Abstract—A method to interpolate the acoustic transfer function (ATF) between regions using kernel ridge regression (KRR) is proposed. Conventionally, the ATF interpolation problem is strongly restricted and situational, depending on knowledge of environmental conditions while not accounting for source position variation. We derive our interpolation function as the solution of an optimization problem defined on a function space where every element holds the acoustic properties of the ATF. By making the space a reproducing kernel Hilbert space (RKHS), we can guarantee that our problem has a known and unique optimizer. The generality of the formulation of this method enables region-to-region estimations, with variable source and receiver within the assigned bounds. The definition of a RKHS also allows for the use of kernel principal component analysis, thereby efficiently providing greater noise robustness to our interpolation function. Our proposed method is compared with a previously established region-to-region interpolation method in numerical simulations where the advantages of the KRR approach are confirmed, showing lower error and greater stability for higher frequencies.

Index Terms—Acoustic transfer function, Helmholtz equation, kernel ridge regression, reproducing kernel Hilbert space, sound field interpolation.

I. INTRODUCTION

The prediction of the variations of a sound signal as it is transmitted between two arbitrary points cannot be performed analytically without prior knowledge of the boundary conditions. Owing to physical phenomena that affect the way sound waves are propagated, the environment may cause changes to the original source signal that must be accounted for. By modeling this effect to be that of a finite impulse response (FIR) filter, we can predict these changes by understanding the impulse response (IR) between the points. This IR is also referred to as the acoustic transfer function (ATF) when in the frequency domain, and it represents the frequency response of the signal recorded at a given receiver point when the source signal was a unit impulse in the time domain. There are many applications of characterizing the acoustics of an environment spatially, such as room acoustic analysis [1], [2], virtual- and augmented-reality systems [3], [4], sound reproduction [5], sound field equalization [6], echo cancellation [7], and speech dereverberation [8].

Most established methods to estimate the ATF spatially are not appropriate for interpolating its values for continuously variable source or receiver position points assuming proper restrictions. As it is a transfer function, one method is to determine its poles and/or zeros, similarly to a standard transfer function of a linear system [9], [10]. However, these models are not properly applicable to the spatial interpolation of the ATF as the relation between the filter coefficients and the source positions is not well understood. Another approach is to interpolate the IR by considering the acoustic field to be a combination of plane waves and then using compressed sensing [11]. In [12], a method of measuring IRs based on a time domain formulation can be used to estimate the IR with high accuracy by utilizing equivalent sources. However, assumptions about the ATF are imposed in [11] and [12] that are not sufficiently general. Moreover, neither method accounts for variable source positions.

The ATF problem is strongly related to understanding sound wave propagation. Once converted to the frequency domain, the function describing the pressure changes caused by a sound wave traveling through the air, now referred to as the sound field, satisfies the Helmholtz equation on the spatial parameters. In the spherical coordinate system, the Helmholtz equation’s spherical wave functions can be determined regardless of boundary conditions [13], meaning that the acoustic characteristics of a room can be expressed as constant coefficients of a series expansion of spherical wave functions of the Helmholtz differential operator. In [14], the sound field is estimated by an empirical truncation of the spherical wave function expansion, allowing for spatial interpolation of sound fields.

In a previous work, Samarasinghe et al. proposed a physical model of the ATF capable of representing it continuously for variable source and receiver positions [15]. This method

Manuscript received 7 March 2022; revised 20 June 2022 and 11 August 2022; accepted 14 August 2022. Date of publication 24 August 2022; date of current version 16 September 2022. This work was supported in part by JSPS KAKENHI under Grant JP19H01116, in part by JST PRESTO under Grant JPMJPR18J4, and in part by JST FOREST Program under Grant JPMJFR216M, Japan. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stefania Cecchi. (Corresponding author: Juliano G. C. Ribeiro.)

Juliano G. C. Ribeiro, Shoichi Koyama, and Hiroshi Saruwatari are with the Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan (e-mail: ribeiro-juliano@g.ecc.u-tokyo.ac.jp; koyama.shoichi@ieee.org; hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp).

Natsuki Ueno was with the University of Tokyo, Tokyo 192-0397, Japan. He is now with the Tokyo Metropolitan University, Tokyo 192-0397, Japan (e-mail: natsuki.ueno@ieee.org).

Digital Object Identifier 10.1109/TASLP.2022.3201368

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
separates the ATF into a direct component, which addresses the direct incidence from the source to the receiver, and a reverberant component, which represents the sound field resulting from interactions with the environment. By considering the direct component to be known and approximating the reverberant component using a finite sum of spherical wave functions, this method can be used to estimate the ATF between variable sources and receivers within regions. However, this model lacks essential properties of the ATF such as reciprocity [16], a property commonly observed in a variety of situations [13], [17]. As the ATF is reciprocal, the enforcement of this essential property is expected to represent real ATF functions. The method is also dependent on reference positions of each region, which are not referenced in the Helmholtz equation at all and introduce another degree of uncertainty.

We propose a method for interpolating ATFs between two regions based on kernel ridge regression (KRR). Instead of truncating the function series, we utilize an infinite dimensional representation of ATFs. A similar approach was used in the sound field estimation problem in [18], [19]. To apply KRR, we define a reproducing kernel Hilbert space (RKHS) that guarantees any estimation holds the physical properties of ATFs. The features of our proposed method are as follows: 1) the reverberant component of the estimated ATF satisfies the homogeneous Helmholtz equation, 2) the estimation does not set a truncation order or an expansion reference position as is required in [15], and 3) the reciprocity of the ATF is taken into consideration.

Preliminary results have been presented in [20]. However, in this paper, we reformulated the RKHS based on the plane wave expansion to make the analysis more straightforward. This formulation is also more representative of the physics of the problem and can be used to expand the concept using directional weighting [21]. In addition, we perform experimental validation with noisy samples, absent from [15], [20], and introduce kernel principal component analysis (KPCA) to increase the robustness against noise.

The remainder of this paper is organized as follows. In Section II, we explain the problem statement and how the interpolation model will be derived. This is followed by Section III, where we explain the properties the ATF is supposed to hold, then define the RKHS model so that the expected properties are always satisfied. Subsequently, in Section IV, we explain the derivation of the kernel ridge regression and kernel principal component analysis procedures, which together compose our proposed method. Then, in Section V, we introduce the truncation-based formulation from [15], which will be compared with our method. This is followed by Section VI, where the simulator conditions are explained, and then the simulation results are discussed. Finally, we conclude in Section VII.

