SPHERICAL HARMONIC REPRESENTATION FOR DYNAMIC SOUND-FIELD MEASUREMENTS

Fabrice Katzberg, Marco Maass, and Alfred Mertins

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

ABSTRACT

Continuously moving microphones produce a high number of spatially dense sound-field samples with low effort in hardware and acquisition time. By interpreting the dynamic procedure as the non-uniform sampling of spatial basis functions, a system of linear equations can be set up. Its solution encodes sound-field parameters that allow for the spatio-temporal reconstruction within the measurement area at bandwidths where static methods would require impractical setups. An existing framework considers such basis functions from a signal processing point of view. It uses sinc-function based interpolation filters which are highly localized around sampled trajectories and may lead to ill-posed problems unless sparsity constraints are made, especially for locations that are away from microphone trajectories. In this paper, we present a new physical interpretation of the dynamic sampling problem. Transferring the problem into frequency domain, we describe samples of a moving microphone in terms of sampled spherical harmonic functions. The use of these global basis functions leads to dynamic measurements that inherently encode expanded sound-field information and, thus, allow for robust reconstruction at off-trajectory positions.

Index Terms— Room impulse responses, moving microphones, spherical harmonics, wave equation.

1. INTRODUCTION

Sound transmission between two points in reverberant environments is described by a convolution with the room impulse response (RIR). The RIR is highly sensitive to spatial shifts due to various sound paths reflecting from room walls and superposing at the receiver position. Thus, a massive number of dense measuring positions is needed for accurately reconstructing the spatio-temporal RIR over an extended volume at large bandwidths. However, this is valuable in many applications related to sound field analysis, auditory scene synthesis, and channel equalization. Typically, RIRs for static emitter-receiver pairs are measured by using deterministic excitation signals that allow for inverting their convolution by correlating the microphone signal with their reverse signal. Signals applicable for such simple deconvolution are perfect sequences [1], maximum-length sequences (MLSs) [2], and sine sweeps [3].

Spatio-temporal recovery of high frequencies makes stationary sampling impractical due to too small spatial intervals required for aliasing-free reconstruction [4]. This includes high effort in calibration and exact positioning of individual microphones. For obtaining qualified sound-field estimates from sub-Nyquist sampling at reduced effort, several static approaches exploit sparse signal structures [5–7] according to the compressed-sensing paradigm [8].

Beyond deterministic excitation, any arbitrary source signal can be used for encoding the convolution into a system of linear equations and estimating involved time-invariant RIRs by least-squares optimization. Such linear measurement equations are also the basis for dynamic approaches where impulse responses at continuously changing positions are modeled in terms of time-varying systems, whose coefficients are tracked by adaptive filtering concepts [9]. This type of methods is commonly used for acoustic echo cancellation [10–13] and for fast acquisition of head-related impulse responses (HRIRs) [14–17]. In [18], a different, analytical approach has been introduced. Using a specially designed excitation signal, the Fourier slice theorem is exploited and the Doppler effect is directly taken into account, in order to reconstruct RIRs along a linear or circular trajectory performed by a microphone at constant speed.

More recently, several methods have been proposed that interpret the reconstruction from dynamic sound-field sampling as spatial interpolation task [19–24]. They explicitly use a multidimensional measurement model where a moving microphone simultaneously generates uniform samples in time domain and, in general, non-uniform samples at varying points in the spatial domain. Such techniques require positional information, based either on a controlled pre-defined trajectory or a tracking of the microphone positions. The method in [19] uses a spatial Fourier basis for the angular reconstruction of HRIRs from continuous-azimuth recordings, and shows more accurate performance than an adaptive-filter based solution. In [20], perfect-sequence excitation is used for the orthogonal expansion of impulse responses, in order to describe the dynamic spatio-temporal sampling by notional static sampling processes of single expansion coefficients. This method, simplifying the problem to pure interpolation in space, has been investigated for reconstructing binaural impulse responses [21] and RIRs along a circle [22]. In [23, 24], a versatile framework has been introduced, where the deconvolution problem in time domain and the interpolation problem in space domain are both integrated into one structured system of linear equations whose solution allows for RIR reconstruction in three-dimensional space. To achieve this, the sound field is parameterized by modeling notional grid points in space and dynamic samples are understood as the result of bandlimited interpolation on that grid using sampled sinc-function approximations.

