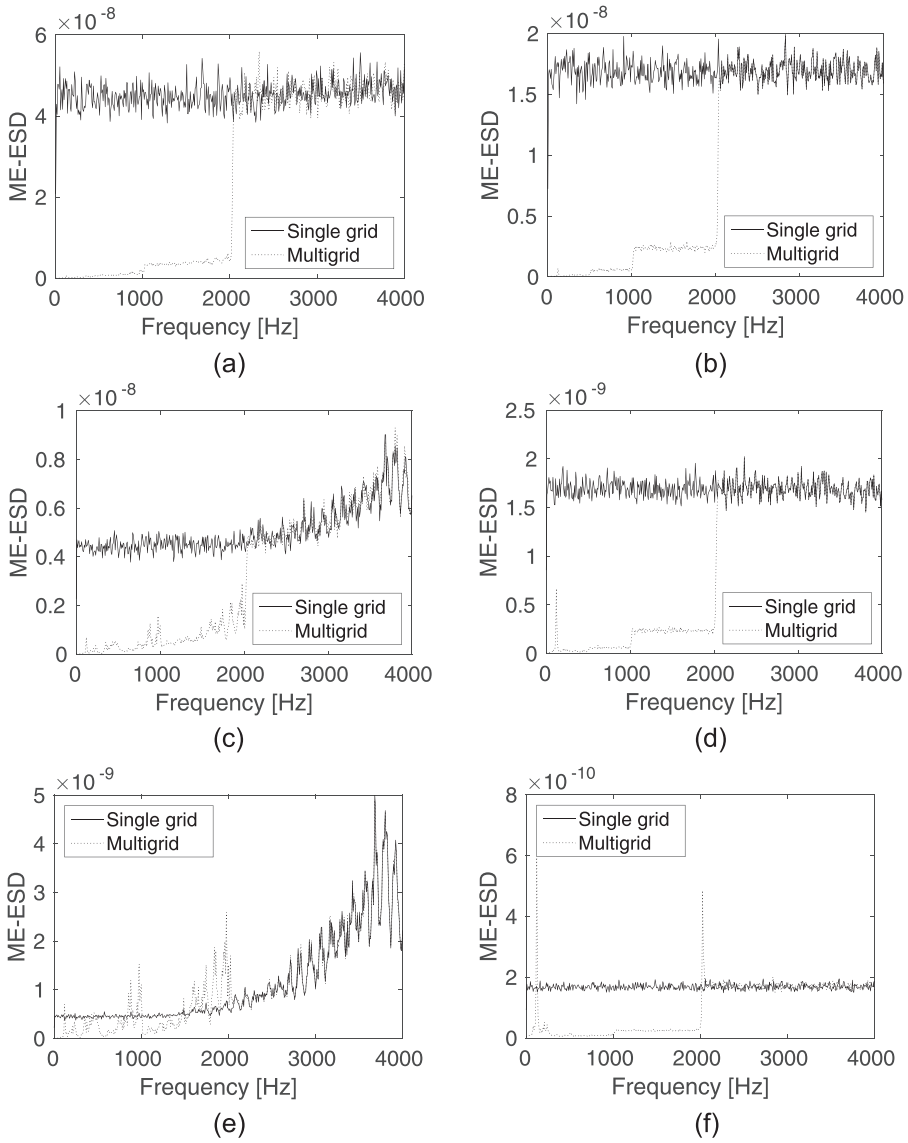

FIG. 6. Frequency-dependent error with and without using the multigrid recovery scheme for different noise levels. (a), (b) SNR = 40 dB, (c), (d) SNR = 50 dB, and (e), (f) SNR = 60 dB. On the left-hand side, the linear interpolation was used, the right-hand side depicts the results of using the Lagrange interpolation.

error level of the direct recovery on the single grid for SNR = 60 dB and SNR = 70 dB. However, the error is so narrow banded that the overall quality of recovery in the sense of the NSM is not impaired when using the multigrid scheme with Lagrange interpolation in the synthesis step.

VII. CONCLUSION

In this work, we proposed a new method for the measurement of sound fields using moving microphones. With known trajectory and excitation signal, a system of linear equations can be set up that leads to the estimation of spatially dependent RIRs on an equidistant virtual grid. A potential application of the proposed method is the high-precision measurement of sound fields using hand-held microphones whose positions are continuously tracked. In order to reduce the computational complexity, perfect sequences have been used. Following the dispersion relation of propagative sound waves, a multiresolution recovery scheme has been suggested for the spatial grid, allowing for both the enhancement of recovery quality under noisy conditions and the fast reconstruction of low frequencies at minor computational cost. We showed that, for the best-case trajectory, our dynamic approach is as good as conventional sampling using static microphones. For non-optimal trajectories, we demonstrated that the use of an oversampled virtual grid substantially improves the recovery results. For trajectories that lead to ill-conditioned linear systems, the extension of the dynamic method to the theory of compressed sensing has been outlined. Relating to this, further approaches that exploit sparse representations of sound fields and use compressed-sensing reconstruction are currently under investigation.

