Bayesian inference and neural estimation of acoustic wave propagation

Yongchao Huang[1] and Yuhang He[2] and and Hong Ge[3]

***Abstract*— In this work, we introduce an integrated framework which combines physics and machine learning methods to analyse acoustic signals. Three methods are developed for this task: a Bayesian inference approach for inferring the spectral acoustics characteristics, a neural-physical model which equips a neural network with forward and backward physical losses, and the non-linear least squares approach which serves as benchmark. The inferred propagation coefficient leads to the room impulse response (RIR) quantity which can be used for relocalisation with uncertainty. The simplicity and efficiency of this framework is empirically validated on simulated data.**

I. Introduction

Analyzing acoustic wave propagation is useful across various fields, including robotic navigation [37], [4], [51], [46], virtual reality [16], [64], autonomous driving [72], [65], and architectural acoustics [44]. Accurate characterization, notably quantifying the *room impulse response* (RIR [45]), is important for precise localization and environment modeling [41], [5]. Conventional physics-based models, such as wave equations [21], [63] and ray tracing [31], [70], are effective but computationally intensive and struggle with uncertainty and real-world acoustic noise. Recent data-driven approaches, including Bayesian inference and neural networks, address these limitations. Bayesian methods [20], [80], [12], [5] utilize prior knowledge for efficient inference and uncertainty quantification, even with sparse data [81], [82], [13]. Neural networks, especially physics-informed variants [52], [12], [79], directly learn acoustic properties from data, enabling rapid real-time inference [18], [35], [1], [2], [75], [62]. However, current methods often require extensive training datasets, lack comprehensive uncertainty estimation, and inadequately incorporate physical insights [6], [69], [2]. This paper proposes an integrated framework combining Bayesian inference, neural-physical modeling, and non-linear least squares for spectral acoustic analysis. Our method bridges physics-based and data-driven methods, enhancing interpretability, accuracy, and robustness for practical robotics and acoustics engineering applications.

II. Estimation of acoustic wave characteristics

The frequency domain sound wave equation can be written as (details see Appendix.B):

$$\frac{\partial^2}{\partial x^2}\tilde{P}(x,\omega)+\frac{\omega^2}{c^2}\tilde{P}(x,\omega)=0 \tag{1}$$

where ω is the angular frequency. We consider a one-direction propagating wave at location x, which can be represented as $\tilde{P}(x,\omega)=\tilde{P}(x_0,\omega)e^{-\gamma(x-x_0)}$. Substituting this single direction wave into 1, we obtain

$$(\alpha^2+i2\alpha\kappa)\tilde{P}(x_0,\omega)e^{-(\alpha+i\kappa)(x-x_0)}=0 \tag{2}$$

where $\alpha(\omega)$ is the frequency-dependent attenuation coefficient (also called damping coefficient), κ is the wave number (also termed spatial velocity). Together they make the *wave propagation coefficient* $\gamma(\omega)=\alpha(\omega)+i\kappa(\omega)$. $\tilde{P}(x_0,\omega)$ are the magnitudes for each frequency components at a reference location x_0 (e.g. x_0=0), determined by initial conditions. The fact that Eq.2 should be satisfied everywhere hints that it actually reflects the property of the medium itself.

Given two wave profiles measured at 2 locations, i.e. a speaker sitting at x_1 and a receiver at x_2, our goal is to estimate the values of the frequency-dependent acoustics characteristics ($\gamma(\omega)$,$\alpha(\omega)$,$\kappa(\omega)$), as they are linked to the property (i.e. impedance, elasticity) of the medium the wave propagates through. Traditionally, wave propagation characteristics are estimated using least squares [30]. Here we introduce two further methodologies, namely Bayesian inference, and neural parameter estimation.

A. Bayesian wave propagation analysis

To obtain the posterior distributions of $\alpha(\omega)$ and $\kappa(\omega)$ in frequency domain using Bayesian inference, incoorporating prior knowledge and data information, we re-formulate the problem as: given a set of $\alpha(\omega)$ and $\kappa(\omega)$ values (e.g. those sampled from prior distributions), we can calculate the wave profile at x_2 as $\tilde{P}(x_2,\omega)=\tilde{P}^m(x_1,\omega)e^{-\gamma(\omega)(x_2-x_1)}$, and under certain noise variation σ^2, this predicted waveform should distribute around the measured wave profile, i.e. $\tilde{P}(x_2,\omega)\sim\mathcal{N}(\tilde{P}^m(x_2,\omega),\sigma^2)$, where the superscript m denotes measurements. The same applies to $\tilde{P}^m(x_1,\omega)$, if the wave were to travel inversely in space from x_2 to x_1. Putting them together, we have:

$$\begin{bmatrix}\tilde{P}^m(x_2,\omega)-\tilde{P}^m(x_1,\omega)e^{-\gamma(\omega)(x_2-x_1)}\\ \tilde{P}^m(x_1,\omega)-\tilde{P}^m(x_2,\omega)e^{\gamma(\omega)(x_2-x_1)}\end{bmatrix}\sim\mathcal{N}(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2\\\sigma^2\end{bmatrix}) \tag{3}$$

The wave equation Eq.2 serves as a constraint, it can be either loosely satisfied (i.e. added to the likelihood Eq.3 to allow noise perturbation) or used as a test condition (hard constraint) to filter out the samples. Instead of inferring parameter values via sampling, alternatively an optimization person may thrive to minimize the LHS of Eq.3, in which nonlinear least squares [67] or maximum likelihood methods [19] can be applied. The detailed procedure of Bayesian inference is described in Algorithm.1.