In this paper, we first present a more general formulation of the dynamic model proposed in [23, 24]. The linear equations are construed in terms of sampling of weighted spatial basis functions, where the weights encode the spatio-temporal RIR and are referred to as sound-field parameters. Then, this generalized model is applied to a new physically based perspective of the dynamic sampling prob-

This work has been supported by the German Research Foundation under Grant No. ME 1170/10-2.

lem, leading to a spherical harmonic representation with frequency parameters describing spherical solutions to the acoustic wave equation. This representation, similarly known from sound-field reproduction techniques [25, 26], is beneficial due to its natural interpretation of sound propagation and is experimentally shown to achieve robust RIR reconstruction within the measurement volume, also at positions being far away from the microphone trajectory.

2. GENERAL DYNAMIC SAMPLING MODEL

By assuming a closed-room scenario with constant atmospheric conditions, the transmission of the sound signal $s(t)$ from a fixed source position to location $\boldsymbol{r} \in \mathbb{R}^3$ subject to time $t \in \mathbb{R}$ is modeled with $p(\boldsymbol{r}, t) = \int_{-\infty}^{\infty} s(t-\tau) h(\boldsymbol{r}, \tau) d\tau$, where $p(\boldsymbol{r}, t)$ is the resulting sound-pressure field and $h(\boldsymbol{r}, t)$ is the spatio-temporal RIR describing a linear time-invariant (LTI) system, i.e., the particular sound-pressure field for a Dirac impulse at $t = 0$.

First, let us consider the temporally sampled version of $p(\boldsymbol{r}, t)$. For a signal having maximum frequency f_{max}, the sampling with $f_\text{s} > 2f_{\text{max}}$ yields uniform samples at equidistant points $t_n = n/f_\text{s}$ ($n \in \mathbb{Z}$) that allow for aliasing-free reconstruction in time. The measurement process subject to the discrete variable n reads $p(\boldsymbol{r}, n) = s(n) * h(\boldsymbol{r}, n) + \eta(n)$, with $\eta(n)$ being the measurement noise.

Instead of capturing the entire field information over all $\boldsymbol{r}$ at each sampling point n, let us now assume that the acquired data is restricted to Q positions $\boldsymbol{r}_q(n) \in \mathbb{R}^3$ changing over time. By defining Q as the number of microphones moving along the trajectories $\boldsymbol{r}_q(n)$ inside a volume of interest Ω, we obtain a dynamic measurement setup that generates samples at uniform points in time and, generally, at varying and non-uniform positions in space.

Without loss of generality, we set $Q = 1$. The sampling process of a single microphone moving along the trajectory $\boldsymbol{r}(n)$ follows the spatio-temporal convolution $p(\boldsymbol{r}(n), n) = s(n) * h(\boldsymbol{r}(n), n)$. This equation holds true for the dynamic case despite of the Doppler effect since both signals, $p(\boldsymbol{r}(n), n)$ and $h(\boldsymbol{r}(n), n)$, live on the joint manifold $\boldsymbol{r}(n)$ leading to shared Doppler shifts [23]. Accordingly, supposing that the amplitude of $h(\boldsymbol{r}, t)$ vanishes into the noise level for $t_n > t_{L-1}$, the dynamic spatio-temporal measurement of N samples reads

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

where $h(\boldsymbol{r}(n), m)$ covers RIRs, each with L uniform delays m, along the non-uniform trajectory $\boldsymbol{r}(n)$. By modeling an appropriate parameterization of the continuous-space and discrete-time RIR within Ω using multidimensional basis functions f_p,

$$h(\boldsymbol{r}, n) \approx \sum_{p=1}^{P} a_p f_p(\boldsymbol{r}, n) \quad \forall\, \boldsymbol{r} \in \Omega, \tag{2}$$

the RIRs along the trajectory $\boldsymbol{r}(n) \in \Omega$ may be represented by the spatially sampled basis functions f_p with respect to the same parameters a_p. Thus, the resulting sampling model is

$$p(\boldsymbol{r}(n), n) = \sum_{m=0}^{L-1} s(n-m) \sum_{p=1}^{P} a_p f_p(\boldsymbol{r}(n), m) + \eta(n). \tag{3}$$

This allows for setting up a system of linear equations describing the dynamic sampling subject to joint parameters. The knowledge of $s(n)$ and $\boldsymbol{r}(n)$ enables the calculation of estimates $\hat{a}_p$ that solve the linear system, e. g. in the least-squares sense with a number of samples $N \geq P$, and may be used for spatial reconstruction inside Ω according to (2).

