KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Jesper Brunnström, Martin Bo Møller, Marc Moonen Title: Bayesian sound field estimation for moving microphones Journal/Conference/Book: IEEE Open Journal of Signal Processing Vol/Book Chapter: Pages: ISBN:
Archived version	■ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ☐ FWO ■ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Published version	
Copyright statement	
Author contact	jesper.brunnstroem@kuleuven.be



IEEE Open Journal of Signal Processing

Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version 01 September, 2024.

Bayesian sound field estimation using moving microphones

JESPER BRUNNSTRÖM[1, 2] (Graduate student member, IEEE), MARTIN BO MØLLER[2], AND MARC MOONEN[1] (Fellow, IEEE)

STADIUS Center for Dynamical Systems, Department of Electrical Engineering (ESAT), KU Leuven, Leuven 3001, Belgium
Bang & Olufsen, Acoustics R&D, Struer, Denmark

CORRESPONDING AUTHOR: JESPER BRUNNSTRÖM (e-mail: jesper.brunnstroem@kuleuven.be).

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: 'Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS'. The scientific responsibility is assumed by the authors.

ABSTRACT A sound field estimation method using moving microphones is proposed. The sound field is modelled with an infinite sequence of spherical harmonic coefficients for each frequency in the discrete Fourier transform domain, and the harmonic coefficients are estimated using Bayesian inference. The proposed method allows for the use of microphones with arbitrary directivities and moving along any trajectory. It can also be adapted for simultaneous measurement with multiple loudspeakers. The approach avoids truncation of the harmonic coefficients, as well as the specification of an expansion center. The proposed method is evaluated through simulations and compared against other state-of-the-art methods. The proposed method is shown to produce estimates with low error, and can be applied in a wider range of situations compared to other methods.

INDEX TERMS Sound field estimation, moving microphone, spherical harmonics, Bayesian inference, Hilbert space

I. INTRODUCTION

Sound field reproduction tasks generally require knowledge of the acoustic response from a set of loudspeakers to a region of interest (ROI) in a room [1]. The response associated with all loudspeakers for a single point can be measured simultaneously using any of the techniques in [2]–[5]. However, the region must be sampled densely in relation to the wave-length of the highest frequency to accurately reconstruct the continuous sound field [6]. For a larger ROI many measurements are required, leading to a time-consuming and error-prone measurement process. Therefore, the development of effective sound field estimation methods is crucial to make the measurement process less costly.

Better sound field estimation generally comes from the exploitation of prior information of the signals, room geometry, or the behaviour of sound. One class of methods impose prior knowledge in terms of sparsity. A signal can be reconstructed from fewer samples if it is sparse in a known basis. There are many ways of choosing a basis as well as performing sparsity-promoting optimization, giving rise to many different methods [7]–[12]. Another class of methods uses machine learning techniques to to learn the structure of the acoustic responses, from prior collected data [13]–[15], or only the data available at measurement time [16], [17].

One important form of prior knowledge is the physical properties of sound. These properties can be encoded in a natural way by representing the sound field in terms of spherical harmonics, plane waves [18], or eigenmodes of an enclosure [19], all of which are solutions to the wave equation. The measurement data can then be used to estimate a set of coefficients representing the sound field in the ROI as a whole [20]–[22]. The coefficient set is in general infinite, and is conventionally truncated or discretized to be computationally tractable. However, there are some methods which avoid truncation, estimating the infinite set of coefficients through a Bayesian approach [23], or estimating a continuous function through kernel ridge regression [24], [25] or Gaussian process regression [26], [27].

The sound field estimation methods discussed so far assume stationary microphones, which requires either many

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

microphones that can be used simultaneously, or a single microphone that is moved between measurement positions. The measurement time or cost could be reduced by using a continuously moving microphone instead. Such methods have been developed in [28]–[32]. A method using multiple loudspeakers simultaneously has been developed in [33], relying on a perfect sequence (PSEQ) as a transmitted signal [34], [35], a signal particularly suited to such acoustic measurements. The method performs well for a 1-dimensional trajectory, but estimating the sound field in 2 and 3-dimensional regions generally requires prohibitively long measurement times. As in the stationary case, the measurement time can be reduced by exploiting prior information, which is done in the form of sparsity in [36], [37], and in the form of a spherical harmonic representation in [38]. With a PSEQ as a transmitted signal, position tracking of the microphone using only the recorded sound can be done [39], leading to minimal cost associated with moving microphone measurements.

The contribution of this paper is a sound field estimation method building on the infinite-dimensional harmonic analysis of [23], using a moving microphone. The method considers the sound recorded by a moving microphone of arbitrary directivity, taking both translation and rotation into account. The sound field is modelled with an infinite sequence of spherical harmonic coefficients for each frequency in the discrete Fourier transform (DFT) domain, which are then estimated using Bayesian inference. The approach avoids truncation of the harmonic coefficients, as well as the specification of an expansion center. The proposed method allows for microphones with arbitrary directivities moving along any trajectory, and can be adapted for simultaneous measurement with multiple loudspeakers. Code for the considered sound field estimation methods is made available under a permissive license at github.com/sounds-research/aspcol.

The paper is structured as follows. In Section II the problem is defined more precisely, along with a fundamental model of the measured data. In Section III a linear measurement model based on spherical harmonics is constructed, and in Section IV a Bayesian inference method for the unknown harmonic coefficients is presented. In Section V the proposed method is compared against existing methods, and finally conclusions are made in Section VI.

A. NOTATION

The set of real numbers is represented by $\mathbb{R}$, complex numbers by $\mathbb{C}$, integers by $\mathbb{Z}$, natural numbers including 0 by $\mathbb{N}$, and the points on the unit sphere in $\mathbb{R}^3$ by $\mathbb{S}^2$. A unit vector derived from a vector $\boldsymbol{a}$ is denoted by $\hat{\boldsymbol{a}} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$. The complex conjugate of a variable a is denoted by a^*, the Hermitian adjoint by a^{H}, and the imaginary unit by $i = \sqrt{-1}$. The Hilbert space of square integrable functions from $\mathbb{S}^2$ to $\mathbb{C}$ is represented by $L^2(\mathbb{S}^2)$. Selecting an element from a variable a is denoted by $(a)_n$, $(a)_{nm}$ or $(a)_n^m$, for vectors, matrices and ℓ^2 sequences respectively.

II. PROBLEM STATEMENT

The typical goal of sound field estimation for sound field reproduction is to determine the response from one or more loudspeakers in a region. For the time being, one loudspeaker is considered, transmitting a known signal $\phi(n) \in \mathbb{R}$, where $n \in \mathbb{Z}$ represents the discrete time index. The loudspeaker has a room impulse response (RIR) associated with each position $\boldsymbol{r} \in \mathbb{R}^3$ in space, representing the acoustic response of the loudspeaker, room, and the microphone placed at $\boldsymbol{r}$. The RIRs are modelled as finite impulse response (FIR) filters of length L, where the lth coefficient is denoted by $h(\boldsymbol{r}, l) \in \mathbb{R}$. The sound recorded by a stationary microphone at position $\boldsymbol{r}$ is

$$p(\boldsymbol{r}, n) = \sum_{l=0}^{L-1} h(\boldsymbol{r}, l)\phi(n-l). \tag{1}$$

The RIR can be represented in terms of its frequency-domain components $H(\boldsymbol{r}, \omega_m) \in \mathbb{C}$ with the inverse DFT as

$$h(\boldsymbol{r}, l) = \frac{1}{L}\sum_{m=0}^{L-1} H(\boldsymbol{r}, \omega_m) e^{-i2\pi lm/L}. \tag{2}$$

The sampled angular frequency ω_m corresponds to the positive continuous angular frequency $\frac{2\pi f_s m}{L}$ for $0 \leq m \leq \frac{L}{2}$ where f_s is the sampling rate. The remaining $\frac{L}{2} < m < L$ frequency-domain components can be defined according to $H(\boldsymbol{r}, \omega_m) = H(\boldsymbol{r}, \omega_{L-m})^*$, due to it being the DFT of a real-valued sequence.