A. Notations

The set of natural numbers, which for our purposes includes 0, is \( \mathbb{N} \). The set of real numbers is \( \mathbb{R} \), and the set of complex numbers is \( \mathbb{C} \). The set of nonnegative real numbers is \( \mathbb{R}_+ \). The application of \( n \) consecutive cartesian products of a set \( A \) will be \( A \times A \times \cdots \times A = A^n \). Italic letters denote scalars and scalar functions, lower-case boldface letters denote vectors, and upper-case boldface letters denote matrices. A vector \( \mathbf{x} \in \mathbb{C}^n \) will have its elements listed as italic letters with indexes, i.e. \( \mathbf{x} = [x_1 \ x_2 \ \ldots \ x_n]^T \). For letter constants, \( i \) is the imaginary unit, \( I \) is the identity matrix, \( e = \exp(1) \) is Euler’s number, and \( 0 \) is the null vector. When the argument of a function is suppressed, we will replace it with a centered dot \( . \). We assume harmonic time dependence \( \exp(-i\omega t) \) with time \( t \) and angular frequency \( \omega \).

All integrals used will be considered to be Lebesgue integrals. The space of square-integrable functions on the Lebesgue integral defined on the domain \( D \) will be referred to as \( L^2(D) \). Functions that only differ in a null measure set will be treated as the same function, as is conventional.

The source and receiver position pairs will be written as \( r/s \), where \( r \) is the position of the receiver and \( s \) the position of the source. For notational simplicity, the letter \( q \) will also be used to refer to a given sound/receiver pair interchangeably whenever convenient to do so.

II. Problem Statement

Let us denote the space where the measurements will be taken as \( \Omega \subset \mathbb{R}^3 \) with stationary acoustic properties. Consider now a pair of positions \((s, r) \in \Omega^2 \) containing a loudspeaker (acting as a monopole source) and a pressure microphone (acting as an omnidirectional receiver). Therefore, for a given wavenumber \( k \), we define the ATM \( h \), \( h : \Omega^2 \times \mathbb{R} \to \mathbb{C} \) as the function that relates the signals between two points. The wavenumber \( k \) is calculated as \( k = 2\pi f/c \), where \( f \) is the frequency and \( c \) is the speed of sound. Henceforth, \( k \) will be omitted from argument lists for notational simplicity, as the calculations are separated by frequency.

Our objective is to interpolate the ATM in a region-to-region manner, that is, to establish a method of interpolating ATM values freely within assigned boundaries. The ATM problem is closely related to the sound field estimation problem. As such, we represent \( h \) as a solution to the inhomogeneous Helmholtz equation with the point source at \( s \) as

\[
(\nabla_r^2 + k^2)h = -\delta(r/s),
\]

where \( \nabla_r^2 \) is the Laplacian operator applied only to the receiver coordinates, and \( \delta(r/s) \) is the Dirac delta function centered at the source point \( s \) and evaluated at the position \( r \). The solution for \( h \) is represented by the sum of particular and homogeneous solutions [13], which can be regarded as the direct component \( h_D \) \( : \Omega^2 \times \mathbb{R} \to \mathbb{C} \) and reverberant component \( h_R \) \( : \Omega^2 \times \mathbb{R} \to \mathbb{C} \), respectively. Thus,

\[
h = h_D + h_R.
\]

This assumption is also used in [15], [20]. The proper representation of sound fields is dependent on the size of the analyzed region and the number of measurement points [14], [16]. Therefore, it is not viable to define an ATF interpolation model that works for the entire environment, but rather between regions within it. We identify two regions: the source region \( \Omega_S \) and the receiver region \( \Omega_R \) as illustrated in Fig. 1.
In [15], the direct component \( h_D \) is assumed to result from the direct incidence of the source signal on the receiver point without encountering obstacles. That is, \( h_D \) is equivalent to a sound field propagated in the free-field, and as such is given by the three-dimensional Green’s function associated with the Helmholtz equation \( G_0 \), as shown in [13].

\[
h_D(r|s) = G_0(r|s) = \frac{e^{ik||r-s||}}{4\pi||r-s||},
\]

(3)

where \( || \cdot || \) is the Euclidean norm. Since the direct component is considered to be known, we aim to approximate the value of \( h_R(r|s) \) \( \forall (r,s) \in \Omega_R \times \Omega_S \) based on multiple ATF measurements. As such, after determining the value of \( h_D \), this component is extracted and only \( h_R \) will be interpolated.

To measure the ATF, we distribute a total of \( L \) loudspeakers at the positions \( \{s_i\}_{i=1}^{L} \) in \( \Omega_S \) and \( M \) microphones at the positions \( \{r_m\}_{m=1}^{M} \) in \( \Omega_R \), and then we measure all \( N = LM \) ATF values. Similarly to the setup described in [15], [20], the ATF values should be measured by taking a source and a receiver devices and determining their frequency response, then using that frequency response to estimate the required ATF values by varying their positions one by one.

The input data vectors \( \{q_n\}_{n=1}^{N} \), referring to the collection of source/receiver pairs measured, the error vector \( e \), and measured output data vector \( y \) will be given as

\[
q_{m+(i-1)M} = r_m|s_i,
\]

(4)

\[
h_R = \begin{bmatrix} h_R(q_1) \ldots h_R(q_N) \end{bmatrix}^T,
\]

(5)

\[
y = h_R + e,
\]

(6)

where \( e \) represents the probabilistic added noise, a random variable with normal distribution \( \mathcal{N}(0, \sigma^2) \), that is, a Gaussian distribution with mean 0 and variance \( \sigma^2 \), which determines the signal-to-noise ratio (SNR). The interpolation function of the reverberant component \( h_R \) will be derived by solving the following optimization problem:

\[
\mathcal{J}(g) = \sum_{n=1}^{N} \left| y_n - g(q_n) \right|^2 + \lambda \|g\|_{\mathcal{H}}^2,
\]

(7)

where \( g \in \mathcal{H} \) is a generic function variable, \( \| \cdot \|_{\mathcal{H}} \) is the norm of the feature space \( \mathcal{H} \) and \( \lambda \in \mathbb{R}_+ \) is the regularization constant. This optimization problem has different solutions depending on the configuration of \( \mathcal{H} \). Therefore, to derive the interpolation function, it is necessary to define the feature space \( \mathcal{H} \) so that it permits a unique minimizer for \( \mathcal{J} \) and holds the relevant acoustic properties of the reverberant component of the ATF.

Having determined the interpolation function for the reverberant component, we can then add back the direct component and obtain the interpolation function for the ATF \( h \) as

\[
h = h_D + h_R.
\]

(9)

III. PHYSICAL FORMULATION OF ATF

The ATF between the points \( r \in \Omega_R \) and \( s \in \Omega_S \) is the signal received at \( r \) when the signal originating from \( s \) is the Dirac delta function in the time domain. It is necessary to consider real world effects that change the signal, such as the reverberation within the enclosure of the environment, making the ATF unpredictable in the general case. In order for our interpolation method to fulfill the necessary properties of the ATF, we must properly define the feature space \( \mathcal{H} \).