3. PHYSICAL PARAMETERIZATION MODEL

The sound pressure $p(\boldsymbol{r}, t)$ in free three-dimensional space follows the homogeneous acoustic wave equation

$$\nabla_{\boldsymbol{r}}^2\, p(\boldsymbol{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\, p(\boldsymbol{r}, t) = 0, \tag{4}$$

where $\nabla_{\boldsymbol{r}}^2$ denotes the Laplace operator with respect to the coordinates in $\boldsymbol{r}$, and c is the speed of sound [27, 28]. By defining the position vector in spherical coordinates $\boldsymbol{r} = [r, \theta, \phi]^T$, solutions to the wave equation can be derived by separating the variables according to

$$p(\boldsymbol{r}, t) = R(r)\, \Theta(\theta)\, \Phi(\phi)\, T(t), \tag{5}$$

which leads to four ordinary differential equations with dependence of the pressure on radius r, polar angle $\theta \in [0, \pi]$, azimuth angle $\phi \in [0, 2\pi)$, and time t, respectively [27,29]. By choosing the fundamental temporal solution $T(t) = e^{j\omega t}$, with the imaginary unit j and the angular frequency $\omega = 2\pi f$, temporal progress of monochromatic sound-pressure fields follows $p(\boldsymbol{r}, t) = p(\boldsymbol{r}) e^{j\omega t}$. Applying this spatio-temporal separation to (4) leads to the Helmholtz equation $\nabla_{\boldsymbol{r}}^2\, p(\boldsymbol{r}, k) + k^2 p(\boldsymbol{r}, k) = 0$, which constitutes a time-independent description of the wave equation respecting the angular wave number $k = \omega/c$.

Solutions to the angular dependencies are conveniently pooled in the set of orthonormal and complete spherical harmonic functions, $Y_v^u(\theta, \phi) = \Theta(\theta)\, \Phi(\phi)$ with order $v \in \mathbb{N}_0$ and mode $u \in \mathbb{Z}$, defined as

$$Y_v^u(\theta, \phi) = \sqrt{\frac{(2v+1)}{4\pi} \frac{(v-u)!}{(v+u)!}}\, P_v^u(\cos\theta)\, e^{ju\phi}, \tag{6}$$

where P_v^u is the associated Legendre function of the first kind with $P_v^u(\cos\theta) = 0$ for all $|u| > v$ [27, 29]. The remaining radial term in the Helmholtz equation leads to the so-called spherical Bessel differential equation. For spherical volumes being free of sound sources, the amount of incoming and outgoing waves is equal and, thus, complete radial solutions are given by $R(r) = A_v b_v(kr)$, where $A_v \in \mathbb{C}$ is a constant and $b_v(kr)$ is the spherical Bessel function of the first kind [27]. By combining all particular solutions, the single-frequency sound pressure in free field reads

$$p(r, \theta, \phi) = \sum_{v=0}^{\infty} \sum_{u=-v}^{v} A_{vu} b_v(kr)\, Y_v^u(\theta, \phi), \tag{7}$$

which may be regarded as the inverse transform of the sound field using basis functions $Y_v^u(\theta, \phi)$. For known pressure $p(r_0, \theta, \phi)$ on a sphere of radius r_0, the forward transform represented by A_{vu} can be uniquely calculated. Due to the orthonormality of (6), multiplying both sides of (7) by $[Y_v^u(\theta, \phi)]^*$ determines the coefficients as

$$A_{vu} = \frac{1}{b_v(kr_0)} \int_0^{2\pi} \int_0^{\pi} p(r_0, \theta, \phi)\, [Y_v^u(\theta, \phi)]^* \sin\theta\, d\theta d\phi. \tag{8}$$