To model the motion of a moving microphone a trajectory $(\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N-1})$ is considered, consisting of N positions $\boldsymbol{r}_n \in \mathbb{R}^3$. Inserting (2) into (1), the microphone signal along the trajectory is

$$p(\boldsymbol{r}_n, n) = \sum_{m=0}^{L-1} H(\boldsymbol{r}_n, \omega_m)\Phi(n, \omega_m), \tag{3}$$

where $\Phi(n, \omega_m) = \frac{1}{L}\sum_{l=0}^{L-1} e^{-i2\pi lm/L}\phi(n-l)$ which can be viewed as a short-time Fourier transform of $\phi(n)$ with window size L and hop size 1. The RIR for any position $\boldsymbol{r}$ is considered stationary, but (3) is time-dependent due to moving through RIRs associated with different positions. The task is to estimate $H(\boldsymbol{r}, \omega_m)$ for all positions $\boldsymbol{r}$ in the ROI, using data recorded by the moving microphone in (3).

III. SPHERICAL HARMONIC MEASUREMENT MODEL

A. SPHERICAL HARMONIC REPRESENTATION

The sound pressure $u : \mathbb{R}^3 \to \mathbb{C}$ for a given frequency ω in a source free region satisfies the homogenous Helmholtz equation $(\nabla^2 + k^2)u = 0$, where k is the wave number $k = \frac{\omega}{c}$. Such a function can be expanded around a position $\boldsymbol{r}_c \in \mathbb{R}^3$, and expressed exactly in a new basis [18, Eq. 6.140]

$$u(\boldsymbol{r}, \omega) = \sum_{\nu,\mu} \alpha_\nu^\mu(\boldsymbol{r}_c, \omega)\varphi_\nu^\mu(\boldsymbol{r} - \boldsymbol{r}_c, \omega), \tag{4}$$

where $\alpha_\nu^\mu(\boldsymbol{r}_c, \omega) \in \mathbb{C}$ is the harmonic coefficient of νth order and μth degree. The summation is shorthand for summing

over $\nu \in \mathbb{N}$ and $\mu \in \mathbb{Z}$ satisfying $-\nu \leq \mu \leq \nu$. The basis function

$$\varphi_\nu^\mu(\boldsymbol{r}, \omega) = \sqrt{4\pi} j_\nu(k\|\boldsymbol{r}\|_2) Y_\nu^\mu(\hat{\boldsymbol{r}}) \tag{5}$$

is constructed from spherical Bessel functions of the first kind j_ν, and spherical harmonic functions Y_ν^μ [18].

The equality in (4) only holds in general for an infinite number of harmonic coefficients. To perform the later analysis, it will be practical to consider all of the coefficients jointly as an infinite sequence. Under some ordering of ν, μ, the harmonic coefficients can be collected into an infinite sequence $\alpha(\boldsymbol{r}, \omega) \in \ell^2(\mathbb{C})$, where $\ell^2(\mathbb{C})$ is the Hilbert space of square summable sequences with complex coefficients and inner product $\langle \alpha, \alpha' \rangle = \sum_{\nu,\mu} \alpha_\nu^\mu (\alpha_\nu'^\mu)^*$ [40, 3.1-6]. The summability constraint is physically motivated [41].

B. RECORDED MICROPHONE SIGNAL

To make use of directional microphones in the measurement process, the directivity is modelled explicitly. In this section such a model will be introduced, and then related to the spherical harmonic representation in (4) following the approach of [23, Supplementary material].

The sound pressure function in (4) can be expressed exactly using a superposition of plane waves expanded around the position $\boldsymbol{r}_c$ [42, Eq. 3.66]

$$u(\boldsymbol{r}, \omega) = \int_{\mathbb{S}^2} \beta_{\hat{\boldsymbol{d}}}(\boldsymbol{r}_c, \omega) e^{-ik(\boldsymbol{r}-\boldsymbol{r}_c)^\top \hat{\boldsymbol{d}}} \, ds(\hat{\boldsymbol{d}}), \tag{6}$$

where $\hat{\boldsymbol{d}} \in \mathbb{S}^2$ represents the incoming plane wave direction, which is opposite of the propagation direction, s is the natural spherical measure, and $\beta_{\hat{\boldsymbol{d}}}(\boldsymbol{r}_c, \omega) \in L^2(\mathbb{S}^2)$ is a weighting function for the particular sound field. Consider a microphone located at $\boldsymbol{r}_c$ with a directivity function $\gamma(\hat{\boldsymbol{d}}, \omega) \in L^2(\mathbb{S}^2)$, representing the microphone signal for a plane wave incoming from direction $\hat{\boldsymbol{d}}$. The recorded microphone signal will then be given as

$$s(\boldsymbol{r}_c, \omega) = \int_{\mathbb{S}^2} \beta_{\hat{\boldsymbol{d}}}(\boldsymbol{r}_c, \omega) \gamma(\hat{\boldsymbol{d}}, \omega) \, ds(\hat{\boldsymbol{d}}). \tag{7}$$

To relate (6) to the spherical harmonic representation (4), the identity $\varphi_\nu^\mu(\boldsymbol{r}) = \int_{\mathbb{S}^2} \frac{i^\nu}{\sqrt{4\pi}} Y_\nu^\mu(\hat{\boldsymbol{d}}) e^{-ik\boldsymbol{r}^\top \hat{\boldsymbol{d}}} \, ds(\hat{\boldsymbol{d}})$ [43, Eq. 3.43] can be inserted into (4) to produce the expression

$$u(\boldsymbol{r}, \omega) = \int_{\mathbb{S}^2} \sum_{\nu,\mu} \alpha_\nu^\mu(\boldsymbol{r}_c, \omega) \frac{i^\nu}{\sqrt{4\pi}} Y_\nu^\mu(\hat{\boldsymbol{d}}) e^{-ik(\boldsymbol{r}-\boldsymbol{r}_c)^\top \hat{\boldsymbol{d}}} \, ds(\hat{\boldsymbol{d}}). \tag{8}$$

Choosing $\beta_{\hat{\boldsymbol{d}}}(\boldsymbol{r}_c, \omega) = \sum_{\nu,\mu} \alpha_\nu^\mu(\boldsymbol{r}_c, \omega) \frac{i^\nu}{\sqrt{4\pi}} Y_\nu^\mu(\hat{\boldsymbol{d}})$ shows that (8) constitutes a plane wave representation of the sound field as in (6). That means that according to (7) the recorded microphone signal will be given as

$$s(\boldsymbol{r}_c, \omega) = \int_{\mathbb{S}^2} \sum_{\nu,\mu} \alpha_\nu^\mu(\boldsymbol{r}_c, \omega) \frac{i^\nu}{\sqrt{4\pi}} Y_\nu^\mu(\hat{\boldsymbol{d}}) \gamma(\hat{\boldsymbol{d}}, \omega) \, ds(\hat{\boldsymbol{d}}). \tag{9}$$

The recorded microphone signal can then be written in a compact form by introducing the directivity coefficients

$$\gamma_\nu^\mu(\omega) = \frac{(-i)^\nu}{\sqrt{4\pi}} \int_{\mathbb{S}^2} \gamma(\hat{\boldsymbol{d}}, \omega)^* Y_\nu^\mu(\hat{\boldsymbol{d}})^* \, ds(\hat{\boldsymbol{d}}), \tag{10}$$

and constructing an infinite sequence $\gamma(\omega) \in \ell^2(\mathbb{C})$. The microphone signal (9) can then be written

$$\begin{aligned} s(\boldsymbol{r}_c, \omega) &= \sum_{\nu,\mu} \alpha_\nu^\mu(\boldsymbol{r}_c, \omega) \gamma_\nu^\mu(\omega)^* \\ &= \langle \alpha(\boldsymbol{r}_c, \omega), \gamma(\omega) \rangle. \end{aligned} \tag{11}$$

C. TRANSLATION

The microphone signal (11) only directly relates to the harmonic coefficients expanded around the microphone position. In order to use measurements from more than one position, it is desirable to translate the harmonic coefficients between expansion centers. This can be done with the unitary operator $T(\boldsymbol{r}, \omega) : \ell^2(\mathbb{C}) \to \ell^2(\mathbb{C})$, defined as [23, Eq. 6], equivalent to the transpose of [43, Theorem 3.26]

$$(T(\boldsymbol{r}, \omega)\alpha(\boldsymbol{r}', \omega))_\nu^\mu = \sum_{\nu',\mu'} T_{\nu\nu'}^{\mu\mu'}(\boldsymbol{r}, \omega) \alpha_{\nu'}^{\mu'}(\boldsymbol{r}', \omega). \tag{12}$$