A. Basic Properties of ATF

We now explain the observed properties of the ATF that will be included in our interpolation model explicitly. Part of our objective is to derive a feature space for the reverberant component, as outlined in Section II. Therefore, we will discuss how these properties impact the characterization of the reverberant component given the known properties of the direct component. The properties are as follows:

1) The reverberant component of the ATF is a solution of the homogeneous Helmholtz equation on the receiver variable.

2) As the Helmholtz operator is invariant to translations and rotations of the coordinate system, the estimation should be invariant to rotations and changes of origin positions.

3) The ATF is reciprocal, meaning that the ATF measured between a source and a receiver point is the same regardless of which is considered the source and which the receiver.

The reverberant component is caused by reflections, diffraction, and any other physical phenomena that do not result from a direct incidence from a source to a receiver. Therefore, it is equivalent to a sound field absent of a source, and as such, the reverberant component must satisfy the homogeneous Helmholtz equation on the receiver variable [13], whatever the source point might be, addressing item 1).

\[
(\nabla_r^2 + k^2)h_R(r|s) = 0, \; \forall r \in \Omega_R \text{ and } \forall s \in \Omega_S.
\]

(10)

Next, we address item 2), that is the invariance of the derivation of the model to rotations and translations. Since the direct component is known and only considers the Euclidean distance between the relative positions of a source and a receiver, it satisfies the second property. Therefore, the reverberant component...
expression must be invariant to rotations and translations of the coordinate system to guarantee the interpolation function will be.

Item 3), which will be referred to henceforth as the reciprocity, signifies that the ATF originated at a source point \( s \) and recorded at a receiver point \( r \) is the same as the ATF measured at \( s \) had the signal originated from \( r \) [13], [16], [17].

\[
h(r|s, k) = h(s|r, k), \quad \forall r, s, k.
\] (11)

Since \( h \) and \( h_D \) are reciprocal, \( h_D \) must be reciprocal as well. By combining (10) with the reciprocity, we have that the Helmholtz equation must also be satisfied on the source variable.

\[
(\nabla^2_r + k^2)h_D(r|s) = (\nabla^2_s + k^2)h_D(r|s) = 0,
\] (12)

where \( \nabla^2_s \) is the Laplacian differential operator applied only to the source position coordinates.

Owing to the differential property, we know that \( h_D \) must be bounded, making it appropriate for interpolation. We will utilize this characterization of the reverberant component to define our feature space properly, as well as devise the KRR and KPCA formulations and explain their merits over the estimation used in [15], which will be used for comparison.

B. RKHS Model for Reverberant Components

A RKHS is a specific class of functional Hilbert spaces that contain a reproducing kernel function. This means that, for a Hilbert space represented by the function set \( H \) and the inner product \( \langle \cdot, \cdot \rangle \) of functions defined on a domain \( D \), there is some kernel function \( \kappa : D^2 \rightarrow \mathbb{C} \) that projects elements of \( H \) on the domain points.

\[
\langle \kappa(\cdot, x), g \rangle = g(x), \quad \forall x \in D \quad \text{and} \quad \forall g \in H.
\] (13)

In [20], we constructed a RKHS formulation based on a series expansion of wave functions, defining the feature space \( H \) and the inner product based on the infinite series. In this paper, however, we will define the RKHS here using an equivalent of a Herglotz wave function [22] for the ease of applicability in RKHS problems [23]. We define the plane wave expansion for the reverberant part of the ATF as

\[
g(r|s) = \int_{\mathbb{S}^2 \times \mathbb{S}^2} e^{ik(u-r+v \cdot s)} G(u, v) \text{dudv}, \quad \text{for} \quad r \in \mathbb{S}^2 \quad \text{and} \quad s \in \mathbb{S}^2.
\] (14)

The space given in (15) is always complete on the induced norm of the accompanying inner product as the space of reciprocal \( L^2 \) functions is a closed set, and the transformation between both sets preserves the proportionality of the norms:

\[
\|g\|_H^2 = \langle g, g \rangle = 16\pi^2 \int_{\mathbb{S}^2 \times \mathbb{S}^2} |G(u, v)|^2 \text{dudv} \propto \|g\|_{L^2}^2, \quad \text{for} \quad g \in H.
\] (17)

where \( \| \cdot \|_{L^2} \) is the norm of the integral space \( L^2 \).

Since the norms are proportional, Cauchy sequences in \( H \) correspond directly to reciprocal Cauchy sequences in the \( L^2 \) space, inheriting their convergence. We expect an approximation based on elements of \( H \) to hold all relevant acoustic properties. This Hilbert space admits the following reproducing kernel function:

\[
\kappa(r|s, r'|s') = \int_{\mathbb{S}^2 \times \mathbb{S}^2} e^{ik(u-r+v \cdot s)} + e^{ik(u-r'+v \cdot s')} \text{dudv} - \frac{1}{2} \left( j_0(k||r-r'||_2)j_0(k||s-s'||_2) + j_0(k||r-s'||_2)j_0(k||r-s'||_2) \right),
\] (18)

where \( j_0 \) is the spherical Bessel function of order 0. The computation of the integral in (18) can be performed by combining the Rayleigh equation with the orthogonality of the spherical harmonic functions [24], as expanded upon in the appendix. We can guarantee that \( \kappa \) is a reproducing kernel function by applying the inner product and checking if it holds the projection property for a generic function \( g \) and generic source/receiver position points \( s \) and \( r \), respectively:

\[
\langle \kappa(\cdot, r|s), g \rangle = 16\pi^2 \int_{\mathbb{S}^2 \times \mathbb{S}^2} \frac{e^{ik(u-r+v \cdot s)} + e^{ik(u+s+v \cdot r)}}{2} G(u, v) \text{dudv}.
\] (19)

We can then simplify the above expression as shown in (14).

\[
\langle \kappa(\cdot, r|s), g \rangle = \frac{1}{2} \langle g(r|s) + g(s|r) \rangle.
\] (20)

Since elements of \( H \) are always guaranteed to be reciprocal, \( g(r|s) = g(s|r) \). Thus,

\[
\langle \kappa(\cdot, r|s), g \rangle = g(r|s), \quad \forall r|s \in \Omega^2 \quad \text{and} \quad g \in H.
\] (21)