Substituting (8) into (7) yields

$$p(r, \theta, \phi) = \sum_{v=0}^{\infty} \sum_{u=-v}^{v} a_{vu}(r_0)\, \frac{b_v(kr)}{b_v(kr_0)}\, Y_v^u(\theta, \phi), \tag{9}$$

where $a_{vu}(r_0)$ is referred to as the spherical wave (SW) spectrum at radius r_0 [27], which allows for wave-field extrapolation to arbitrary radii according to $a_{vu}(r) = a_{vu}(r_0) b_v(kr)/b_v(kr_0)$.

4. PROPOSED APPROACH

The spherical volume of interest Ω is free of sound sources and obstacles. As usual, the impact of the microphone on the sound field is considered negligible. The origin of the spherical coordinate system is set to the center of Ω. Feasible positions are restricted by $r \in [0, r_{\max}]$, with $r_{\max}$ being the maximum radius inside Ω. The acoustical environment is supposed to be reverberant, which is consistent with the free-field model within Ω. This can be shown by reference to the LTI assumption and geometrical room acoustics [30].

4.1. RIR Parameters in Frequency Domain

The sound field for Dirac delta excitation at $t = 0$ is described by particular solutions to the acoustic wave equation. By analogy with (2), (5), and (9), we propose a frequency-wise parameterization of the temporally sampled RIR inside continuous space Ω according to

$$h(\boldsymbol{r}, n) = \frac{1}{L} \sum_{l=0}^{L-1} h_{\omega_l}(\boldsymbol{r}, n) = \frac{1}{L} \sum_{l=0}^{L-1} h_{\omega_l}(\boldsymbol{r})\, e^{j\omega_l t_n}, \qquad (10)$$

with SW parameters $a_{vu}(r_{\max}, l)$ composing

$$h_{\omega_l}(\boldsymbol{r}) = \sum_{v=0}^{M_l} \sum_{u=-v}^{v} a_{vu}(r_{\max}, l)\, \frac{b_v(k_l r)}{b_v(k_l r_{\max})}\, Y_v^u(\theta, \phi) \qquad (11)$$

for sampled frequencies $\omega_l = 2\pi f_s l / L$ and $k_l = \omega_l / c$ with $l \in \{0, \ldots, L-1\}$. The equality in (11) is valid for $M_l = \infty$, however, the magnitudes of the Bessel functions decay rapidly for $v > kr$ and accurate approximations are obtained by limiting the order of the spherical harmonics to $M_l \geq k_l r_{\max}$ [29]. Basically, (11) represents the discrete Fourier transform (in time dimension) of the spatially expanded RIR, $H(\boldsymbol{r}, l) = \sum_{n=0}^{L-1} h(\boldsymbol{r}, n)\, e^{-2\pi j n l / L} = h_{\omega_l}(\boldsymbol{r})$, and (10) is the corresponding inverse discrete Fourier transform.

4.2. Physical Formulation of the Dynamic Sampling Problem

By substituting (10) into the dynamic model (1) and considering the error-free case, the signal recorded by the microphone moving along the trajectory $\boldsymbol{r}(n) = [r_n, \theta_n, \phi_n]^T \in \Omega$ is

$$\begin{aligned} p(\boldsymbol{r}(n), n) &= \sum_{m=0}^{L-1} s(n-m)\, \frac{1}{L} \sum_{l=0}^{L-1} h_{\omega_l}(\boldsymbol{r}(n))\, e^{2\pi j \frac{l}{L} m} \\ &= \sum_{l=0}^{L-1} S_l(n) \sum_{v=0}^{M_l} \sum_{u=-v}^{v} a_{vu}(r_{\max}, l) \\ &\qquad \cdot \frac{b_v(k_l r_n)}{b_v(k_l r_{\max})}\, Y_v^u(\theta_n, \phi_n), \end{aligned} \qquad (12)$$

with $S_l(n) = 1/L \sum_{m=0}^{L-1} s(n-m)\, e^{2\pi j m l / L}$ corresponding to the (inverse) discrete short-time Fourier transform of $s(n)$ with window length L and hop size of one. The resulting linear model reads