The translation coefficients $T_{\nu\nu'}^{\mu\mu'}(\boldsymbol{r}, \omega) \in \mathbb{C}$ can be calculated as

$$\begin{aligned} &T_{\nu\nu'}^{\mu\mu'}(\boldsymbol{r}, \omega) = 4\pi(-1)^\mu i^{\nu'-\nu} \\ &\sum_{q=0}^{\nu+\nu'} i^q j_q(k\|\boldsymbol{r}\|_2) Y_q^{\mu'-\mu}(\hat{\boldsymbol{r}})^* \mathcal{G}(\nu, \mu; \nu', -\mu'; q), \end{aligned} \tag{13}$$

where $\mathcal{G}$ represents the Gaunt coefficient [43, Eq. 3.71]. The translation operator can then be used as

$$\alpha(\boldsymbol{r}, \omega) = T(\boldsymbol{r} - \boldsymbol{r}', \omega)\alpha(\boldsymbol{r}', \omega). \tag{14}$$

The operator satisfies the following important identities

$$\begin{aligned} T(-\boldsymbol{r}, \omega) &= T(\boldsymbol{r}, \omega)^{-1} = T(\boldsymbol{r}, \omega)^{\mathsf{H}} \\ T(\boldsymbol{r} + \boldsymbol{r}', \omega) &= T(\boldsymbol{r}, \omega) T(\boldsymbol{r}', \omega). \end{aligned} \tag{15}$$

D. TIME-DOMAIN HARMONIC COEFFICIENTS

The frequency-domain components $H(\boldsymbol{r}, \omega_m)$ in (3) is defined in the DFT domain for sampled frequencies ω_m, while the microphone model of the previous sections is defined for continuous frequencies ω. In this section the microphone model is considered in the DFT domain, in order to relate it to the data model (3). Due to (3), the frequency-domain components must be considered jointly, meaning care should be taken to ensure both the measurement model and inference procedure leads to a valid sequence of frequency-domain components which corresponds to real-valued RIRs.

Given real-valued time-domain RIRs $h(\boldsymbol{r}, l)$, their frequency-domain components $H(\boldsymbol{r}, \omega_m)$ can be fully represented using only $0 \leq m \leq \frac{L}{2}$. The remaining components for $\frac{L}{2} < m < L$ are defined by the conjugate symmetry $H(\boldsymbol{r}, \omega_m) = H(\boldsymbol{r}, \omega_{L-m})^*$. Using the symmetry to define the frequency-domain components, both basis functions and harmonic coefficients satisfies the same symmetry, $\alpha(\boldsymbol{r}, \omega_m) = \alpha(\boldsymbol{r}, \omega_{L-m})^*$, and $\varphi_\nu^\mu(\boldsymbol{r}, \omega_m) = \varphi_\nu^\mu(\boldsymbol{r}, \omega_{L-m})^*$.

Constructing a measurement model in terms of $\alpha(\boldsymbol{r}, \omega_m)$ using only $0 \leq m \leq \frac{L}{2}$ would make the model nonlinear, due to the required complex conjugate operation.

To simplify the resulting measurement model, a real-valued time-domain sequence of harmonic coefficients is introduced and estimated in lieu of $\alpha(\boldsymbol{r},\omega)$.

The harmonic coefficients associated with all L sampled frequencies are collected into one sequence $\alpha(\boldsymbol{r}) \in \ell^2_L(\mathbb{C})$ defined as

$$\begin{aligned}\alpha(\boldsymbol{r}) &= \big(\alpha(\boldsymbol{r},\omega_0),\dots,\alpha(\boldsymbol{r},\omega_{L-1})\big)\\ \ell^2_L(\mathbb{C}) &= \underbrace{\ell^2(\mathbb{C})\oplus\cdots\oplus\ell^2(\mathbb{C})}_{L\text{ times}},\end{aligned} \tag{16}$$

where $\oplus$ denotes the direct sum, which implies $\ell^2_L(\mathbb{C})$ is also a Hilbert space when equipped with the inner product $\langle\alpha(\boldsymbol{r}),\alpha'(\boldsymbol{r})\rangle = \sum_{m=0}^{L-1}\langle\alpha(\boldsymbol{r},\omega_m),\alpha'(\boldsymbol{r},\omega_m)\rangle$ [44, Definition 6.1]. The projection from the composite space to one of the components $\Pi_m : \ell^2_L(\mathbb{C}) \to \ell^2(\mathbb{C})$ is defined as $\Pi_m\alpha(\boldsymbol{r}) = \alpha(\boldsymbol{r},\omega_m)$. The time-domain harmonic coefficients $\tilde{\alpha}(\boldsymbol{r},l) \in \ell^2(\mathbb{R})$ are defined accordingly

$$\tilde{\alpha}(\boldsymbol{r}) = \big(\tilde{\alpha}(\boldsymbol{r},0),\dots,\tilde{\alpha}(\boldsymbol{r},L-1)\big) \in \ell^2_L(\mathbb{R}). \tag{17}$$

The DFT operator $\mathcal{F} : \ell^2_L(\mathbb{C}) \to \ell^2_L(\mathbb{R})$ is used to relate the two quantities as $\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}) = \alpha(\boldsymbol{r})$, and is defined as

$$\alpha(\boldsymbol{r},\omega_m) = \Pi_m\mathcal{F}\tilde{\alpha}(\boldsymbol{r}) = \sum_{l=0}^{L-1} e^{-i2\pi lm/L}\tilde{\alpha}(\boldsymbol{r},l). \tag{18}$$

Due to (18) and because $\tilde{\alpha}(\boldsymbol{r})$ is real-valued, $\alpha(\boldsymbol{r})$ will be conjugate symmetric.

Because the harmonic coefficients at an arbitrary position should be conjugate symmetric, the translation operator must only produce harmonic coefficients satisfying the constraint. The sampled frequencies ω_m for $0 \le m \le \frac{L}{2}$ corresponds to positive frequencies, and so the harmonic coefficients can be translated using $T(\boldsymbol{r},\omega_m)$, while for $\frac{L}{2} < m < L$ the translated harmonic coefficients can be determined from the conjugate symmetry. The translation operator $T(\boldsymbol{r},\omega)$ in (14) can be extended to $T(\boldsymbol{r}) : \ell^2_L(\mathbb{C}) \to \ell^2_L(\mathbb{C})$ to translate the harmonic coefficients for all sampled frequencies by defining it as

$$\begin{aligned}&\alpha(\boldsymbol{r},\omega_m) = \Pi_m T(\boldsymbol{r}-\boldsymbol{r}')\alpha(\boldsymbol{r}')\\ &= \begin{cases} T(\boldsymbol{r}-\boldsymbol{r}',\omega_m)\alpha(\boldsymbol{r}',\omega_m) & \text{if } 0 \le m < \frac{L}{2}\\ T(\boldsymbol{r}-\boldsymbol{r}',\omega_{L-m})^*\alpha(\boldsymbol{r}',\omega_m) & \text{if } \frac{L}{2} < m < L\\ \Re\{T(\boldsymbol{r}-\boldsymbol{r}',\omega_m)\}\alpha(\boldsymbol{r}',\omega_m) & \text{if } m = \frac{L}{2}.\end{cases}\end{aligned} \tag{19}$$

where $\Re\{\cdot\}$ is the real part operator, projecting the complex argument onto the real line. Conjugate symmetry forces the imaginary part of $\alpha(\boldsymbol{r},\omega_m)$ for $m = 0$ to be zero, and in the case of an even L for $m = \frac{L}{2}$ as well. The translation operator must then be real-valued for those frequencies. For $m = 0$ the operator $T(\boldsymbol{r},0)$ is guaranteed to be real-valued [43, Eq. 3.83]. The operator $\Re\{T(\boldsymbol{r},\omega_{\frac{L}{2}})\}$ does not necessarily produce translated harmonic coefficients that correspond to their continuous frequency counterpart. Nevertheless, under the assumption that sampling is preceded by ideal anti-aliasing filtering with a zero output at the Nyquist frequency, there will be no error incurred.