We have shown that \( \kappa \) is indeed the reproducing kernel of the Hilbert space, making \( H \) as constructed a reproducing kernel Hilbert space. This reproducing kernel satisfies the homogeneous Helmholtz equation for any source/variable pair, it is invariant in regards to rotation and translation of the coordinate system, and is reciprocal in regards to the first source/receiver pair input.
IV. INTERPOLATION FUNCTION OF REVERBERANT COMPONENT

RKHSs are widely applied in physics and machine learning applications [23], [25] because of their adaptability and the existence of the representer theorem [26], which guarantees solutions of minimization problems such as (8) with solutions expressed as linear expansions of kernel functions anchored on the domain points of the data set [27]. In other words, the solution to our optimization problem is given as

\[
\hat{h}_R(r|s) = \sum_{n=1}^{N} \alpha_n \kappa(r|s, q_n).
\]  

(22)

Therefore, optimizing the defined loss is equivalent to finding the complex coefficient vector \( \alpha = [\alpha_1 \alpha_2 \ldots \alpha_N]^T \in \mathbb{C}^N \) for the interpolation function of the reverberant component \( \hat{h}_R \). We will adopt the alternative notation below for ease of operation:

\[
\hat{h}_R(r|s) = \kappa(r|s)\alpha,
\]  

(23)

where the function vector \( \kappa \) is

\[
\kappa(r|s) = [\kappa(r|s, q_1) \kappa(r|s, q_2) \ldots \kappa(r|s, q_N)].
\]  

(24)

Therefore the cost \( J \) and the optimization problem (8) can be rewritten as

\[
J(\alpha) = \|K\alpha - y\|^2 + \lambda \alpha^H K\alpha,
\]  

(25)

\[
\hat{h}_R(r|s) = \kappa(r|s) \left( \arg\min_{\alpha \in \mathbb{C}^N} J(\alpha) \right),
\]  

(26)

where \( K \) is the Gram matrix given as

\[
K = \begin{bmatrix}
\kappa(q_1, q_1) & \ldots & \kappa(q_1, q_N) \\
\vdots & \ddots & \vdots \\
\kappa(q_N, q_1) & \ldots & \kappa(q_N, q_N)
\end{bmatrix}.
\]  

(27)

This problem can be solved by minimization on \( \mathbb{C}^N \). By taking the gradient of \( J \) in regards to \( \bar{\alpha} \), we can see that

\[
\frac{\partial J}{\partial \bar{\alpha}} = (K^2 + \lambda K)\alpha - Ky,
\]  

(28)

and the Hermitian matrix of \( J \) is given by

\[
H(J) = K^2 + \lambda K.
\]  

(29)

Since \( K \) is a Gram matrix, it is positive-definite, meaning that \( H(J) \) is also positive-definite. Therefore, any critical points of \( J \) will be minimum points. To find those minimum points, we find the point with a null gradient:

\[
\frac{\partial J}{\partial \bar{\alpha}} = (K^2 + \lambda K)\alpha - Ky = 0
\]  

\[
\Rightarrow (K^2 + \lambda K)\alpha = Ky.
\]  

(30)

Since \( K \) is positive-definite, it is also nonsingular. Thus, the coefficient vector can be computed.

\[
\alpha = (K + \lambda I)^{-1}y.
\]  

(31)

This means \( J \) only has one minimizer, meaning it is a global minimizer as well. Therefore, the reverberant component of the interpolation function that optimizes our problem is given by

\[
\hat{h}_R(r|s) = \kappa(r|s)\alpha = \kappa(r|s)(K + \lambda I)^{-1}y.
\]  

(32)

This estimation is referred to as kernel ridge regression [27]. As the interpolation function is a linear combination of kernel functions, it holds the pertinent acoustic properties we embedded when creating the RKHS. Finally, the interpolation function is given by adding back the direct component to the estimations.

A. Application of KPCA for Increasing Noise Robustness

The interpolation function \( \hat{h} \) necessarily satisfies the fundamental properties of the ATF. The addition of noise into the problem is sure to cause performance degradation, and as such, some form of criterion that diminishes its effect was also applied. We opted for the KPCA, an established and widely-used method of dimensionality reduction [28] which requires the data to be projected into a higher dimensional RKHS. As the ATF has already been characterized in terms of a RKHS, it is natural to consider the use of the KPCA for the ATF interpolation problem.

Another property that makes the definition of the RKHS attractive for approximation problems is the characterization of the data and its features. The KPCA is a nonlinear extension of the principal component analysis method, a technique of dimensionality reduction and feature extraction that involves capturing the essential properties of a data set on the basis of an extended notion of the covariance function applied to each pair of samples when the feature space they are in is a RKHS. The KPCA requires the data to be mean-centered, or centered, so the first step in applying it is to make the mean of our data set 0:

\[
x = (I - 1_N)y,
\]  

(33)

where \( x \) is our centered data and \( 1_N \in \mathbb{C}^{N \times N} \) is defined as

\[
1_N = \frac{1}{N} \begin{bmatrix}
1 & \ldots & 1 \\
\vdots & \ddots & \vdots \\
1 & \ldots & 1
\end{bmatrix}.
\]  

(34)

After our data have been centered, we calculate the covariance matrix of the data, defined as the matrix of the inner products between every pair of elements of \( x \).

\[
\text{Cov}(x) = \begin{bmatrix}
x_1 \cdot x_1 & x_1 \cdot x_2 & \ldots & x_1 \cdot x_N \\
\vdots & \vdots & \ddots & \vdots \\
x_N \cdot x_1 & x_N \cdot x_2 & \ldots & x_N \cdot x_N
\end{bmatrix}.
\]  

(35)

For the standard principal component analysis, the product used is the · product between vector samples of data in \( \mathbb{R}^d \) spaces, \( d \in \mathbb{N} \). This is somewhat limited when the dimension of the vector space is much smaller than the number of data samples. For the kernel principal component analysis, however, we assume that the measurement matrix \( y \) is actually made of elements of a RKHS. Thus, instead of the usual Euclidean inner product, the inner product assumed here is given by the kernel function [29]. If the measurement vector \( y \) was already centered, the covariance matrix is straightforwardly the Gram matrix \( K \). If the data were not centered, the matrix will be given by the
centered Gram matrix $K'$:

$$\text{Cov}(x) = K' = (I - 1_N)K(I - 1_N). \quad (36)$$

From here, the KPCA and the standard principal component analysis behave identically. We first apply the singular value decomposition (SVD). Since $K$ is a Gram matrix, it is positive-definite and symmetrical, meaning that it has a complete eigenspace, all of its singular values are positive, and the SVD is equivalent to its diagonalization. The same can be said of $K'$ except that the singular values are only expected to be nonnegative. Therefore, there is an orthogonal matrix $V$ and a diagonal matrix $\Sigma$ such that

$$K' = V\Sigma V^H. \quad (37)$$

The nonzero elements of $\Sigma$ form a vector $\sigma$, where they are arranged in a descending order. Each singular value $\sigma_n$ is related to a column of $V$. For a given threshold $\sigma_{th}$, to apply the KPCA on our data, we must determine $n' \leq N$ such that $\sigma_{n'} > \sigma_{th}$ and $\sigma_{n'+1} \leq \sigma_{th}$. In doing so, we build the new lower-dimensional space $V_{\text{red}}$ into which the data will be projected, acting as a noise filter. The columns of $V_{\text{red}}$ correspond to the first $n'$ eigenvectors.