ACKNOWLEDGMENTS

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

[1]J. N. Mourjopoulos, "Digital equalization of room acoustics," J. Audio Eng. Soc. **42**(11), 884–900 (1994).
[2]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. **22**(10), 1522–1532 (2014).
[3]M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (2012), pp. 13–16.
[4]V. P. Ipatov, "Ternary sequences with ideal periodic autocorrelation properties," Radio Eng. Electron. Phys. **24**, 75–99 (1979).
[5]H.-D. Lüke, "Sequences and arrays with perfect periodic correlation," IEEE Trans. Aerosp. Electron. Syst. **24**(3), 287–294 (1988).
[6]J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," J. Audio Eng. Soc. **31**(7/8), 478–488 (1983).
[7]D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**(6), 419–444 (1989).
[8]A. Farina, "Advancements in impulse response measurements by sine sweeps," in *122nd AES Convention* (2007), pp. 1–21.
[9]A. Telle, C. Antweiler, and P. Vary, "Der perfekte Sweep—Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme" ("A perfect sweep–A new excitation signal for the adaptive system identification of time-variant acoustic systems"), in *Proceedings of the German Annual Conference on Acoustics*, Berlin (2010), pp. 341–342.
[10]M. Sondhi, "An adaptive echo canceller," Bell Syst. Tech. J. **46**(3), 497–511 (1967).
[11]J. Benesty, Y. Huang, J. Chen, and P. Naylor, "Adaptive algorithms for the identification of sparse impulse responses," in *Topics in Acoustic Echo and Noise Control*, edited by E. Hänsler and G. Schmidt (Springer, Berlin, 2006), pp. 125–153.
[12]D. L. Duttweiler, "Proportionate normalized least-mean-squares adaption in echo cancelers," IEEE Trans. Speech Audio Process. **8**(5), 508–517 (2000).
[13]Y. Kaneda, S. Makino, and N. Koizumi, "Exponentially weighted step-size NLMS adaptive filter based on the statistics of a room impulse response," IEEE Trans. Speech Audio Process. **1**(1), 101–108 (1993).
[14]P. A. Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation," J. Sign. Process. **86**(6), 1182–1192 (2006).
[15]K. Helwani, H. Buchner, and S. Spors, "Multichannel adaptive filtering with sparseness constraints," in *International Workshop on Acoustic Signal Enhancement* (2012), pp. 1–4.
[16]J. Benesty, D. R. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," IEEE Trans. Speech Audio Process. **6**(2), 156–165 (1998).
[17]T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (2003), pp. 616–619.
[18]T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," IEEE Trans. Sign. Process. **54**(10), 3790–3804 (2006).
[19]T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," J. Acoust. Soc. Am. **122**(3), 1636–1645 (2007).
[20]T. Ajdler, "The plenacoustic function and its applications," Ph.D. thesis, École Polytechnique Fédérale de Lausanne (2006).
[21]K. H. Kuttruff, "Sound in enclosures," in *Handbook of Acoustics*, edited by M. J. Crocker (Wiley-Interscience, New York, 1998), pp. 935–938.
[22]J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950 (1979).
[23]P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**(5), 1527–1529 (1986).
[24]R. Mignot, L. Daudet, and F. Ollivier, "Interpolation of room impulse responses in 3d using compressed sensing," in *Proceedings of Acoustics*, Société Francaise d'Acoustique (2012), pp. 2943–2948.
[25]R. Gribonval, G. Chardon, and L. Daudet, "Blind calibration for compressed sensing by convex optimization," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (2012), pp. 2713–2716.
[26]S. M. Kay, "Linear Bayesian estimator," in *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, Upper Saddle River, NJ, 1993), pp. 379–418.
[27]D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
[28]E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006).
[29]B. K. Nataratjan, "Sparse approximate solutions to linear systems," SIAM J. Comput. **24**(2), 227–234 (1995).
[30]D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," IEEE Trans. Inf. Theory **47**(7), 2845–2862 (2001).
[31]E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," Ann. Stat. **35**(6), 2313–2351 (2007).
[32]R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Stat. Soc. **58**, 267–288 (1996).
[33]S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Sci. Comput. **20**(1), 33–61 (1998).
[34]D. R. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," IEEE Sign. Process. Lett. **5**(7), 174–176 (1998).
[35]M. M. J. Yekta, "A frequency domain proof for the equivalence of the maximally flat FIR fractional delay filter and the Lagrange interpolator," Digital Sign. Process. **21**(1), 13–19 (2011).
[36]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. **54**(2), 385–392 (2009).