The time-domain harmonic coefficients can be translated by transforming the harmonic coefficients between the time- and frequency-domain before and after the translation. This can be explicitly expressed as

$$\tilde{\alpha}(\boldsymbol{r}) = \tilde{T}(\boldsymbol{r}-\boldsymbol{r}')\tilde{\alpha}(\boldsymbol{r}') = \mathcal{F}^{-1}\,T(\boldsymbol{r}-\boldsymbol{r}')\,\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}'). \tag{20}$$

Both $T(\boldsymbol{r})$ and $\tilde{T}(\boldsymbol{r})$ satisfy the properties shown in (15).

The directivity coefficients for all frequencies $\gamma \in \ell^2_L(\mathbb{C})$, defined in the same way as the harmonic coefficients, i.e., $\gamma = (\gamma(\omega_0),\dots,\gamma(\omega_{L-1}))$, should be constructed to satisfy conjugate symmetry. If the coefficients $\gamma(\omega_m)$ are chosen for $0 \le m \le \frac{L}{2}$, the coefficients for $\frac{L}{2} < m < L$ are then determined by $\gamma(\omega_m) = \gamma(\omega_{L-m})^*$.

E. MEASUREMENT MODEL

The RIR $h(\boldsymbol{r},l)$ represents the sound recorded by a microphone from a transmitted unit impulse, therefore the model in (11) is also valid for the RIR. The frequency-domain components of the RIR along the trajectory can be expressed in terms of the time-domain harmonic coefficients using (11), (14) and (18) as

$$H(\boldsymbol{r}_n,\omega_m) = \big\langle \Pi_m T(\boldsymbol{r}_n-\boldsymbol{r})\,\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}_c), \gamma(\omega_m,n)\big\rangle. \tag{21}$$

Inserting (21) into the dynamic measurement model (3), the resulting model is

$$p(\boldsymbol{r}_n,n) = \big\langle \Phi(n)T(\boldsymbol{r}_n-\boldsymbol{r})\,\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}), \gamma(n)\big\rangle, \tag{22}$$

where $\Phi(n) : \ell^2_L(\mathbb{C}) \to \ell^2_L(\mathbb{C})$ is defined as $\Pi_m\Phi(n)\alpha(\boldsymbol{r}) = \Phi(n,\omega_m)\alpha(\boldsymbol{r},\omega_m)$. The operator $\Phi(n)$ is conjugate symmetric by construction, as the DFT of a real-valued sequence $\phi(n)$. The time-varying directivity $\gamma(n)$ is introduced to allow for one set of directivity coefficients γ per time-index n to allow for a rotation as the microphone moves.

A measurement operator $\Xi(\boldsymbol{r}) : \ell^2_L(\mathbb{R}) \to \mathbb{R}^N$ is defined as

$$\Xi(\boldsymbol{r})\tilde{\alpha}(\boldsymbol{r}) = \begin{bmatrix} \big\langle \Phi(0)T(\boldsymbol{r}_0-\boldsymbol{r})\,\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}), \gamma(0)\big\rangle \\ \vdots \\ \big\langle \Phi(N-1)T(\boldsymbol{r}_{N-1}-\boldsymbol{r})\,\mathcal{F}\,\tilde{\alpha}(\boldsymbol{r}), \gamma(N-1)\big\rangle \end{bmatrix}, \tag{23}$$

which leads to the linear measurement model

$$\boldsymbol{p} = \Xi(\boldsymbol{r})\tilde{\alpha}(\boldsymbol{r}) + \boldsymbol{\epsilon}, \tag{24}$$

where $\boldsymbol{p} \in \mathbb{R}^N$ is the recorded samples with $(\boldsymbol{p})_n = p(\boldsymbol{r}_n,n)$ and $\boldsymbol{\epsilon} \in \mathbb{R}^N$ models the measurement noise.

IV. ESTIMATION OF HARMONIC COEFFICIENTS

A. BAYESIAN INFERENCE

A Bayesian approach will be used to solve the inverse problem of estimating $\tilde{\alpha}(\boldsymbol{r})$ in (24). The noise $\boldsymbol{\epsilon}$ is assumed to be distributed according to a zero-mean Gaussian distribution $\mathcal{N}(0,\boldsymbol{\Sigma})$, where the noise covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ is assumed to be known. A prior distribution is chosen for the time-domain harmonic coefficients $\tilde{\alpha}(\boldsymbol{r}) \sim \mathcal{N}(0,\lambda I)$, where I is the identity operator, which does not encode any preference for either position or harmonic order. It is an improper prior,

because the covariance operator is not of trace-class and therefore does not define a proper probability measure [45, Proposition 1.8]. However, the posterior is still guaranteed to be proper [46], [47], which is sufficient for the purposes of Bayesian inference.

Given the linear measurement model (24) and the chosen Gaussian priors, the posterior distribution of the harmonic coefficients is [48, Example 6.23]

$$\rho_{\tilde{\alpha}(\boldsymbol{r})|\boldsymbol{p}} = \mathcal{N}(m_{\tilde{\alpha}}(\boldsymbol{r}), \Sigma_{\tilde{\alpha}}(\boldsymbol{r})). \tag{25}$$

The mean and variance are

$$\begin{aligned} m_{\tilde{\alpha}}(\boldsymbol{r}) &= \Xi(\boldsymbol{r})^{\mathsf{H}}(\boldsymbol{\Psi} + \lambda^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{p} \\ \Sigma_{\tilde{\alpha}}(\boldsymbol{r}) &= \lambda\Big(I - \Xi(\boldsymbol{r})^{\mathsf{H}}(\boldsymbol{\Psi} + \lambda^{-1}\boldsymbol{\Sigma})^{-1}\Xi(\boldsymbol{r})\Big). \end{aligned} \tag{26}$$

The adjoint measurement operator $\Xi^{\mathsf{H}}(\boldsymbol{r}) : \mathbb{R}^N \to \ell^2_L(\mathbb{R})$ can be written explicitly as

$$\Xi^{\mathsf{H}}(\boldsymbol{r})\boldsymbol{a} = \sum_{n=0}^{N-1} \mathcal{F}^{-1}\, T(\boldsymbol{r} - \boldsymbol{r}_n)\Phi(n)^{\mathsf{H}}\gamma(n)(\boldsymbol{a})_n \tag{27}$$

for any $\boldsymbol{a} \in \mathbb{R}^N$. The matrix $\boldsymbol{\Psi} = \Xi(\boldsymbol{r})\Xi^{\mathsf{H}}(\boldsymbol{r}) \in \mathbb{R}^{N\times N}$ is defined as

$$(\boldsymbol{\Psi})_{nn'} = \langle \Phi(n)T(\boldsymbol{r}_n - \boldsymbol{r}_{n'})\Phi(n')^{\mathsf{H}}\gamma(n'), \gamma(n)\rangle. \tag{28}$$

Most microphone directivities can be modelled well by only a few non-zero coefficients in the sequence (10), in which case $\boldsymbol{\Psi}$ can be calculated without error. The harmonic coefficients can then be estimated as the mean of the posterior distribution $m_{\tilde{\alpha}}(\boldsymbol{r})$.