Our filtered signal $y'$ is given as

$$y' = V_{\text{red}}^H(I - 1_N)y + 1_Ny. \quad (38)$$

Therefore, the average has to be returned to the data post-processing. Indeed, the choice of a filter carries a considerable risk: if the data are not consistent with the devised physical model, the choice to discard the eigenspace related to the eigenvalues ordered higher than $n'$ would mean eliminating an arbitrary portion of the data that may or may not be representative of the whole. Conversely, if the method does work, it shows the power of the physical model given to the ATF by showing the correlation between the data and the expected result.

Currently, the KRR already employs regularization, another form of coping with noisy data sets. Regularization has the effect of shrinking the influence lower-correlation components have on the estimation, while using the KPCA to project the data completely eliminates their contribution to the data. These components tend to be relatively noisier compared to the more highly correlated vectors associated with higher principal values. Therefore, if the noise level is significant, we expect the KPCA to be more effective in filtering noise than the KRR, while for clean signals, their performance should be similar.

V. COMPARISON WITH TRUNCATION-BASED ESTIMATION

We now explain the previously established method that will be used as a standard of comparison. To estimate the acoustic transfer function, that is, to constrain the method by upholding the physical characteristics of the ATF, we must analyze forms of representing such functions. In [15], an interpolation method that only relies on properties of sound fields and does not rely on knowledge of the geometrical or acoustic properties of the environment was proposed.

By expressing the coordinate source and receiver points in spherical coordinates, one can determine the solution to the sound field problem by solving the homogeneous Helmholtz equation, which is a series expansion of spherical wave functions. The positions $r$ and $s$ will be

$$r = \begin{bmatrix} r \cos(\phi_r) \sin(\theta_r) \\ r \sin(\phi_r) \sin(\theta_r) \\ r \cos(\theta_r) \end{bmatrix} + r_c, \quad (39)$$

$$s = \begin{bmatrix} s \cos(\phi_s) \sin(\theta_s) \\ s \sin(\phi_s) \sin(\theta_s) \\ s \cos(\theta_s) \end{bmatrix} + s_c, \quad (40)$$

where $r_c$ and $s_c$ are chosen reference positions in $\Omega_R$ and $\Omega_S$. $r$ and $s$ are the respective radii, $\theta_r$ and $\phi_r$ the zenith angles and $\phi_r$, and $\phi_s$ the azimuth angles. The wave functions are products of wave functions for the single variable case.

Since the reverberant component satisfies the Helmholtz equation on both variables, it is possible to express it as a series expansion on coefficients $\gamma_{\nu,\mu}$, which are functions of the reference positions:

$$h_R(r|s) = \sum_{\nu \in \mathbb{N}} \sum_{\nu' \in \mathbb{N}} \sum_{\mu' = -\nu}^{\nu'} \gamma_{\nu,\mu'}(r_c, s_c) \varphi_{\nu'}(s) \varphi_{\nu}(r), \quad (41)$$

where $\varphi_{\nu}(r) = j_{\nu}(k_r r) Y_{\nu}^{\mu}(\phi_r, \theta_r)$ is the spherical wave function of order $\nu$ and degree $\mu$, $j_{\nu}$ is the spherical Bessel function of order $\nu$, and $Y_{\nu}^{\mu}$ is the spherical harmonic function of order $\nu$ and degree $\mu$. In [20], we have shown this infinite modal expansion can be expressed in terms of a RKHS with the same kernel function.

In [15], the reverberant component is given by truncating (41) at two separate empirical orders $N_R$ and $N_S$. These orders are dependent on the frequency and the size of the region. The relation between the number of orders $\nu_{\text{ord}}$, the wavenumber $k$ and the radius of the region $R$ is given in [16] as

$$\nu_{\text{ord}} = \frac{keR}{2}, \quad (42)$$

Therefore, the interpolation function is given by estimating the coefficients up to finite orders:

$$\hat{h}_R(r|s) = \sum_{\nu=0}^{N_R} \sum_{\nu' = -\nu}^{\nu} \sum_{\mu' = -\nu'}^{\nu'} \gamma_{\nu,\mu'}(r_c, s_c) \varphi_{\nu'}(s) \varphi_{\nu}(r). \quad (43)$$

This estimation will be referred to as the truncation (TR) hereafter. The method of coefficient extraction is described in detail in [15]. The environment is entirely characterized by the calculated coefficients, meaning that knowledge of the shape and physical properties of the room is not necessary. This method also satisfied the Helmholtz equation on both source and receiver coordinates, as the ATF is expected to. For our application, these components were estimated sequentially using a regularized least-squares criterion in each region [14].

However, this model still has issues that should be addressed. The first is the fact the estimation is dependent on defining centers and orientation for the coordinate system. The Helmholtz operator is invariant in regards to rotations and translations of the
coordinate system, however these factors do change the derived truncated expansion. This means the centers of the estimation and the orientation of the coordinate systems could also be seen as hyperparameters of the model, with unpredictable effects. Another issue is that the process is divided in stages, meaning error from the first estimation accumulates with the second estimation. The reciprocity is also not enforced in any way.

VI. NUMERICAL EXPERIMENTS

The proposed methodology was experimented on using the image source method [30]. The experimental conditions were chosen for a projected reverberation time of $T_{60} = 0.45$ s. For that end, we adopted a rectangular room with dimensions of $3.2 \times 4.0 \times 2.7$ m in the $x, y,$ and $z$ directions respectively. The speed of sound was $c = 340$ m/s.

The proposed and truncation-based methods were compared in two different experimental settings: when the regions are separate as shown in Fig. 2, and when they overlap as shown in Fig. 3. In both cases, the regions were represented by spheres of 0.2 m radius with centers $r_0$ for $\Omega_R$ and $s_0$ for $\Omega_S$. In the case where the regions are separate ($\Omega_R \cap \Omega_S = \emptyset$), the chosen centers were $r_0 = [-0.35, -0.43, -0.29]^T$ m and $s_0 = [0.35, 0.43, 0.29]^T$ m. For the case where the regions overlap ($\Omega_R \cap \Omega_S \neq \emptyset$), the chosen centers were $r_0 = [-0.10, -0.12, -0.08]^T$ m and $s_0 = [0.10, 0.12, 0.08]^T$ m. In either case, the regions were assumed to be spherical with a radius of 0.20 m.