[1]Yongchao Huang is with Dept. of Computing Science, University of Aberdeen, UK yongchao.huang@abdn.ac.uk

[2]Yuhang He is with Dept. of Computer Science, University of Oxford, UK yuhang.he@cs.ox.ac.uk

[3]Hong Ge is with Dept. of Engineering, University of Cambridge, UK hg344@cam.ac.uk

B. Neural parameter estimation

Instead of using Bayesian inference for parameter estimation, we could also build a neural network for predicting the parameter values. These neural estimated physical quantities, i.e. the attenuation coefficient α and phase speed κ, are then plugged into the physics to yield a predicted waveform profile at the receiver position, i.e. $\tilde{P}(x_2,\omega) = \tilde{P}(x_1,\omega)e^{-\gamma(x_2-x_1)}$. The gap between the predicted waveform $\tilde{P}(x_2,\omega)$ and the measurement $\tilde{P}^m(x_2,\omega)$, typically the sum of magnitude discrepancies over the frequency spectrum across training instances, can be minimized by adjusting the network weights via optimisation routines. Fig.1 sketches the design of the the neural-physical model.

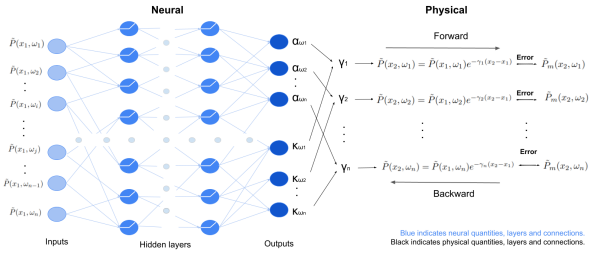


Fig. 1: The neural-physical model architecture.

A wise design of the loss functional could accelerate the learning phase and improve accuracy. As the physical process is reversible in space, i.e. if the received signal is sent back to the speaker, then same input wave should be derived using the neural net predicted propagation coefficient. This motivates our forward-backward loss design, where the forward loss accounts fro the mismatch between the received signals, while the backward loss measures the discrepancies between the input signals if we reverse the physical process. We can further implement the principle of minimum entropy of the predicted coefficients across all instance predictions, as the environment or medium in which the wave propagating in stays static over the experimental time horizon. With these principles in mind, the following target functional is constructed to guide the training of the neural-physical model:

$$\begin{aligned}\text{Loss}(\theta;x_1,x_2) = \frac{1}{N\times n}\sum_{i=1}^{N}\sum_{j=1}^{n}\{[\tilde{P}_i^m(x_2,\omega_j)-\tilde{P}_i(x_2,\omega_j,\theta)]^2 \\ +[\tilde{P}_i^m(x_1,\omega_j)-\tilde{P}_i(x_1,\omega_j,\theta)]^2+[C-I(\frac{1}{N}\sum_{i=1}^{N}C_{i,\cdot})^T]^2\}\end{aligned} \tag{4}$$

where θs are the neural net weights. ω is the angular frequency, n is the total number of symmetric frequency components after discrete Fourier transform (DFT). N is the number of training data point (each data point is a row vector of Fourier magnitudes with length n). $\tilde{P}^m(x_1)$ and $\tilde{P}^m(x_2)$, both with size (N,n), are the Fourier magnitude matrix of the *measured* signals, $\tilde{P}(x_1)$ and $\tilde{P}(x_2)$ are their corresponding *predicted* counterparts [1]. I is a N-length column vector with identity elements. $C_{N\times n}$ is the propagation coefficient matrix from the outputs of the neural network:

$$C_{N\times n} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \dots & \alpha_{1,n} & \kappa_{1,1} & \kappa_{1,2} & \dots & \kappa_{1,n} \\ \alpha_{2,1} & \alpha_{2,2} & \dots & \alpha_{2,n} & \kappa_{2,1} & \kappa_{2,2} & \dots & \kappa_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{N,1} & \alpha_{N,2} & \dots & \alpha_{N,n} & \kappa_{N,1} & \kappa_{N,2} & \dots & \kappa_{N,n} \end{bmatrix}$$

with each column corresponds to a frequency and each row represents the coefficients prediction for input instance i.

The first two terms in the curl brackets of Eq.4 measure the forward and backward discrepancies, respectively. The third term measures the row concentration: each row of the propagation coefficient coefficient matrix is de-meaned by the row average to form a residual matrix; minimizing the magnitude of the residual matrix encourages all rows to be close to each other (as they should be). Note all matrix squares in Eq.4 denote element-wise squares (i.e. squared Frobenius norm). The acoustic wave characteristic coefficients in matrix $C_{N\times n}$, predicted by the neural net using N data points, naturally provides uncertainty quantification. Uncertainty can also be obtained using neural net ensembling, with each network exposed to a sub-set of data.