$$\boldsymbol{p} = \sum_{l=0}^{L-1} \boldsymbol{\mathcal{S}}_l \boldsymbol{Y}_{M_l} \odot \boldsymbol{B}_l \boldsymbol{D}_l\, \boldsymbol{a}_l + \boldsymbol{\eta}, \qquad (13)$$

where $\odot$ denotes the Hadamard product, $\boldsymbol{p} \in \mathbb{R}^N$ is the measurement vector, $\boldsymbol{\eta} \in \mathbb{R}^N$ is the error term due to noise and order-limiting, $\boldsymbol{\mathcal{S}}_l = \operatorname{diag}\{S_l(0), \ldots, S_l(N-1)\}$ renders the temporal evolution of discrete frequencies in $s(n)$, diagonal matrices $\boldsymbol{D}_l$ with entries $1/b_v(k_l r_{\max})$ connect to the reference sphere of radius $r_{\max}$, the matrices $\boldsymbol{Y}_{M_l} \in \mathbb{C}^{N \times P_l}$ and $\boldsymbol{B}_l \in \mathbb{R}^{N \times P_l}$ with $P_l = (M_l + 1)^2$ carry the sampled spherical harmonic and Bessel functions, respectively, and the vectors $\boldsymbol{a}_l \in \mathbb{C}^{P_l}$ contain the corresponding SW parameters for each sampled frequency ω_l, $\boldsymbol{a}_l = [a_{00}(r_{\max}, l), \ldots, a_{M_l M_l}(r_{\max}, l)]^T$. The $N \times P$ matrix

$$\boldsymbol{\Lambda} = \left[\boldsymbol{\mathcal{S}}_0 \boldsymbol{Y}_{M_0} \odot \boldsymbol{B}_0 \boldsymbol{D}_0, \ldots, \boldsymbol{\mathcal{S}}_{L-1} \boldsymbol{Y}_{M_{L-1}} \odot \boldsymbol{B}_{L-1} \boldsymbol{D}_{L-1}\right]$$

and the vector concatenation $\boldsymbol{a} = \left[\boldsymbol{a}_0^T, \boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_{L-1}^T\right]^T \in \mathbb{C}^P$ of $P = \sum_{l=0}^{L-1} P_l$ parameters build up the system of linear equations

$$\boldsymbol{p} = \boldsymbol{\Lambda} \boldsymbol{a} + \boldsymbol{\eta}. \qquad (14)$$

4.3. Interpretation and Comparison

The moving microphone acquires the convolution results of the excitation sequence with the spatio-temporal RIR inside Ω. In general, the radii of the sampling points vary continuously, and, thus, each sample results from an RIR belonging to an individual SW spectrum. However, by knowing the trajectory of the dynamic microphone, the samples of various SW spectra are jointly represented as the extrapolation of one shared SW spectrum on the sphere of radius $r_{\max}$. These aligning RIR parameters may be estimated by solving the linear system (14), and, in turn, allow for RIR reconstruction at arbitrary positions inside Ω according to (10) and (11). The parameters $a_{vu}(r_{\max}, l)$ encode the continuous RIR sound-field for discrete frequencies in time: they represent the SW spectrum (spatial frequencies) on the sphere of radius $r_{\max}$ for sampled temporal frequencies ω_l. The corresponding field information inside Ω is provided by spherical harmonics (angular reconstruction) and Bessel functions (radial extrapolation), and the discrete delay of RIRs is modeled by temporal solutions $e^{2\pi j l n / L}$. For reconstructing sound waves at ω_l inside a spherical volume Ω_{S} of size V, the number of parameters for the SW model is at least ($M_l = k_l r_{\max}$)

$$P_{\Omega_{\mathrm{S}}}(V, \omega_l) = \left\lceil \omega_l \sqrt[3]{3V\,(4\pi)^{-1}} c^{-1} + 1 \right\rceil^2. \qquad (15)$$