Both the source and receiver arrays used were concentric double-layer spherical shell arrays. The distribution of measurement points in either layer was chosen with the $t$-design criterion of evaluation [31], [32], as they achieve near uniform sampling [33]. The number of points in a $t$-design rig is given by a chosen $t \in \mathbb{N}$ as $(t + 1)^2$. The arrays in the source and receiver regions always had the same number of points. This model was adopted due to the fact the TR relies on two consecutive applications of sound field analysis. If the number of points in the source or receiver region is insufficient to represent sound fields for a specific frequency, the estimation is unreliable.

There was one virtual array setting utilized in the experiments: a concentric dual shell array with an internal radius of 0.19 m and an outer radius of 0.20 m centered at the center of the regions. These arrays had a total of $M = L = 32$ points each, corresponding to $t$-design arrays of $t = 3$, giving a total of $N = 1024$ points from which the coefficients for each method are derived from. These test points were only used to derive the model, not to validate it, as doing so would condition the method to the training data.

In order to avoid spatial aliasing for the TR, we avoided having more total modes for the truncation than the total number of measurement points in each region. As the total number of modes up to an order $\nu_{\max}$ is $(\nu_{\max} + 1)^2 \leq 32$, we can determine a maximum of $N_S = N_R = [\sqrt{32} - 1] = 4$ orders. From (42), this means the maximum frequency for which the TR will be considered reliable is $f_{\max} = N_R c / \pi r_{\text{out}} = 796.28$ Hz. We compared both methods up to a maximum frequency of 1000 Hz, beyond this theoretical limit.

A. Evaluation Criteria for Experiments

The proposed methodology for estimation of the ATF in noisy environments was tested by comparing the truncation and the KRR with and without application of the KPCA. We experimented with two different noise levels: 20 dB and 40 dB. The noise used had a complex Gaussian distribution, with mean 0 and variance determined by the SNR. There were four different experimental settings: TR without the KPCA, KRR without the KPCA, TR with the KPCA (TR+KPCA), and KRR with the KPCA (KRR+KPCA). As mentioned previously, the KPCA was applied with a simple filter criterion where the principal values would be the eigenvalues surpassing $\sigma_0 = \epsilon$. The regularization constant $\lambda$ for the KRR was given by the SNR as $10^{-\frac{\text{SNR}}{10}}$, under the assumption of a 0 mean complex Gaussian

![Fig. 2. Experimental setup where the regions are separate. Red crosses represent sources, blue circles represent receivers.](image1)

![Fig. 3. Experimental setup where the regions overlap. Red crosses represent sources, blue circles represent receivers.](image2)
distribution for the noise and a Gaussian prior with correlation matrix given by the Gram matrix for the KRR, which has been shown to be a reasonable model for sound fields in [34]. For the truncation, four hyperparameters were considered: the centers of the coordinate systems and the hyperparameters associated with each region. These hyperparameters were optimized using a standard optimization tool [35] and a leave-one-out cross validation (LOO-CV) criterion with the quadratic error, here adopted for its unbiased nature and closed form solution [36].

The first comparison criterion is an evaluation of the frequency-domain performance of each method, expressed by the normalized mean square error (NMSE) defined in (44). The experimental results for this test can be seen in Fig. 4.

\[
\text{NMSE}(f) = 10 \log_{10} \left( \frac{\sum_{n=1}^{N'} |h(q_{n}', k) - \hat{h}(q_{n}', k)|^2}{\sum_{n=1}^{N'} |h(q_{n}', k)|^2} \right),
\]

(44)

where \(N' = 9025\) is the number of test points and the pairs \(\{q_{n}'\}_{n=1}^{N'}\) represent the collection of all the test source/receiver points. The test points represent all possible pairs between 95 source points and 95 receiver points picked for model validation, both randomly chosen with a uniform distribution within their respective regions. This validation set is completely separate from the 1024 pairs used to derive the interpolation functions, and none of the sources or receivers in the validation set coincides with any of the source or receivers in the virtual arrays. The frequencies evaluated were 50 sampled frequencies between 0 Hz and 1000 Hz without including either end.

We also evaluated the performance of each method in performing pointwise estimations with the visualized reconstructed pressure distribution from a single source located at \(s_0\) on a regularly spaced \(0.72\) m × \(0.72\) m square grid centered in \(r_0\), for a total of 10201 points and analyze the capability of the method to reproduce that sound field for a frequency of 804 Hz. Since this frequency exceeds \(f_{\text{max}}\), the TR pressure distribution was only shown for the cleaner 40 dB SNR case, with application of the KPCA. The KRR estimations are still shown with both levels of noise to better observe how well the application of the KPCA improved the noisier sample. We also analyze the pointwise error of the estimations with the normalized error (NE) defined in (45).

\[
\text{NE}(r|s) = 10 \log_{10} \left( \frac{|h(r|s) - \hat{h}(r|s)|^2}{|h(r|s)|^2} \right).
\]

(45)
B. Experimental Results

As shown in Fig. 4, the KRR outperformed the TR for every frequency, for both region configurations and the KPCA was successful in mitigating the effects of noise in the samples. Special attention should be given to the 20 dB SNR case, where the TR diverged radically without the KPCA, but was actually able to perform better with it. Particularly for higher frequencies, when the models start to lose interpolation power, the KPCA showed itself to be effective. We conclude that the derived RKHS is an accurate representation of the analyzed data and the KRR is both more apt for lower frequency interpolations and is valid for a higher maximum frequency.

We also evaluated the pointwise interpolation performance of the methods. The attempted reconstructions can be seen in Figs. 5 and 6, and the corresponding pointwise errors are shown in Figs. 7 and 8. The KRR is compared with the best case scenario of the TR, as indicated by the NMSE results: TR+KPCA with 40 dB SNR. In both configurations, the KRR was shown to perform better. After the application of the KPCA, the KRR of the 20 dB SNR case shows notably better performance than the TR selected. This shows that the KRR not only achieves lower mean error, but also the reconstructed field diverges less from the expected on a pointwise basis as well.

Since the chosen method of feature extraction uses a filter, the successful application of the KPCA further validates the KRR interpolation: even without prior knowledge of the signal itself, projecting the data into the principal vectors resulted in a cleaner sample, showing that the method was indeed mostly filtering out noise and that the enforcement of reciprocity enforces behavior expected to be observed in the data.