C. Non-linear least squares

To minimize the LHS in Eq.3 (same as the first two terms in Eq.4), we use non-linear least squares without regularization. We can first apply the log trick for exponential functions, which gives $\log\tilde{P}^m(x_{i2},\omega) - \log\tilde{P}^m(x_{i1},\omega) \approx -\gamma(\omega)(x_{i2}-x_{i1})$. Expanding the equation to host rows as (x_{i1},x_{i2}) pair instances and use frequency as column index, we obtain the matrix representation $\Delta\log\tilde{P} \approx -\Delta\mathbf{x}\gamma^T$, where $\Delta\log\tilde{P}_{N\times n}(x_{i1},x_{i2},\omega_j) = \tilde{P}_{N\times n}(x_{i2},\omega_j) - \tilde{P}_{N\times n}(x_{i1},\omega_j)$ is the measurements difference matrix, $\Delta\mathbf{x}_{N\times 1} = \{x_{i2}-x_{i1}\}_{i=1}^N$ is the collection of travel distances. $\gamma_{N\times 1}(\omega_i) = \alpha(\omega_i) + i\kappa(\omega_i)$ is a vector of propagation coefficients corresponding to each of the n frequency components.

As $\Delta\log\tilde{P}_{N\times n}$ is matrix-valued, we have a system of linear equations: $\Delta\log\tilde{P}_i = -\Delta x_i \times \gamma^T$. If we make x as the diagonal entries of a diagonal matrix $\Delta X_{N\times N}$, and construct a new matrix $\Gamma_{N\times n}$ by dulicating γ^T along each row, then the following matrix least squares problem can be formulated:

$$\hat{\gamma} = \arg\min_{\gamma}\|\Delta\log\tilde{P}_{N\times n} - (-\Delta X_{N\times N}\Gamma_{N\times n})\|_2^2 \tag{5}$$

by solving the normal equations, we have

$$\hat{\Gamma}=-(\Delta X^T\Delta X)^{-1}\Delta X^T\Delta\log\tilde{P}=-(\Delta X)^{-1}\Delta\log\tilde{P}$$

This least squares solution implies that $\hat{\Gamma}_{i,\cdot} = -\frac{\Delta\log\tilde{P}_{i,\cdot}}{x_{i2}-x_{i1}}$, i=1, 2, ..., N, which, not surprisingly, matches exactly the proposed wave equation solution [30] $\tilde{P}^m(x_{i2},\cdot) = \tilde{P}^m(x_{i1},\cdot)e^{-\gamma(x_{i2}-x_{i1})}$. In an ideal scenario (i.e. noise free), all rows of $\hat{\Gamma}$ should be equal; when noise is present, for each input pair, we obtain an estimated version of $\gamma(\omega)$, and least squares in this way yields an interval estimate. This uncertainty, induced by noise, is different from the Bayesian credible interval which is generated directly by sampling (noise is encoded in the likelihood and the MH step though).

[1] Note that, as all inputs and outputs are in frequency domain, these matrix entries are mostly complex-valued. For two complex numbers, when evaluating their discrepancy, contributions from both real and imaginary parts are aggregated.

III. Estimating Room Impulse Response (RIR)

The room impulse response $RIR(x,t)$ is a convolutional basis signal in time domain which, after convolved with the input signal $P(x,t)$, can be used to generate the output time domain signal:

$$P(x_1,t) \circledast RIR(x_2-x_1,t) = P(x_2,t) \tag{6}$$

where $\circledast$ denotes convolution operator. $RIR(x,t)$ has the same size as the input signal (e.g. fixed sampling rate), and is generally an indicator of the acoustics characteristics of the room (e.g. air density), given the ambient conditions (e.g. temperature, humidity, etc). Eq.6 is another representation of the wave propagation phenomenon: given the room geometry and conditions, it assumes a fixed convolutional operand $RIR(x,t)$ between two fixed locations. It reflects the state of the medium in which the wave propagates in.

Applying Fourier transform to Eq.6 gives the frequency domain representation[2]:

$$\tilde{P}(x_1,\omega) \odot \tilde{RIR}(x_2-x_1,\omega) = \tilde{P}(x_2,\omega) \tag{7}$$

from which we obtain

$$\tilde{RIR}(x_2-x_1,\omega) = \tilde{P}(x_2,\omega)./\tilde{P}(x_1,\omega) = e^{-\gamma(x_2-x_1)} \tag{8}$$

where $./$ denotes element-wise division (i.e. the *Hadamard* division). We observe that, if $\tilde{P}(x_1,\omega) := \mathbf{1}$, which corresponds to a Dirac delta impulse in time domain [3], then we have $\tilde{RIR}(x_2-x_1,\omega) = \tilde{P}(x_2,\omega)$. That is, $\tilde{RIR}(x_2-x_1,\omega)$ equals the response (output of the system) measured at x_2, when the excitation (input of the system) at x_1 is a Dirac unit impulse. This implies that, with he learned parameters (α,κ) of the system, we can purposely inject a Dirac unit impulse into the system and obtain $\tilde{RIR}(x_2-x_1,\omega)$ by analysing the received signal at x_2. The time domain $RIR(x_2-x_1,t)$ can be obtained by inverting the frequency domain signal.

Note that, in a homogeneous environment, i.e. the properties of the medium (e.g. density) in which wave propagates are constant over time and space (e.g. humidity is not changing over time or temperature), the wave propagation coefficient γ remains the same for all locations (i.e. orientation isotropic). The RIR, however, is a function of both the wave propagation coefficient and the wave travel distance x_2-x_1 (the absolute coordinates are not relevant).