B. RECONSTRUCTION

Although (26) provides a simple estimate for each harmonic coefficient, the full infinite sequence cannot be calculated, as it would require infinite memory. A subset can be calculated, where the number of coefficients required to well represent a sound field depends on the bandwidth and size of the region [41]. However, there exists a closed form expression for the estimated sound field in a given position for a microphone with given directivity coefficients $\gamma_{\text{rec}} \in \ell^2(\mathbb{C})$, which avoids truncation of the harmonic coefficients entirely. Note that the reconstruction directivity coefficients γ_{rec} are not related to the directivity of the microphone used for measurement, but instead allows for reconstruction of the sound field as a microphone with directivity γ_{rec} would have recorded it. Using (11), the RIR a microphone with directivity γ_{rec} would see for any position $\boldsymbol{r}$ can be estimated as

$$H(\boldsymbol{r}, \omega_m) = \langle \alpha(\boldsymbol{r}, \omega_m), \gamma_{\text{rec}}\rangle. \tag{29}$$

By inserting the expression for $\tilde{\alpha}(\boldsymbol{r})$ from (26) and the measurement operator (23), the expression becomes

$$\begin{aligned} H(\boldsymbol{r}, \omega_m) &= \langle \Pi_m\, \mathcal{F}\, \tilde{\alpha}(\boldsymbol{r}), \gamma_{\text{rec}}\rangle \\ &= \langle \Pi_m\, \mathcal{F}\, \Xi(\boldsymbol{r})^{\mathsf{H}}\boldsymbol{v}, \gamma_{\text{rec}}\rangle, \end{aligned} \tag{30}$$

where $\boldsymbol{v} = (\boldsymbol{\Psi} + \lambda^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{p}$. The definition of the adjoint measurement operator can be used, simplifying the expression to

$$H(\boldsymbol{r}, \omega_m) = \sum_{n=0}^{N-1} \langle \Pi_m T(\boldsymbol{r} - \boldsymbol{r}_n)\gamma(n), \gamma_{\text{rec}}\rangle \Phi(n, \omega_m)^*(\boldsymbol{v})_n \tag{31}$$

which can be calculated for $0 \le m \le \frac{L}{2}$, and the other frequencies determined by $H(\boldsymbol{r}, \omega_m) = H(\boldsymbol{r}, \omega_{L-m})^*$.

The expression (31) can be calculated exactly for finite order microphone directivities. A finite order directivity means that the two directivities satisfy the property $(\Pi_m\gamma(n))^\mu_\nu = 0$ and $(\gamma_{\text{rec}})^\mu_\nu = 0$ for $\nu > D$ for some D. Then the inner product in (31) can be computed as

$$\begin{aligned} &\langle \Pi_m T(\boldsymbol{r})\gamma(n), \gamma_{\text{rec}}\rangle \\ &= \sum_{\nu,\nu'=0}^{D} \sum_{\substack{|\mu|\le\nu \\ |\mu'|\le\nu'}} T^{\mu\mu'}_{\nu\nu'}(\boldsymbol{r}, \omega_m)(\gamma(n, \omega_m))^{\mu'}_{\nu'}(\gamma^*_{\text{rec}})^\mu_\nu, \end{aligned} \tag{32}$$

which is a finite sum of known quantities, and can therefore be computed exactly. For the case of $m = \frac{L}{2}$ only the real part of $T^{\mu\mu'}_{\nu\nu'}(\boldsymbol{r}, \omega_m)$ should be used, analogous to (19). A similar expression can also be used to calculate the matrix $\boldsymbol{\Psi}$ in (28) exactly.

Commonly the quantity of interest is the sound pressure, in which case an omnidirectional directivity $\gamma_{\text{rec}} = \gamma_0$ should be chosen, which is defined as $(\gamma_0)^\mu_\nu = \delta(\nu)$.

C. OMNIDIRECTIONAL MICROPHONES

In this section closed form expressions are given for the common special case where the sound pressure is estimated using an omnidirectional microphone, meaning that $\gamma(n, \omega_m) = \gamma_0$ for all n and m. The expressions are simplified using the relationship $\langle T(\boldsymbol{r}, \omega)\gamma_0, \gamma_0\rangle = j_0(\frac{\omega}{c}\|\boldsymbol{r}\|_2)$, which corresponds to the spatial correlation of a diffuse sound field [49]. Assuming that L is even for ease of presentation, expression (28) can be simplified to

$$\begin{aligned} (\boldsymbol{\Psi})_{nn'} = &\sum_{m=0}^{\frac{L}{2}} j_0\big(\frac{\omega_m}{c}\|\boldsymbol{r}_n - \boldsymbol{r}_{n'}\|_2\big)\Phi(n, \omega_m)\Phi(n', \omega_m)^* \\ &+ \sum_{m=1}^{\frac{L}{2}-1} j_0\big(\frac{\omega_m}{c}\|\boldsymbol{r}_n - \boldsymbol{r}_{n'}\|_2\big)\Phi(n, \omega_m)^*\Phi(n', \omega_m). \end{aligned} \tag{33}$$

The sound field reconstruction expression (31) can also be simplified to

$$H(\boldsymbol{r}, \omega_m) = \sum_{n=0}^{N-1} j_0\big(\frac{\omega_m}{c}\|\boldsymbol{r} - \boldsymbol{r}_n\|_2\big)\Phi(n, \omega_m)^*(\boldsymbol{v})_n. \tag{34}$$

The zeroth order spherical Bessel function is the unnormalized sinc function, and is easy to compute. From (33) and (34), the sound field estimate can be computed in closed form. Similar, but slightly more complicated, expressions can be obtained for other low-order microphone directivities.

D. FIRST ORDER DIRECTIONAL MICROPHONES

Many directional microphones can be characterized by the directionality function

$$\gamma(\hat{\boldsymbol{d}}, \omega) = A + (1 - A)\hat{\boldsymbol{d}}^\top \hat{\boldsymbol{d}}_{\text{mic}}, \tag{35}$$

where $A \in [0, 1]$ is a parameter and $\hat{\boldsymbol{d}}_{\text{mic}} \in \mathbb{S}^2$ is the direction of the microphone. This model includes many common microphone directivities such as the cardioid, figure-8, and the supercardioid, depending on the value of the parameter A. Translated into the spherical harmonic domain via (10) the coefficients of $\gamma \in \ell^2(\mathbb{C})$ are [23, Supplementary material]

$$\gamma_\nu^\mu = \begin{cases} A & \text{if } \nu = 0 \\ \frac{-Ai\sqrt{4\pi}}{3} Y_1^\mu(\hat{\boldsymbol{d}}_{\text{mic}})^* & \text{if } \nu = 1 \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

For directionality functions that can be expressed as higher order polynomials of $\hat{\boldsymbol{d}}^\top \hat{\boldsymbol{d}}_{\text{mic}}$, the integral (10) can still be solved analytically using the Funk-Hecke formula [43, Eq. 3.45]. Otherwise, the integral can be computed numerically.

E. MULTIPLE LOUDSPEAKERS AND MICROPHONES

The proposed method is not limited to measurements for a single loudspeaker at a time. One common signal ϕ can be chosen, and each loudspeaker assigned a delayed version of the same signal such that $\phi_s(n) = \phi(n - sL)$ for $s = 0, \ldots, S - 1$ where S is the number of loudspeakers. Then the measurement model in (1) for multiple loudspeakers can be written as

$$p(\boldsymbol{r}, n) = \sum_{l=0}^{SL-1} \underline{\mathrm{h}}(\boldsymbol{r}, l)\phi(n - l), \tag{37}$$

where $\underline{\mathrm{h}}$ is a concatenation of the RIRs associated with all loudspeakers. The multiple loudspeaker system appears as an equivalent single-source system, consisting of a single spatially distributed source with internal delays. Note that (31) estimates the frequency components of the concatenated RIR. To separate the components associated with each loudspeaker, the estimate (31) is computed for all L sampled frequencies and then the inverse DFT returns the estimate of the time-domain concatenated RIR $\underline{\mathrm{h}}(\boldsymbol{r}, l)$.

Depending on the choice of signal ϕ, the equivalent single-source system is not necessarily identifiable [2]. One way to ensure a well-defined problem is to choose ϕ to be a PSEQ with a period of SL samples [3], [33], meaning it should satisfy $\sum_{n=0}^{SL-1} \phi(n)\phi(n + l) \propto \delta(l \bmod SL)$, where $\propto$ denotes proportionality.

The method is not limited to a single microphone at a time. Sound can be recorded with M microphones using a periodic signal ϕ for an integer number of periods. Then the trajectories and recorded audio signals from the M microphones can be concatenated, appearing like a single microphone with a potentially non-continuous trajectory.

If the sound pressure is of interest instead of the RIRs, the signal ϕ in the model can be set to $\phi(n) = \delta(n \bmod L)$ regardless of the true loudspeaker signal. Under a periodic sound field with period L, the choice results in $p(\boldsymbol{r}_n, n) = h(\boldsymbol{r}_n, n)$, so the derivations correspond to estimating the sound pressure.