In contrast, the uniform-grid model from [23, 24] uses the parameterization $h(\boldsymbol{x}, n) \approx \sum_{i=1}^{G} h(\boldsymbol{g}_i, n)\, f_i(\boldsymbol{x})$, where the parameters $h(\boldsymbol{g}_i, n)$ are RIRs at notional grid points $\boldsymbol{g}_i$ with spacing $\Delta < c/(2 f_{\max})$, allowing for bandlimited interpolation by separable sinc-function approximations $f_i(\boldsymbol{x}) = f_i(x)\, f_i(y)\, f_i(z)$ subject to Cartesian coordinates $\boldsymbol{x} = [x, y, z]^T$. This model leads to a very simple, structured linear system that enables straightforward error analyses and extensions. It provides full scalability over room dimensions and requires $P_{\Omega_{\mathrm{C}}}(V, \omega_l) = \lceil \omega_l \sqrt[D]{V}\,(2\pi c)^{-1} + 1 \rceil^D$ parameters at a minimum for recovering sampled frequency ω_l inside a D-dimensional cubical region Ω_{C} of size V. The high number of parameters for $D = 3$ and the dynamic-sample representation in terms of spatially concentrated, fast decaying interpolation filters most likely lead to ill-posed or even underdetermined problems in practical applications. The solution to this issues is the compressed-sensing based recovery as proposed in [24]. This involves addition sparsity constraints and sampling conditions.

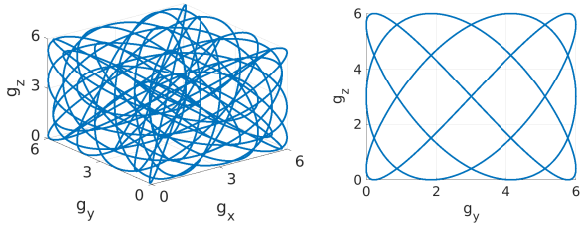
In the experimental part of this paper, we compare non-sparse and sparse uniform-grid recovery from dynamic samples with the proposed approach. The spherical harmonic functions span the entire sphere around the measurement volume. So, each spatio-temporal sample contributes a global amount of wanted sound-field parameters allowing for natural rendering of sound propagation inside Ω, and, thus, for spatially robust sound-field reconstruction. For this, sparse recovery is not mandatory. Results in Sec. 5 are obtained by least-squares solutions of (14), with orders $M_l = \lceil k_l r_{\max} \rceil$.

5. EXPERIMENTS AND RESULTS

By use of the image source method [31], we simulated dynamic measurements in fifty reverberant environments, randomly chosen according to the uniform distribution of room sizes $[2; 10]\,\mathrm{m} \times [2; 10]\,\mathrm{m} \times [2; 5]\,\mathrm{m}$, reverberation times $T_{60} \in [0.05; 0.25]\,\mathrm{s}$, and sampling frequencies $f_\mathrm{s} \in [2000; 8000]\,\mathrm{Hz}$ (with corresponding cutoffs). In each scenario, both the source position and the location of the cubical measurement volume Ω with side lengths $4c/f_\mathrm{s}$ were randomly selected on condition of minimum distances of $0.3\,\mathrm{m}$ to each other and to the room walls. For excitation, we used R repetitions of an MLS with power $\sigma_s^2 = 1$ and period length $L_p \in \{511, 1023, 2047\}$ chosen subject to $L_p \geq L$ where $t_{L-1} = T_{60}$. The number of repetitions was varied with $R = 900$ and $R = 450$, in order to test dynamic recovery for long and short acquisition times. Sampling of the moving microphone was simulated along the Lissajous trajectory shown in Fig. 1, scaled to the particular size of the target volume. This kind of trajectory has been proposed for the original approach in [23]. For the recordings, white Gaussian measurement noise was added at signal-to-noise ratios (SNRs) of $10\,\mathrm{dB}$, $20\,\mathrm{dB}$, and $30\,\mathrm{dB}$, respectively. To evaluate spatial RIR recovery, we use the normalized system misalignment (NSM) as in [23, 24], which measures the energy ratio between the error signal and the corresponding true RIR. Accordingly, the mean NSM (MNSM) is used to describe the reconstruction error over Ω.