IV. Relocalisation

We demonstrate an application of the aforementioned method for relocalisation in robotics. We have the scenario where a moving robot would like to evaluate its distance from known speaker positions via collecting and processing information from fixed-position speakers. Given two wave profiles $\tilde{P}_1(\omega)$ and $\tilde{P}_2(\omega)$, we can use Eq.7 to inversely calculate the distance $\Delta x = x_2 - x_1$:

$$\Delta x = -\frac{1}{\gamma}\ln \tilde{RIR}(x_2-x_1,\omega) = -\frac{1}{\gamma}\ln\frac{\tilde{P}(x_2,\omega)}{\tilde{P}(x_1,\omega)} \tag{9}$$

For 1D navigation, knowing the speaker location x_1 (e.g. a fixed speaker), we are able to estimate the current position x_2 of the mobile robot equipped with a receiver. Due to noise, however, the resulted Δx ratio may not be constant across frequencies. We can therefore obtain a distribution of Δx and take a mean estimate:

$$\widehat{\Delta x} = \frac{1}{n}\sum_{i=1}^{n} -\frac{1}{\gamma}\ln\frac{\tilde{P}(x_2,\omega_i)}{\tilde{P}(x_1,\omega_i)} \tag{10}$$

where n is the number of frequency components in the Fourier spectrum.

For 2D navigation problem [4], however, we can only locate x_2 on the circle with radius Δx centering the speaker position; to locate the absolute location of the robot, we need signals from extra two speakers, calculate each radius and take the intersection as the search area, see Figure.2.

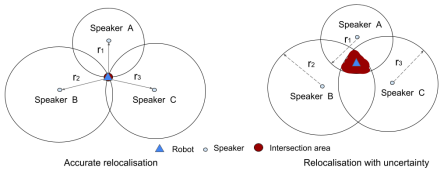
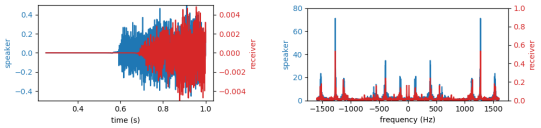


Fig. 2: 2D/3D relocolisation diagram.

V. Experiments

We use the *SoundSpaces 2.0* [11] simulator with the *Matterport3D* dataset [10] to simulate a homogeneous room closely resembling a real environment. A fixed-position speaker emits a constant signal, while 9 spatially distributed receivers capture the resulting 1-second, 16kHz signals without interference. For Bayesian inference and least squares, only one speaker-receiver pair is used for training; for the neural-physical model, 8 pairs are used for training and 1 for testing. See Appendix D for full details.



(a) Time domain. (b) Frequency domain.

Fig. 3: A simulated speaker-receiver signal pair.

A. Bayesian inference results

Priors of the Bayesian model are specified as follows:

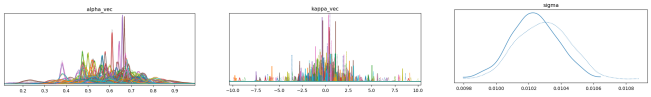
$$\alpha \sim \mathcal{N}(\mathbf{1},I), \kappa \sim \mathcal{N}(\mathbf{0},10^2 I), \sigma \sim \textit{HalfNormal}(std=1)$$

[2] $\odot$ here denotes element-wise multiplication (also known as the Hadamard product or Schur product). Here the time-frequency transform trick applies: convolution in time domain is equivalent to multiplication in frequency domain, and vice versa.

[3] The frequency domain response of the Dirac delta function is a constant with a magnitude of 1 at all frequencies.

[4] In 3D, the circles become balls, but the same principle follows.

and the forward likelihood specified previously in the first half of Eq.3 is used. The *NUTS* sampler [42], a variant Hamiltonian Monte Carlo method, is used to perform Bayesian inference on the posterior density of (α, κ, σ). The MCMC samples are shown in Fig.4, which suggests the success of MCMC inference of the posterior.



(a) α (b) κ (c) σ

Fig. 4: MCMC sample trajectories. Two MCMC chains, each with 1000 iterations, are generated.

After inferring the wave propagation coefficient γ, we plug it into Eq.8 and obtain RIR (showon in Fig.11 in Appendix.D). To validate Bayesian learning, given the speaker signal $\tilde{P}(x_1, \omega)$, we use the wave solution $\tilde{P}(x_2, \omega) = \tilde{P}(x_1, \omega)e^{-\gamma(\omega)(x_2 - x_1)}$ to predict the wave profile $\tilde{P}(x_2, \omega)$ and compare it with ground truth. This is shown in Fig.5, where we observe a good fit of the training wave, while the predicted test wave is reasonably good. The uncertainties, generated by sampling γ from the posterior, give credible interval for the predictions.

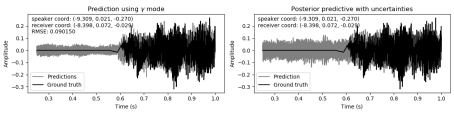
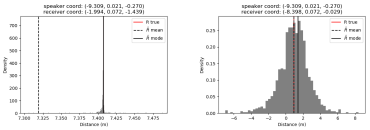


Fig. 5: Test wave profile prediction using γ posterior MAP and MCMC samples.