V. EXPERIMENTS

A. SIMULATION

Simulations are performed, with ideal linear microphones and loudspeakers. Modelling an ideal sampling procedure, the moving microphone is simulated by generating one RIR for each sampled spatial position, convolving the full signal with each RIR, and selecting one output sample per sampled position. The room impulse responses are generated according to the image-source method [50] using [51], for a room of size $5.4\,\text{m} \times 3.2\,\text{m} \times 2.1\,\text{m}$. Cardioid microphones are simulated as first order differential microphone arrays [52, Section 3.3] using the response of two omnidirectional microphones spaced by $1\,\text{mm}$.

All mean squared error (MSE) values shown are normalized by the mean square of the true value. The zeroth frequency bin is not considered in any of the experiments. Frequencies below $50\,\text{Hz}$ are not considered for experiment with cardioid microphones, as the differential array method has numerical problems for very low frequencies, and so the results would not be accurate.

B. METHODS AND PARAMETERS

1) Nearest neighbour

This method uses stationary omnidirectional microphones, and returns the unchanged identified RIR associated with the microphone closest to the evaluation position. This method provides a baseline against which other methods should provide a performance improvement.

2) Stationary Bayesian

This method uses stationary microphones, and can be considered the stationary microphone equivalent of the proposed method [23]. Both omnidirectional and cardioid microphones are considered, referred to as *stationary omni* and *stationary cardioid*. The RIRs at the microphone positions are first identified, and are then used to estimate the sound field at the evaluation positions. The prior variance of the harmonic coefficients is set to $\lambda = 10$. *Stationary omni* is equivalent to kernel interpolation presented in [24].

3) Moving Bayesian

This method is the proposed sound field estimation method with moving microphones. Both omnidirectional and cardioid microphones are considered, referred to as *moving omni* and *moving cardioid*. The prior variance of the harmonic coefficients is set to $\lambda = 10$.

4) Spatial spectrum

This method uses a moving omnidirectional microphone, with another spherical harmonic representation, where the sound field is expressed as the spatial spectrum on a reference sphere. Some changes to the literal description of the method in [38] have been made to reduce computational cost and improve estimation performance. For shorter acquisition times as considered here, the system matrix is very poorly conditioned, and the linear system must therefore be solved with regularized least squares. The regularization parameter is denoted by $\rho \in \mathbb{R}_{\geq 0}$. Using notation defined in [38], instead of estimating the coefficients a_{vu}, the modified coefficients $\frac{a_{vu}}{b_v(k_l r_{\max})}$ are estimated, which simplifies both estimation and reconstruction, as well as reducing numerical problems by avoiding division by zero. With the modified coefficients, the signal model in [38, Eq. 12], is closely related to (22), although [38, Eq. 12] is restricted to omnidirectional microphones, has a truncated sequence of harmonic coefficients, and includes coefficients for frequencies above the Nyquist frequency. Frequencies above the Nyquist frequency increase the computational cost considerably without benefit, so in these experiments, the system matrix is only constructed for frequency bins $l = 0, 1, \ldots, \frac{L}{2}$. The suggested truncation order of $\lceil \frac{\omega_m}{c} r_{\text{ref}} \rceil$ for frequency ω_m is used, where r_{ref} is chosen as the distance between the origin and the furthest point on the trajectory.

5) 1D interpolation

This method uses a moving omnidirectional microphone, and is presented in [31] for a single loudspeaker and [33] for multiple loudspeakers. It is particularly suited to estimation on the 1-dimensional trajectory along which the microphone moves. The method relies on the orthogonal properties of a PSEQ as a training signal. Interpolation is performed along the trajectory using a 512 point Lagrange interpolation. The open source implementation from [53] is used.

C. COMPUTATIONAL COST

The computational cost of sound field estimation using moving microphones can be high. The proposed method requires solving a linear system of size $N \times N$. The dependence on N means that the cost becomes high for long acquisition times or high sampling rates. In practice, computing the matrix $\mathbf{\Psi}$ can be the dominant computational cost, requiring the translation operator to be calculated $(\frac{L}{2}+1)N^2$ times. Note that when calculating the translation operator, only the spherical Bessel function is dependent on both position and frequency, while the spherical harmonics are dependent on position, and the Gaunt coefficients on neither.

As a comparison, *spatial spectrum* requires solving a linear system of size $N \times P$, where P is the number of coefficients for all frequencies. The number of coefficients is $P = \sum_{l=0}^{L-1}(M_l+1)^2$, where it is suggested that $M_l = \lceil \frac{\omega_l}{c} r_{\text{ref}} \rceil$. For shorter acquisition times where P is closer to or smaller than N, the system is singular or very poorly conditioned, requiring a regularized least squares method, leading to a system matrix of size $P \times P$.

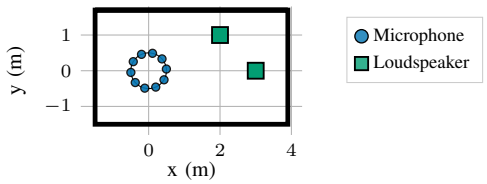


FIGURE 1. The position of the loudspeakers and circular trajectory on a circle. The stationary microphone positions for $M = 10$ are indicated by circles. The black lines shows the walls in the simulated reverberation.

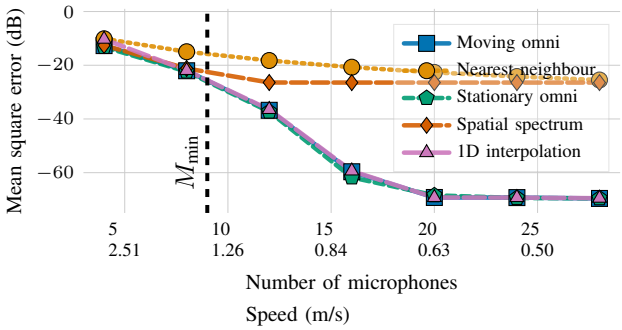


FIGURE 2. Mean square error in dB as a function of the speed of the moving microphone.

D. ESTIMATION ON A CIRCLE

The proposed method is validated by comparing it against existing methods and theoretical results on the simpler task of estimating the sound field on a circle using omnidirectional microphones. The moving microphone moves at a constant speed around a circle of radius $r = 0.5\,\text{m}$. Both the stationary microphones and the 200 evaluation positions are placed equiangularly along the circle. Two loudspeaker are present, for which the RIRs should be simultaneously estimated using the approach in Section IV-E. The position of the circle, loudspeakers, and the stationary microphones are shown in Fig. 1. The simulated room has a reverberation time of $\text{RT}_{60} = 0.2\,\text{s}$, simulated at a sampling rate of $1\,\text{kHz}$, giving a maximum frequency of $f_{\max} = 500\,\text{Hz}$. The experiment uses data without added noise, but a small regularization can still be helpful due to numerical errors, so the noise covariance matrix for both *moving omni* and *stationary omni* is set to $\mathbf{\Sigma} = 10^{-11}\boldsymbol{I}$, and the regularization parameter for *spatial spectrum* is set to $\rho = 10^{-12}$.

The speed of the moving microphone can be compared to the number of stationary microphones using the concept of effective number of microphones M_{eff}, as introduced in [33]. It is related to the speed along the circumference of the circle by $v = \frac{2\pi r f_s}{M_{\text{eff}} L}$. The minimum number of effective microphones for a moving microphone of constant speed is $M_{\min} = 2\lceil \frac{2\pi f_{\max} r}{c} \rceil + 1$, where $\lceil \cdot \rceil$ is the ceiling function [33].