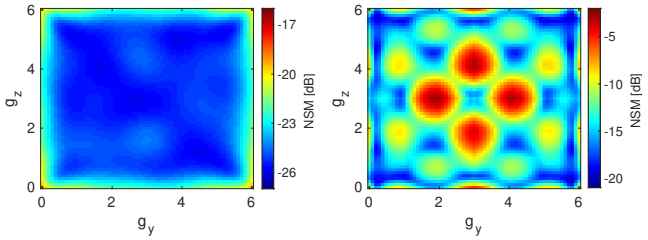
For the uniform-grid parameterization according to [23, 24], we modeled 7^3 grid points in three dimensions covering Ω entirely. Regarding the volume sizes, this corresponds to spatial oversampling by factor 1.5. Lagrange interpolation filters of order five were used in each spatial dimension. Non-sparse recovery of the grid parameters in the least-squares sense leads to MNSM $\gg 1$ in each experiment, and, thus, is unsuitable, although the performed trajectory as in Fig. 1 seems to be spatially dense, interpolation filters have wide support of six grid points in each dimension, and high numbers of $N \gg P$ spatio-temporal samples are provided. Nevertheless, the condition numbers of the involved sampling matrices explode for the uniform-grid model. By doubling the number of dynamic samples ($R = 1800$) along the same trajectory, adequate ranges of condition numbers are obtained. However, in that case good performance in reconstruction is spatially concentrated around the measurement trajectory. At the same time, the physical interpretation of the dynamic sampling model leads to accurate RIR reconstruction within the entire measurement volume, also for less samples and lower SNR. This is shown in Fig. 2 for a particular experiment.

For the compressed-sensing based grid recovery, we used the fast update scheme from [24] based on iterative hard thresholding. Spatial discrete Fourier transforms were used for the sparse representation. The initial sparsity parameter was $K_0 = 4L$, successively relaxed after every five iterations according to $K_{i+1} = 1.5K_i$, and the step size $\mu = 0.15/L$ was chosen. In Fig. 3, reconstruction results in terms of MNSM are shown over all experiments by use of boxplots. For long acquisition times, the physical approach (Fig. 3(a)) and the sparse grid recovery (Fig. 3(b)) achieve comparable accuracy in RIR reconstruction. The proposed method performs slightly better in 30%, 80%, and 64% of the experiments for SNRs of $10\,\mathrm{dB}$, $20\,\mathrm{dB}$, and $30\,\mathrm{dB}$, respectively. However, for short acquisition times, the sparse grid recovery (Fig. 3(d)) outperforms the least-squares solutions of the presented spherical harmonics framework (Fig. 3(c)). This result motivates compressed-sensing based extensions of the physically interpreted dynamic model in future. Existing stationary methods already use similar representations for sparse sound-field analysis, synthesis, and interpolation [7, 32–34].



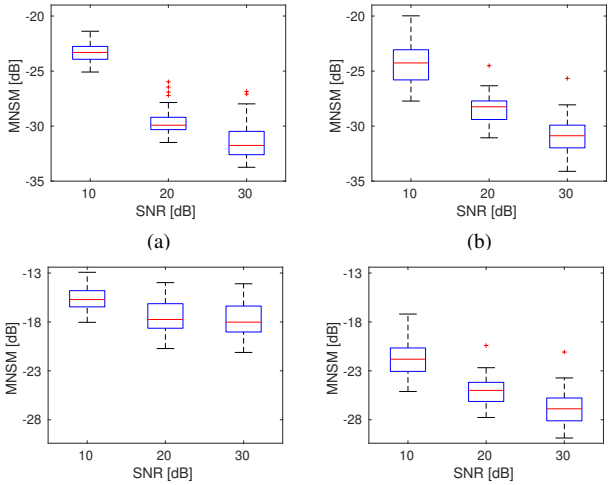
(a) (b)

Fig. 1. Lissajous trajectory used for dynamic sampling. (a) Three-dimensional view and (b) y-z view.



(a) (b)

Fig. 2. Errors of spatially reconstructed RIRs in the centered yz-plane of Ω for dynamic measurements at 7250 Hz inside a room with $T_{60} = 0.2\,\mathrm{s}$ using the trajectory in Fig. 1. Results of (a) the proposed model for $R = 900$, SNR $= 10\,\mathrm{dB}$, and (b) the uniform-grid model for $R = 1800$, SNR $= 30\,\mathrm{dB}$. Note the different colorbar ranges.