Having the two waves measured at the speaker and the robot positions, we can use Eq.9 to obtain a distribution of the distance Δx; uncertainty is also propagated from parameter estimation. Fig.6 shows the estimated distances for two receivers. In both cases, with noise presence, we still observe good consistency with ground truth.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 6: Bayesian inference: estimating the distance between speaker and receiver using MAP γ.

B. Neural parameter estimation results

A fully connected, multi-layer perceptron (MLP) network [5], with *ReLU* activations and layer sizes (4800, 128, 256, 256, 128, 4800), is constructed to consume the speaker spectral waveform and outputs the propagation coefficient.

[5]We also presented results using an auto-encoder architecture in Appendix.D-D.

Fig.7 compares the training and test performances in time domain. We observe that the test performance, quantified by the RMSE metric, is slightly worse than the Bayesian method. The distance estimations are presented in Fig.8.

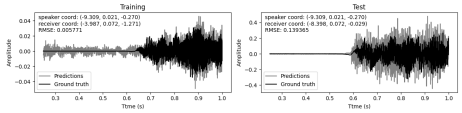
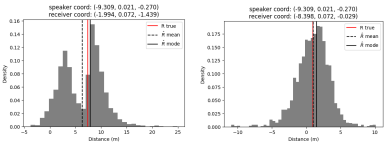


Fig. 7: Neural-physical model performances in time domain on training and test samples.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 8: Estimating the distance between speaker and receiver using neural estimated γ (posterior MAP).

C. Comparison

A comparison of the estimated distances are made in Table.I, in which we also include results from Bayesian mean estimation (Appendix.D-B) and neural autoencoder estimation (Appendix.D-D). Among these three distance estimation methods, Bayesian inference wins in both cases, although the other two methods also show comparable performances.

	Test receiver 1			Test receiver 2		
Statistics	Mean	Mode	Std	Mean	Mode	Std
Bayesian MAP	**7.319**	**7.407**	**0.466**	**0.936**	1.425	1.715
Bayesian mean	2.749	2.896	1.311	0.416	0.487	**0.855**
Neural (MLP)	6.290	7.883	3.955	0.986	1.509	2.395
Neural (Autoencoder)	6.308	7.623	3.880	0.961	1.869	2.375
Least squares	5.458	4.998	2.968	0.854	**1.318**	2.016
Ground truth	**7.407**	-	-	**0.943**	-	-

TABLE I: Comparison of three distance estimation methods.

VI. Conclusions

This work combines machine learning and physics to infer acoustic wave characteristics in the frequency domain. We evaluate three methods, namely Bayesian inference, neural-physical estimation, and non-linear least squares, all shown to be effective in this task. Bayesian inference leverages small data, prior knowledge, and uncertainty quantification, while the neural and least squares approaches offer speed and physically consistent outputs (i.e. symmetry for the physical quantities). These physics-informed machine learning models enhance sample efficiency, expressivity, and interpretability, and accelerate learning via embedded dynamics. The proposed framework is robust, efficient, and applicable to tasks such as RIR estimation and robot relocalisation using acoustics-only measurements.

APPENDIX

A. Background, related work and notations

Acoustic wave propagation analysis is fundamental to a wide range of applications, including robotic navigation [37], [4], [51], [46], immersive virtual reality [16], [64], autonomous driving [72], [65], and architectural acoustics [44]. A key objective in this domain is the precise characterization of acoustic signals, particularly the room impulse response (RIR [45]), which is critical for accurate robot localization and environmental modeling [41], [5].

Traditionally, acoustic wave propagation has been modeled using physics-based approaches, including wave equations [21], [63] and geometry-inspired methods such as ray tracing [31], [70]. In room acoustics modelling, this bifurcation is reflected in wave-based modelling [7], [49], [36], [8] and geometry-based techniques [61]. Geometry-based methods include ray tracing [66], the image source method (ISM) [3], beam tracing [23], and acoustic radiosity [27], [47], with the primary goal of capturing room reverberation effects comprising direct-path, specular reflections, and late reverberation.

While effective, these conventional methods face computational challenges in dynamic or complex environments and often lack robust handling of uncertainty and noise which are inherent in real-world acoustic signals. To overcome these limitations, data-driven methods have gained traction. Bayesian inference approaches [20], [80], [12], [5] leverage prior knowledge to infer propagation parameters efficiently and offer robust predictions with quantified uncertainty, even with limited or sparse data [81], [82], [13]. Neural networks [18], [35], [1], especially physics-informed ones [52], [12], [79], demonstrate strong potential in learning complex acoustic characteristics directly from data [52], [2], [75], [62], and offer rapid inference suitable for real-time applications.

Recent developments have also focused on learning room impulse responses directly with deep learning models [54], [68], [55], [53], [15], [39], [56], [40]. However, these methods generally require large datasets of source-to-receiver RIRs (s2r-RIRs). In contrast, approaches like *SoundNeRirF* exploit robot-recorded sounds from various positions to learn implicit receiver-to-receiver RIRs (r2r-RIRs), which are more practical to collect in real-world scenarios.

Relocalisation through acoustic signals plays a vital role in tasks such as positioning and mapping in robotics [74], as well as enhancing immersive experiences in VR and gaming [32], [73]. While traditional methods rely on data like LiDAR [74], [22], RGB imagery [71], and spatial audio [25], [26], data-driven acoustic modeling offers a complementary or alternative pathway with strong potential.