In Fig. 2 the MSE for different speeds of the moving microphone are compared. For these estimates only a single

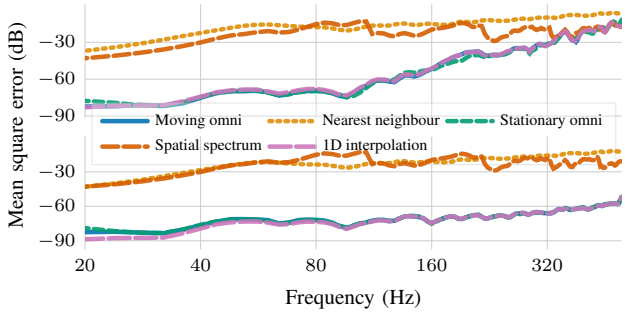


FIGURE 3. **Mean square error in dB as a function of frequency for two loudspeakers, with $M_{\text{eff}} = 10$ (top) and $M_{\text{eff}} = 20$ (bottom).**

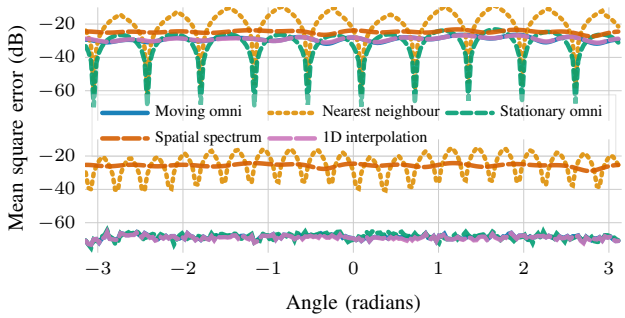


FIGURE 4. **Mean square error in dB as a function of angle for two loudspeakers with $M_{\text{eff}} = 10$ (top) and $M_{\text{eff}} = 20$ (bottom). The lines associated with the *moving omni* and *1D interpolation* overlap.**

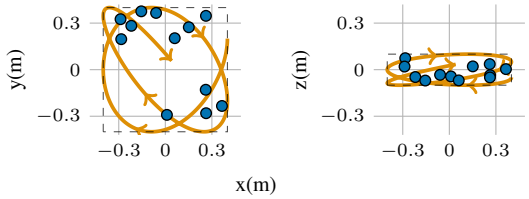


FIGURE 5. **Position of the stationary microphones and the trajectory of the moving microphone. The dashed line shows the boundary of the ROI.**

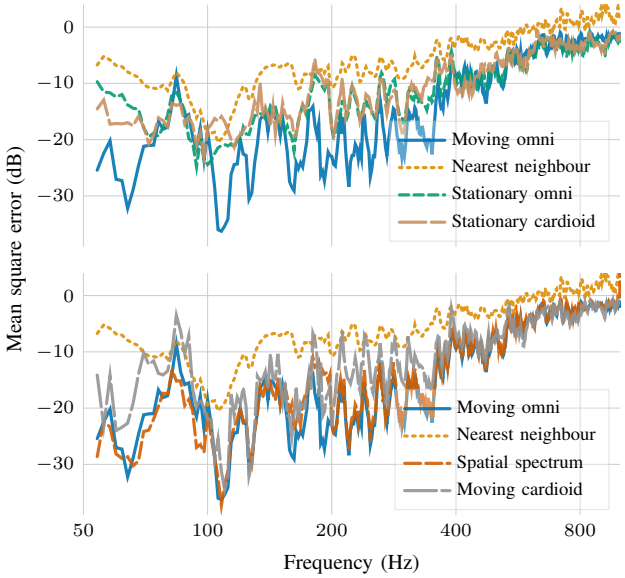


FIGURE 6. **Mean square error in dB as a function of frequency for estimation over the 3-dimensional region. The same experiment is shown on top and bottom, but is split into two for readability purposes.**

loudspeaker is used. Equally many equiangularly spaced stationary microphones are used as the number of effective microphones for the moving microphone. *Moving omni*, *stationary omni*, and *1D interpolation* consistently show the same performance. Note that the error continues to decrease when M_{eff} is chosen even larger than M_{min}, consistent with experiments in [33].

The MSE across all evaluation positions as a function of frequency is shown in Fig. 3 for $M_{\text{eff}} = 10$ and $M_{\text{eff}} = 20$, with a measurement time of 5 s and 10 s respectively. The estimates are now made for two loudspeakers simultaneously. The performance of *moving omni* is similar to *stationary omni* and *1D interpolation*. As the speed decreases, and therefore the amount of data increases, the MSE decreases primarily for the higher frequencies. At the very lowest frequencies, *1D interpolation* performs slightly better than *moving omni* and *stationary omni*, which is due to the regularization introduced in *moving omni* and *stationary omni*. The behaviour is not related to the frequency being low, but rather that the MSE at those frequencies is extremely low, and so then the regularization begins to take effect.

The MSE as a function of angle is shown in Fig. 4. For $M_{\text{eff}} = 10$, all methods except *spatial spectrum* to some extent exhibit a lower error at 10 angles, corresponding to the position of the stationary microphones or one specific part of the periodic training signal. This pattern is only present for *nearest neighbour* for $M_{\text{eff}} = 20$, as the spatial sampling is sufficient for the other methods, as can be seen in Fig. 2.

E. ESTIMATION IN A 3-DIMENSIONAL REGION

In this experiment a more difficult task is considered, where the ROI is a cuboid of size $0.8\,\text{m} \times 0.8\,\text{m} \times 0.2\,\text{m}$, and the sampling rate is set to 2000 Hz, giving a maximum frequency of $f_{\text{max}} = 1000\,\text{Hz}$. Within the ROI a total of 1024 evaluation positions are placed on a uniform grid at a 5 cm distance. White Gaussian noise is added to each microphone at a signal-to-noise ratio of 30 dB. The noise covariance matrices for the Bayesian methods are set to $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ where σ^2 is the measured variance of the noise. The regularization parameter for *spatial spectrum* is set to $\rho = 10^{-5}$.

The positions of the $M = 12$ stationary microphones are sampled from a uniform distribution over the ROI. Cardioid microphones are placed at the same positions as the omnidirectional microphones, with microphone directions each sampled from a uniform distribution on the unit sphere. The moving microphone trajectory is a Lissajous curve, as was used in [29], [38], defined as $\boldsymbol{r}_n = \boldsymbol{a} \circ \cos(\frac{t_n}{f_s}\boldsymbol{f} + \boldsymbol{\theta})$ for a set of frequencies $\boldsymbol{f} = \begin{bmatrix}3, 3.5, 2.5\end{bmatrix}^\top$, a fixed phase-offset $\boldsymbol{\theta} = \begin{bmatrix}0, \frac{\pi}{3}, \frac{2\pi}{3}\end{bmatrix}^\top$, an amplitude $\boldsymbol{a} = \begin{bmatrix}0.4 & 0.4 & 0.1\end{bmatrix}^\top$, an element-wise cos function, and with $\circ$ representing element-wise multiplication. The time steps t_n are chosen such that the speed is approximately constant at $0.8\,\text{m/s}$. The moving cardioid microphone has the same trajectory, with a time-

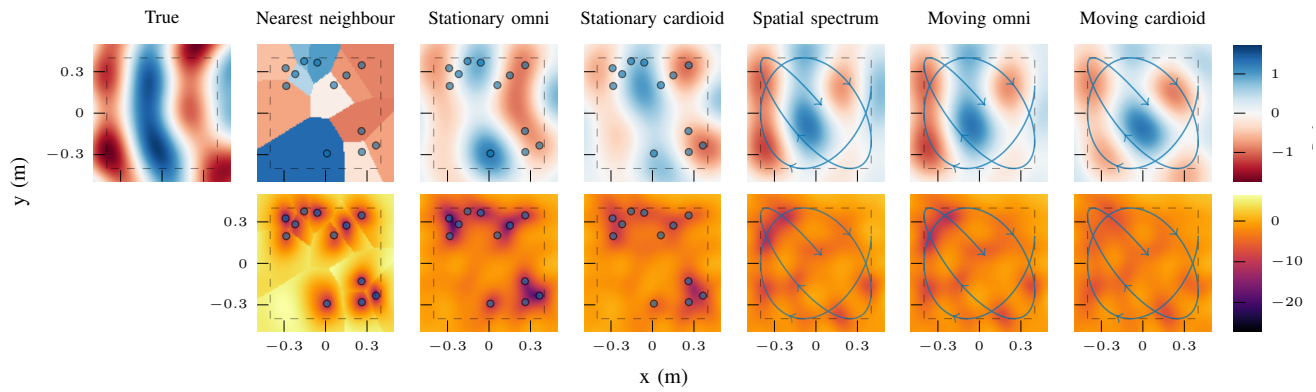


FIGURE 7. The real part of the estimated complex sound pressure at 668 Hz (top row), and the mean square error in decibels over all frequencies (bottom row), at the central horizontal plane in the ROI. The dashed line shows the size of the ROI. The microphone positions projected onto the plane are shown as well.

TABLE 1. Mean normalized mean square error over all frequencies in decibels.