(a) (b) (c) (d)

Fig. 3. Errors over all experiments regarding long (first row) and short (second row) acquisition times for (a), (c) the proposed physical model and (b), (d) the sparse recovery on a uniform grid.

6. CONCLUSION

In this paper, we first presented a more general formulation of the dynamic sampling problem in terms of spatially sampled basis functions. Based on that, we introduced a physical model of continuously sampled RIRs. This model allows for robust RIR reconstruction at off-trajectory positions and achieves state-of-the-art results in simulated experiments for sufficiently long acquisition times. Future works will be directed to compressed-sensing based extensions for reducing sampling time.

7. REFERENCES

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

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

[3] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122th Conv. Audio Eng. Soc.*, Conv. Paper 7121, 2007.

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

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

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

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

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

[9] A. Carini, "Efficient NLMS and RLS algorithms for perfect and imperfect periodic sequences," *IEEE Trans. Signal Process.*, vol. 85, no. 4, pp. 2048–2059, 2010.

[10] M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497–511, 1967.

[11] J. Benesty, Y. Huang, J. Chen, and P. A. Naylor, "Adaptive algorithms for the identification of sparse impulse responses," in *Topics in Acoustic Echo and Noise Control*, E. Hänsler and G. Schmidt, Eds., pp. 125–153. Berlin: Springer, 2006.

[12] P. A. Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation," *J. Sign. Process.*, vol. 86, no. 6, pp. 1182–1192, 2006.

[13] K. Helwani, H. Buchner, and S. Spors, "Multichannel adaptive filtering with sparseness constraints," in *Proc. Int. Workshop on Acoustic Signal Enhancement*, 2012.

[14] G. Enzner, "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 393–396.

[15] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 325–328, 2009.

[16] C. Antweiler and G. Enzner, "Perfect sequence LMS for rapid acquisition of continuous-azimuth head related impulse responses," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011, pp. 281–284.

[17] J. He, R. Ranjan, W.-S. Gan, N. K. Chaudhary, N. D. Hai, and R. Gupta, "Fast continuous measurement of HRTFs with unconstrained head movements for 3D audio," *J. Audio Eng. Soc.*, vol. 66, no. 1, pp. 884–900, 2018.

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

[19] C. Urbanietz and G. Enzner, "Spatial-Fourier retrieval of head-related impulse responses from fast continuous-azimuth recordings in the time-domain," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019.

[20] N. Hahn and S. Spors, "Comparison of continuous measurement techniques for spatial room impulse responses," in *Europ. Signal Process. Conf.*, 2016, pp. 1638–1642.

[21] N. Hahn, W. Hahne, and S. Spors, "Dynamic measurement of binaural room impulse responses using an optical tracking system," in *Proc. Int. Conf. Spatial Audio*, 2017, pp. 16–21.

[22] N. Hahn and S. Spors, "Simultaneous measurement of spatial room impulse responses from multiple sound sources using a continuous moving microphone," in *Europ. Signal Process. Conf.*, 2018.

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

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

[25] J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in *Proc. 114th Conv. Audio Eng. Soc.*, Conv. Paper 5788, 2003.

[26] S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with higher order ambisonics - objective measurements and validation of a 4th order spherical microphone," in *Proc. 120th Conv. Audio Eng. Soc.*, Conv. Paper 6857, 2006.

[27] E. G. Williams, *Fourier Acoustics – Sound Radiation and Nearfield Acoustical Holography*, Academic Press, 1999.

[28] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications*, Springer, third edition, 2019.

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

[30] H. Kuttruff, *Room Acoustics*, Spon Press, sixth edition, 2016.

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

[32] B. Masiero and M. Pollow, "A review of the compressive sampling framework in the lights of spherical harmonics: Applications to distributed spherical arrays," in *Proc. 2nd Int. Symposium on Ambisonics and Spherical Acoustics*, 2010.

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

[34] Y. Takida, S. Koyama, and H. Saruwatari, "Exterior and interior sound field separation using convex optimization: Comparison of signal models," in *Europ. Signal Process. Conf.*, 2018.