Despite progress, existing methods still face challenges: they often depend on extensive training data, struggle with uncertainty quantification, and inadequately incorporate physical knowledge [6], [69], [2]. In response, this paper proposes a novel framework that integrates Bayesian inference, neural-physical modeling, and classical non-linear least squares to infer spectral acoustic characteristics, estimate RIR and relocate the robot efficiently. This approach bridges the gap between data-driven and physics-based models, enhances the interpretability, accuracy and robustness of acoustic parameter estimation for practical applications in robotics and acoustics engineering.

Notations. We mainly concern about spatial and spectral notations. Frequency domain quantities are denoted with a tilde hat, e.g. $\tilde{P}^m(x_{i1}, \omega_j)$. The superscript m denotes measurement, distinguishing from its predicted counterpart without superscript. Speaker signal is always implied by the subscript "1", while receiver signal by "2". Whenever double indices is necessary, e.g. matrix entries, the first subscript indicates instance index, e.g. x_{i1} corresponds to the position of speaker i. As each speaker-receiver pair gives a vector signal, aggregating all signals together forms a $N \times n$ matrix, where N is the total number of signals and n is the number of frequency components for example. Therefore, row i accommodates the signal corresponding to the (x_{i1}, x_{i2}) pair, and column j refers to the magnitudes corresponding to (angular) frequency ω_j. Vectors and matrices are represented in bold whenever convenient.

B. Preliminaries

Bayesian inference

One of the major tasks in Bayesian inference is to infer the posterior distribution of some parameters of interest via the *Bayes rule*:

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})} \tag{11}$$

where $\mathbf{z} = z_{1:m} \in \mathbb{R}^m$ are the latent variables that we are interested in performing inference on, $\mathbf{x} = x_{1:n} \in \mathbb{R}^d$ are the observations, $p(\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ are the prior and posterior distributions, respectively, $p(\mathbf{x}|\mathbf{z})$ is the data likelihood given $\mathbf{z}$, and $p(\mathbf{x}) = \int_{\mathbb{R}^m} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$ is the marginal likelihood, which can be hard to compute or intractable. Sampling methods such as Markov chain Monte Carlo (MCMC) [14] and variational inference (VI [24]) methodologies can be used to perform approximate inference [43], and they provide guarantees at stationary states. However, MCMC methods are known to be computational intensive [78] in large data settings [28], and high dimensions [17], [34], or when the model is complex [43]. Bayesian inference has the advantage of sample efficiency and uncertainty quantification, it also offers the flexibility of continuous inference. Besides, it helps detect any inconsistency between the specified probabilistic model and data.

Sound wave propagation

In time domain, the propagating waves are governed by the wave equation [6]

$$\frac{\partial^2 P(x,t)}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 P(x,t)}{\partial t^2} = 0 \tag{12}$$

[6]This is, however, a very simplified version of more complex wave propagation theories which may involve wave inference, overlapping, reflection and refraction, etc.

where $P(x,t)$ is the sound pressure which can be directly measured. Eq.12 describes a wave varying in space and time (i.e. a spatial-temporal model [7]). We seek to find a solution $P(x,t)$ to this wave equation. In acoustics, we concern about the particle displacement u(x,t) which is normally proportional to the sound pressure $P(x,t)$. For simplicity, we assume the equality $P(x,t) = u(x,t)$. And the time domain solution $u(x,t)$ for a 1D propagating sound wave can be represented as [30]

$$u(x,t) = Ae^{\pm\alpha x}e^{i(\omega t \pm \kappa x)} \tag{13}$$

where ω is the angular frequency, $\alpha(\omega)$ is the frequency-dependent attenuation coefficient (also called damping coefficient), κ is the wave number (also termed spatial velocity). Together they make the *wave propagation coefficient* $\gamma(\omega) = \alpha(\omega) + i\kappa(\omega)$. Notes that [30], $\alpha(\omega)$ is a positive even function, and $\kappa(\omega)$ is an odd function and positive for $\omega > 0$. The $\pm$ sign indicates the wave propagating in two directions. Energy dissipation is induced when wave propagating in, for example, viscoelastic medium, which is captured by the attenuation term $e^{\pm\alpha x}$, i.e. wave magnitude attenuation over the travel distance x; phase changes are determined by the term $e^{i(\omega t \pm \kappa x)}$. In acoustics, we also have the relation

$$|\kappa(\omega)| = \frac{2\pi}{\lambda(\omega)} = \frac{\omega}{c(\omega)} \tag{14}$$

where c is the wave speed. Note that, α, κ, c are both frequency-dependent and material-dependent.

More frequently, wave propagation phenomenon is analysed in frequency domain [30]. We could apply Fourier transform to the time domain wave equation, which gives [30]

$$\frac{\partial^2}{\partial x^2}\tilde{P}(x,\omega) + \frac{\omega^2}{c^2}\tilde{P}(x,\omega) = 0 \tag{15}$$

and the frequency domain solution

$$\tilde{P}(x,\omega) = \tilde{P}(x_0,\omega)e^{-\gamma x} + \tilde{P}_0'(\omega)e^{\gamma x} \tag{16}$$

where $\tilde{P}(x_0,\omega)$ and $\tilde{P}_0'(\omega)$ are the magnitudes for each frequency components propagating in left and right directions (they are analogous to the initial time domain magnitude A in Eq.13). They can be determined by initial conditions.