Fig. 5. Reconstructed ATF distribution (real part) when $\Omega_R \cap \Omega_S = \emptyset$ for a frequency of 804 Hz. The black circle represents the bounds of $\Omega_R$. (a) Original; (b) KRR for 20 dB SNR; (c) KRR for 40 dB SNR; (d) KRR+KPCA for 20 dB SNR; and (e) TR+KPCA for 40 dB SNR.

Fig. 6. Reconstructed ATF distribution (real part) when $\Omega_R \cap \Omega_S \neq \emptyset$ for a frequency of 804 Hz. (a) Original; (b) KRR for 20 dB SNR; (c) KRR for 40 dB SNR; (d) KRR+KPCA for 20 dB SNR; and (e) TR+KPCA for 40 dB SNR.

Fig. 7. Normalized error in dB when $\Omega_R \cap \Omega_S = \emptyset$ for a frequency of 804 Hz. (a) KRR for 20 dB SNR; (b) KRR for 40 dB SNR; (c) KRR+KPCA for 20 dB SNR; (d) TR+KPCA for 40 dB SNR.
We infer that the ATF function being approximated is thus an element of the RKHS, or at least close to it, which means the derivations presented in this work are effective and plausible representations of the ATF functions analyzed. Consequently, the enforcement of basic properties of the ATF is effective in extracting properties of the data set and provides greater robustness to noisy samples.

VII. CONCLUSION

We have introduced a kernel ridge regression interpolation method for the acoustic transfer function that is invariant to translations and rotations of the coordinate system, admits variable source and receiver positions, enforces reciprocity on its estimations, and is capable of interpolating ATF values for an arbitrary environment without prior knowledge of its physical or acoustic properties. This is a region-to-region method and is only constrained by the physical properties of the ATF. The derived estimation surpassed the previously established truncation-based method for every frequency. The application of the KPCA to increase noise robustness improves results further, even for the truncation method. The KRR when applied in conjunction to increase noise robustness improves results further, even for a frequency of 804 Hz. The KPCA representation of the ATF functions analyzed. Consequently, we want to compute the following integral:

$$\int_{S^2} e^{ikr \cdot v} \, dv$$

$$= 4\pi \int_{S^2} \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} j_{\nu}(k \| r \|) Y^\nu_{\mu}(\phi_r, \theta_r) Y^\mu_{\nu}(\phi_v, \theta_v) \, dv$$

$$= 4\pi \int_{S^2} \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} j_{\nu}(k \| r \|) Y^\nu_{\mu}(\phi_r, \theta_r) \int_{S^2} Y^\mu_{\nu}(\phi_v, \theta_v) \, dv. \quad (47)$$

We also know that the spherical harmonic functions are orthogonal in regards to the above integral [24], and that $$Y^\nu_{0}(\phi_v, \theta_v) = 1/\sqrt{4\pi}$$. Thus:

$$\int_{S^2} Y^\nu_{0}(\phi_v, \theta_v) Y^\mu_{\nu}(\phi_v, \theta_v) \, dv = \begin{cases} 1, & \nu = \mu = 0 \\ 0, & \text{otherwise} \end{cases}. \quad (48)$$

By combining both expressions we have that:

$$\int_{S^2} e^{ikr \cdot v} \, dv = 4\pi j_0(k \| v \|), \quad (49)$$

meaning the kernel function can then be derived by reorganizing the following integral:

$$\int_{S^2 \times S^2} e^{ik[u \cdot (r - r') + v \cdot (s - s')]} \, dudv = \int_{S^2} e^{ik(r - r') \cdot u} \, du \int_{S^2} e^{ik(s - s') \cdot v} \, dv$$

$$\quad (50)$$

We can now use (49) to compute the integrals:

$$\int_{S^2 \times S^2} e^{ik[u \cdot (r - r') + v \cdot (s - s')]} \, dudv = 16\pi^2 j_0(k \| r - r' \|) j_0(k \| s - s' \|), \quad (51)$$

and substituting this expression into (18).

REFERENCES

[1] H. Morgenstern, B. Rafaely, and M. Noiernig, “Joint design of spherical microphone and loudspeaker arrays for room acoustic analysis,” in Proc. Annu. German Congr. Acoust. (DAGA), 2015, pp. 475–478.

[2] D. Khaykin and B. Rafaely, “Acoustic analysis by spherical microphone array processing of room impulse responses,” J. Acoustical Soc. Amer., vol. 132, no. 1, pp. 261–270, 2012.

[3] T. Lentz, D. Schröder, M. Vorlaender, and I. Assenmacher, “Virtual reality system with integrated sound field simulation and reproduction,” EURASIP J. Adv. Signal Process., vol. 2007, 2007, Art. no. 187.

[4] S. Lopes, A. Oliveira, J. Vieira, G. Campos, P. Dias, and R. Costa, “Real-time audio augmented reality system for pervasive applications,” in Proc. Audio Eng. Soc. Brazil Conf., 2013, pp. 94–97.

[5] F. Borra, L. D. Gebru, and D. Markovic, “Soundfield reconstruction in reverberant environments using higher-order microphones and impulse response measurements,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2019, pp. 281–285.

[6] R. Mazur, F. Katsberg, and A. Mertins, “Robust room equalization using sparse sound-field reconstruction,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2019, pp. 4230–4234.

[7] P. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. Jensen, “Multimicrophone acoustic echo cancellation using multi-channel warped linear prediction of common acoustical poles,” in Proc. Eur. Signal Process. Conf., 2010, pp. 2121–2125.
[8] N. Mohran, R. Velmurugan, and P. Rao, “Speech dereverberation using NMF with regularized room impulse response,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2017, pp. 4955–4959.

[9] Y. Haneda, Y. Kaneda, and N. Kitawaki, “Common-acoustical-pole and resonance modeling of spatial interpolation and extrapolation of a room transfer function,” IEEE Trans. Speech Audio Process., vol. 7, no. 6, pp. 709–717, Nov. 1999.

[10] 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. 22, no. 1, pp. 205–216, Jan. 2014.

[11] R. Mignot, G. Chardon, and L. Daudet, “Low frequency interpolation of an ARMA model and application to spatial interpolation and extrapolation of a room transfer function,” IEEE Trans. Speech Audio Process., vol. 7, no. 6, pp. 709–717, Nov. 1999.

[12] 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. 22, no. 1, pp. 205–216, Jan. 2014.

[13] E. G. Williams, Fourier Acoustics, London, U.K.: Academic Press, 1999.

[14] M. A. Poletti, “Three-dimensional surround sound systems based on spherical harmonics,” J. Audio Eng. Soc., vol. 53, no. 11, pp. 1004–1025, 2005.