	Omni	Cardioid
Stationary	−5.57	−5.02
Moving	−5.06	−5.00
Nearest neighbour	−0.81	-
Spatial spectrum	−4.82	-

varying microphone direction chosen as $\hat{\boldsymbol{d}}_n = \frac{\boldsymbol{r}_n}{\|\boldsymbol{r}_n\|_2}$. The moving microphone collects samples for 12 periods of the periodic training signal, an equal amount of data as collected by the stationary microphones, giving a measurement time of 6 s. The trajectory and the stationary microphone positions are shown in Fig. 5.

The MSE across all evaluation positions as a function of frequency is shown in Fig. 6. *Moving omni* performs well in comparison to the stationary methods for lower frequencies, but slightly worse for high frequencies. Similar performance is obtained by the use of omnidirectional and cardioid microphones for both moving and stationary methods. *Spatial spectrum* produces similar estimates to *moving omni*, although with slightly higher error. The mean MSE over all frequencies are compared in Table 1, where it can be seen that the *stationary omni* has the lowest error out of all methods, but also that both *moving omni* and *moving cardioid* has a slightly lower error compared to *spatial spectrum*. The MSE was computed in the horizontal plane at $z = 0\,\mathrm{m}$, shown in Fig. 7.

VI. CONCLUSIONS

The task of estimating a sound field using a moving microphone has been considered. Modelling the measured signal using infinite-dimensional spherical harmonic analysis, the sound field can be estimated using Bayesian inference. The method allows microphones with arbitrary directivities moving along any trajectory, as well as simultaneous measurement with multiple loudspeakers. The proposed method was evaluated through simulations and compared against state-of-the-art methods. The proposed method has been shown to produce estimates with low error, and the proposed method can also be applied in a wider range of situations. The formulation of the estimation problem as a Bayesian inference problem allows for the exploitation of additional prior knowledge, which will be considered in future research.

ACKNOWLEDGMENT

The authors would like to thank Natsuki Ueno for very helpful comments on probability measures in infinite-dimensional spaces.

REFERENCES

[1] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, no. 5, May 2017, Art. no. 532.

[2] Y. Huang, J. Benesty, and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Processing*, vol. 86, no. 6, pp. 1278–1295, June 2006.

[3] C. Antweiler, A. Telle, and P. Vary, "NLMS-type system identification of MISO systems with shifted perfect sequences," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2008.

[4] M. Schneider and W. Kellermann, "Large-scale Multiple Input/Multiple Output system identification in room acoustics," *Proc. Mtgs. Acoust.*, vol. 19, no. 1, June 2013.

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

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

[7] N. Antonello, E. D. 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, Oct. 2017.

[8] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, July 2017.

[9] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3770–3779, June 2018.

[10] F. Katzberg, R. Mazur, M. Maass, M. Böhme, and A. Mertins, "Spatial interpolation of room impulse responses using compressed sensing," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2018, pp. 426–430.

[11] S. Koyama and L. Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 1, pp. 172–184, Mar. 2019.

[12] S. Damiano, F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Soundfield reconstruction in reverberant rooms based on compressive sensing and image-source models of early reflections," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2021, pp. 366–370.

[13] F. Lluís, P. Martínez-Nuevo, M. Bo Møller, and S. Ewan Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *J. Acoust. Soc. Am.*, vol. 148, no. 2, pp. 649–659, Aug. 2020.

[14] M. S. Kristoffersen, M. B. Møller, P. Martínez-Nuevo, and J. Østergaard, "Deep sound field reconstruction in real rooms: Introducing the ISOBEL sound field dataset," Feb. 2021, arXiv preprint, arXiv:2102.06455.

[15] K. Shigemi, S. Koyama, T. Nakamura, and H. Saruwatari, "Physics-informed convolutional neural network with bicubic spline interpolation for sound field estimation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2022.

[16] M. Pezzoli, D. Perini, A. Bernardini, F. Borra, F. Antonacci, and A. Sarti, "Deep prior approach for room impulse response reconstruction," *Sensors*, vol. 22, no. 2710, Apr. 2022.

[17] J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, "Kernel interpolation of acoustic transfer functions with adaptive kernel for directed and residual reverberations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, June 2023.

[18] E. G. Williams and J. A. Mann, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, vol. 108, Academic Press, 1999.

[19] T. P. Vu and H. Lissek, "Low frequency sound field reconstruction in a non-rectangular room using a small number of microphones," *Acta Acust.*, vol. 4, no. 2, 2020, Art. no. 5.

[20] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2002, vol. 2, pp. II–1949–II–1952.

[21] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "3D spatial soundfield recording over large regions," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2012.

[22] H. Chen, T. D. Abhayapala, and W. Zhang, "3D sound field analysis using circular higher-order microphone array," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 1153–1157.

[23] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 135–139, Jan. 2018.

[24] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*. Sept. 2018, pp. 436–440.

[25] N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.*, vol. 69, pp. 2383–2395, Apr. 2021.

[26] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 149, no. 2, pp. 1107–1119, Feb. 2021.

[27] E. Fernandez-Grande, D. Caviedes-Nozal, M. Hahmann, X. Karakonstantis, and S. A. Verburg, "Reconstruction of room impulse responses over extended domains for navigable sound field reproduction," in *Proc. Int. Conf. Immersive 3D Audio*, Sept. 2021.

[28] 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, Sept. 2007.

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

[30] 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, May 2017.

[31] N. Hahn and S. Spors, "Continuous measurement of impulse responses on a circle using a uniformly moving microphone," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 2536–2540.

[32] N. Hahn and S. Spors, "Continuous measurement of spatial room impulse responses using a non-uniformly moving microphone," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*. Oct. 2017, pp. 205–208.

[33] N. Hahn and S. Spors, "Simultaneous measurement of spatial room impulse responses from multiple sound sources using a continuously moving microphone," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Sept. 2018, pp. 2180–2184.

[34] H. Luke and H. Schotten, "Odd-perfect, almost binary correlation sequences," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 495–498, Jan. 1995.

[35] A. Telle, C. Antweiler, and P. Vary, "Der perfekte Sweep - Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme," in *Fortschritte der Akustik : 36. Deutsche Jahrestagung für Akustik*. 2010, pp. 341–342.

[36] F. Katzberg and A. Mertins, "Sparse recovery of sound fields using measurements from moving microphones," in *Compressed Sensing in Information Processing*, G. Kutyniok, H. Rauhut, and R. J. Kunsch, Eds., Applied and Numerical Harmonic Analysis, pp. 471–505. Springer International Publishing, 2022.

[37] 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, Nov. 2018.

[38] F. Katzberg, M. Maass, and A. Mertins, "Spherical harmonic representation for dynamic sound-field measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, June 2021, pp. 426–430.

[39] F. Katzberg, M. Maass, R. Pallenberg, and A. Mertins, "Positional tracking of a moving microphone in reverberant scenes by applying perfect sequences to distributed loudspeakers," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sept. 2022.

[40] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1978.

[41] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, June 2007.

[42] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering*, Applied Mathematical Sciences. 4th edition, 2019.

[43] P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves with N Obstacles*, vol. 107 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, 2006.

[44] J. B. Conway, *A Course in Functional Analysis*, Graduate Texts in Mathematics. Springer, 2nd edition, Sept. 1990.

[45] G. Da Prato, *An Introduction to Infinite-Dimensional Analysis*, Springer, 2006.

[46] G. Taraldsen and B. Lindqvist, "Improper priors are not improper," *The American Statistician*, vol. 64, pp. 154–158, May 2010.

[47] B. T. Knapik, A. W. Van Der Vaart, and J. H. Van Zanten, "Bayesian inverse problems with Gaussian priors," *Ann. Statist.*, vol. 39, no. 5, Oct. 2011.

[48] A. M. Stuart, "Inverse problems: A Bayesian perspective," *Acta Numerica*, vol. 19, pp. 451–559, May 2010.

[49] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, Jr., "Measurement of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1072–1077, June 2005.

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

[51] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.

[52] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*, vol. 6 of *Springer Topics in Signal Processing*, Springer, 2013.

[53] N. Hahn, "Impulse response measurement using a moving microphone," Nov. 2017, [Online]. Available: https://github.com/narahahn/continuous_measurement.