Substituting the wave solution Eq.16 into the wave equation Eq.15, we have

$$(\alpha^2 + i2\alpha\kappa)\tilde{P}(x_0,\omega)e^{-(\alpha+i\kappa)x} + (\alpha^2 + i2\alpha\kappa)\tilde{P}_0'(\omega)e^{(\alpha+i\kappa)x} = 0 \tag{17}$$

For the interest of our problem, we would like to just consider a one-direction propagating wave $\tilde{P}(x,\omega) = \tilde{P}(x_0,\omega)e^{-\gamma x}$, which simplifies Eq.17 to be $(\alpha^2 + i2\alpha\kappa)\tilde{P}(x_0,\omega)e^{-(\alpha+i\kappa)x} = 0$, which is a complex-valued equality.

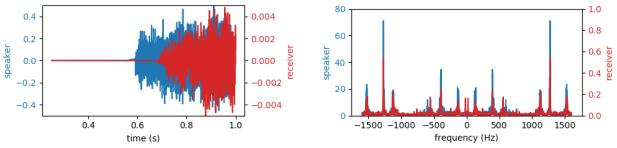
[7]Similar natural phenomena include heat transfer which is governed by the heat equation [76].

C. Algorithm: Bayesian inference of acoustic wave characteristics

The algorithm for Bayesian inference of the wave propagation coefficient is presented in in Algorithm.1.

D. Experiments: full details

Experimental setup: we use a room simulator *SoundSpaces 2.0* [11] to simulate a homogeneous room with *Matterport3D* dataset [10]. The simulated room maximally resembles a real room environment. In the simulated environment, a static speaker with fixed position constantly sends out a fixed signal, and 9 receivers, located at different positions, collect the corresponding signals without inference. All signals are 1 second long with sampling rate 16k. For Bayesian inference, only one pair of speaker-receiver signals, shown in Fig.9, is used as training data; for the neural-physical model, 8 signal pairs are used in training, and the 1 remaining signal pair serves as test set. Least squares utilizes just 1 pair of signals for estimation.



(a) Time domain. (b) Frequency domain.

Fig. 9: Simulated speaker-receiver signal pair used in Bayesian training. Speaker coordinate: (-9.308,0.021,-0.270), receiver coordinate: (-1.994,0.072,-1.439), $\Delta x = 7.408$.

A. Bayesian inference results

The probabilistic model is specified as follows: multivariate Gaussian priors are used for α and κ, and half normal distribution is applied for noise variance, i.e.

$$\alpha \sim \mathcal{N}(\mathbf{1}, I), \kappa \sim \mathcal{N}(\mathbf{0}, 10^2 I), \sigma \sim \textit{HalfNormal}(std = 1)$$

We use the forward likelihood specified previously in the first half of Eq.3. The *NUTS* sampler [42], a variant Hamiltonian Monte Carlo method, is used to perform Bayesian inference on the posterior density of (α, κ, σ). Table.II shows the statistics of 6 parameters [8] and the noise variance. The convergence diagnostic $\hat{R}$ (i.e. the Gelman-Rubin statistic) [9], which evaluates the between and within-chain mixing, concentrates around 1, and the effective sample size (ESS, the number of "independent" samples) is sizeable. In Fig.10, the damping coefficient α is positive, and the wave number symmetrically distributed around 0, which is consistent with the physics. This suggests that MCMC inference of the posterior has been successful.

[8]As the number of frequency components is large, we arbitrarily select these 6 parameters.

Algorithm 1: Bayesian inference of wave propagation coefficients (Metropolis-Hastings sampling as an example)

Inputs: two time series waveform measurements $P^m(x_1, t_1)$, $P^m(x_2, t_2)$.

Calculate the travel distance $\Delta x = x_2 - x_1$. Convert time domain measurements $P^m(x_1, t_1)$, $P^m(x_2, t_2)$ into frequency domain, obtain frequency domain data $\tilde{P}^m(x_1, \omega)$, $\tilde{P}^m(x_2, \omega)$.

Set prior distributions $p_r(\alpha)$, $p_r(\kappa)$ and noise level $p_r(\sigma)$.

Set sampling distributions $p_{lik}(\tilde{P}(x_2, \omega)|\tilde{P}^m(x_1, \omega), \alpha, \kappa)$ and $p_{lik}(\tilde{P}(x_1, \omega)|\tilde{P}^m(x_2, \omega), \alpha, \kappa)$ to be Gaussians with variance σ^2, where $\tilde{P}(x_2, \omega) = \tilde{P}^m(x_1, \omega)e^{-\gamma(\omega)(x_2 - x_1)}$ and $\tilde{P}(x_1, \omega) = \tilde{P}^m(x_2, \omega)e^{\gamma(\omega)(x_2 - x_1)}$ are the predicted wave profiles.

Choose proposal distributions q_α, q_κ and q_σ, e.g. symmetric Gaussians.

Sample an initial sample α_0 and κ_0 from prior distributions. Note α_0 and κ_0 are vectors with length equaling the number of frequencies.

For each iteration $l = 0, 1, 2, ..., L-1$, repeat:

(1) generate the next sample candidate $(\alpha_{l+1}, \kappa_{l+1}, \sigma_{l+1})$ by sampling from the proposal distributions $q_{\alpha_{l+1}|\alpha_l}$, $q_{\kappa_{l+1}|\kappa_l}$ and $q_{\sigma_{l+1}|\sigma_l}$, respectively.