[15] P. N. Samarasinghe, T. D. Abhayapala, M. A. Poletti, and T. Betlehem, “An efficient parameterization of the room transfer function,” IEEE/ACM Trans. Audio Speech Lang. Process., vol. 23, no. 12, pp. 2217–2227, Dec. 2015.

[16] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, “3D spatial soundfield recording over large regions,” in Proc. Int. Workshop Acoust. Signal Enhancement, 2012, pp. 1–4.

[17] J. W. Strutt, “Some general theorems relating to vibrations,” Proc. London Math. Soc., vol. 1, no. 1, pp. 357–368, 1871.

[18] 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, 2018, pp. 436–440.

[19] 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.

[20] J. G. C. Ribeiro, N. Ueno, S. Koyama, and H. Saruwatari, “Kernel interpolation of acoustic transfer function between regions considering reciprocity,” in Proc. IEEE 11th Sensor Array Multichannel Signal Process. Workshop, 2020, pp. 1–5.

[21] J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, “Region-to-region kernel interpolation of acoustic transfer function with directional weighting,” in Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2022, pp. 576–580.

[22] M. Ikehata, “The Herglotz wave function, the Vekua transform and the enclosure method,” Hiroshima Math. J., vol. 35, pp. 485–506, 2005.

[23] D. Colton and P. Monk, “Herglotz wave functions in inverse electromagnetic scattering theory,” in Topics in Computational Wave Propagation: Direct and Inverse Problems, M. Ainsworth, P. Davies, D. B. D. Ryne, and P. Martin, Eds. Berlin, Germany: Springer, 2003, pp. 367–394.

[24] R. Mehrem, “The plane wave expansion, infinite integrals and identities involving spherical bessel functions,” Appl. Math. Comput., vol. 217, no. 12, pp. 5360–5365, 2011.

[25] T. Nguyen, Z. Chen, and J. Lee, “Dataset meta-learning from kernel ridge-regression,” 2020, arXiv:11.00050.

[26] F. Dinuzzo and B. Schölkopf, “The representer theorem for Hilbert spaces: A necessary and sufficient condition,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, vol. 25, pp. 189–196.

[27] K. P. Murphy, Probabilistic Machine Learning. Cambridge, MA, USA: MIT Press, 2022.

[28] S. Y. Kung, Kernel Methods and Machine Learning, Cambridge, U.K.: Cambridge Univ. Press, 2014.

[29] R. Rosipal, M. Girolami, L. Trejo, and A. Cichocki, “Kernel PCA for feature extraction and de-noising in nonlinear regression,” Neural Comput. Appl., vol. 10, pp. 231–243, 2001.

[30] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” J. Acoustical Soc. Amer., vol. 65, no. 4, pp. 943–950, 1979.

[31] X. Chen and R. S. Womersley, “Spherical T-design with $d = (t + 1)^2$ points.” Accessed: Feb. 03, 2022. [Online]. Available: http://www.polyu.edu.hk/ama/staff/jchen/sphdesigns.html

[32] X. Chen and R. S. Womersley, “Existence of solutions to systems of underdetermined equations and spherical designs,” SIAM J. Numer. Anal., vol. 44, no. 6, pp. 2320–2341, 2006.

[33] F. Zotter, M. Frank, and A. Sontacchi, “The virtual T-design ambiotics-rig using VBAP,” in Proc. EAA EuroRegio Cong. Sound Vib., 2010, pp. 1–4.

[34] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunsksog, P. Gerstoft, and E. Fernandez-Grande, “Gaussian processes for sound field reconstruction,” J. Acoustical Soc. Amer., vol. 149, no. 2, pp. 1107–1119, 2021.

[35] P. K. Mogensen and A. N. Riseth, “Optim: A mathematical optimization package for Julia,” J. Open Source Softw., vol. 3, no. 24, 2018, Art. no. 615.

[36] G. A. F. Seber and A. J. Lee, Linear Regression Analysis, 2nd ed. Hoboken, NJ, USA: Wiley, 2003.

Juliano G. C. Ribeiro (Graduate Student Member, IEEE) received the B.E. degree in electric engineering from the Instituto Militar de Engenharia, Rio de Janeiro, Brazil, in 2017, and the M.S. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2021. He is currently working toward the Ph.D. degree in information science and technology from the University of Tokyo. His research interests include spatial audio, acoustic signal processing, and physics informed machine learning.

Natsuki Ueno (Member, IEEE) received the B.E. degree in engineering from Kyoto University, Kyoto, Japan, in 2016, the M.S. and Ph.D. degrees in information science and technology from the University of Tokyo, Tokyo, Japan, in 2018 and 2021, respectively. He is currently a Project Assistant Professor with Tokyo Metropolitan University, Tokyo, Japan. His research interests include spatial audio and acoustic signal processing.

Shoichi Koyama (Member, IEEE) received the B.E., M.S., and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 2007, 2009, and 2014, respectively. He joined Nippon Telegraph and Telephone Corporation in 2009 as a Researcher in acoustic signal processing. In 2014, he moved to the University of Tokyo, where he has been a Lecturer since 2018. From 2016 to 2018, he was also a Visiting Researcher with Paris Diderot University (Paris7), Institut Langevin, Paris, France. His research interests include audio signal processing, acoustic inverse problems, and spatial audio. He was the recipient of RIKEN Prize Innovative Young Researcher Award by the Acoustical Society of Japan in 2015, and the Research Award by Fumio Foundation for Information Technology in 2018.

Hiroshi Saruwatari (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Nagoya University, Nagoya, Japan, in 1991, 1993, and 2000, respectively. In 1993, he joined SECOM IS Laboratory, Tokyo, Japan, and in 2000, Nara Institute of Science and Technology, Ikoma, Japan. Since 2014, he has been a Professor with The University of Tokyo, Tokyo, Japan. His research interests include statistical audio signal processing, blind source separation, and speech enhancement. He has put his research into the world's first commercially available independent-component-analysis-based BSS microphone in 2007. He was the recipient of several paper awards from IEICE in 2001 and 2006, from TAF in 2004, 2009, 2012, and 2018, from IEEE-IROS2005 in 2006, and from APSIPA in 2013 and 2018. He was also the recipient of the DOCOMO Mobile Science Award in 2011, the Ichimura Award in 2013, The Commendation for Science and Technology by the Minister of Education in 2015, the Achievement Award from IEICE in 2017, and the Hoko-Award in 2018. He has been professionally involved in various volunteer works for IEEE, EURASIP, IEICE, and ASJ. Since 2018, he has been an APSIPA Distinguished Lecturer.