(2) calculate the predicted wave profile $\tilde{P}(x_2, \omega) = \tilde{P}^m(x_1, \omega)e^{-\gamma(x_2 - x_1)}$ and $\tilde{P}(x_1, \omega) = \tilde{P}^m(x_2, \omega)e^{\gamma(x_2 - x_1)}$, where $\gamma = \alpha + i\kappa$.

(3) evaluate the acquisition value: $r = \frac{p_r(\alpha_{l+1}) \times p_r(\kappa_{l+1}) \times p_r(\sigma_{l+1}) \times p_{lik}(\tilde{P}(x_2,\omega)|\tilde{P}^m(x_1,\omega), \alpha_{l+1}, \kappa_{l+1}, \sigma_{l+1}) \times p_{lik}(\tilde{P}(x_1,\omega)|\tilde{P}^m(x_2,\omega), \alpha_{l+1}, \kappa_{l+1}, \sigma_{l+1})}{p_r(\alpha_l) \times p_r(\kappa_l) \times p_r(\sigma_l) \times p_{lik}(\tilde{P}^m(x_1,\omega)|\tilde{P}(x_2,\omega), \alpha_l, \kappa_l, \sigma_l) \times p_{lik}(\tilde{P}^m(x_2,\omega)|\tilde{P}(x_1,\omega), \alpha_l, \kappa_l, \sigma_l)}$.

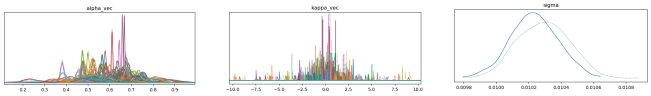
(4) evaluate the LHS of the wave equation at x_2: $LHS = (\alpha^2 + i2\alpha\kappa)\tilde{P}(x_1, \omega)e^{-(\alpha + i\kappa)(x_2 - x_1)}$.

(5) generate a uniform random number $r' \in [0, 1]$. If $r' \leq r$, $\alpha > 0$ and $LHS = 0$, accept the candidate. Else, reject the candidate and set $\alpha_{l+1} = \alpha_l$, $\kappa_{l+1} = \kappa_l$ and $\sigma_{l+1} = \sigma_l$ instead.

Return sample trajectories $\alpha_0, \alpha_1, ...\alpha_{L-1}$ and $\kappa_0, \kappa_1, ...\kappa_{L-1}$.

Quantity	α_0	α_{500}	α_{1000}	κ_0	κ_{500}	κ_{1000}	σ
Mean	1.561	1.603	1.536	0.348	-0.689	0.156	0.010
Std	0.704	0.709	0.728	10.789	9.818	10.133	0.000
$\hat{R}$	1.000	1.000	1.000	1.010	1.000	1.010	1.050
ESS (tail)	569	493	321	569	750	486	255

TABLE II: Summary statistics of MCMC samples.



(a) α (b) κ (c) σ

Fig. 10: MCMC sample trajectories. Two MCMC chains, each with 1000 iterations, are generated.

After obtaining the propagation coefficient $\gamma(\omega) = \alpha(\omega) + i\kappa(\omega)$ using sample modes, we plug it into Eq.8 and obtain RIR, as shown in Fig.11. To validate Bayesian learning, given the speaker signal $\tilde{P}(x_1, \omega)$, we use the wave solution $\tilde{P}(x_2, \omega) = \tilde{P}(x_1, \omega)e^{-\gamma(\omega)(x_2 - x_1)}$ to predict the wave profile $\tilde{P}(x_2, \omega)$ and compare it with ground truth. This is shown in Fig.12, where we observe a good fit of the training wave, while the predicted test wave is reasonably good. The uncertainties, generated by sampling γ from the posterior, give credible interval for the predictions.

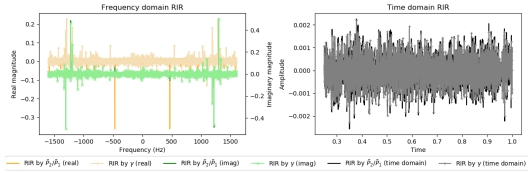
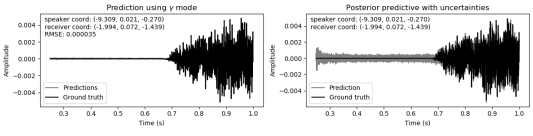
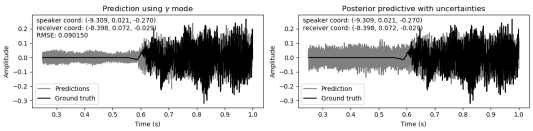


Fig. 11: Posterior MAP estimated RIR for the training signal pair.



(a) Predicting the training wave profile.

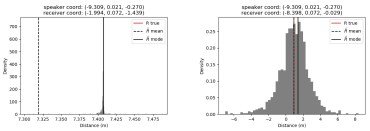


(b) Predicting a test wave profile.

Fig. 12: Predictions using posterior MAP and samples from MCMC inference results.

Having the two waves measured at the speaker and the robot positions, we can use Eq.9 to obtain a distribution of the distance Δx; uncertainty is also propagated from parameter estimation. Fig.13 shows the estimated distances for two receivers. In both cases, with noise presence, we still observe good consistency with ground truth. In parallel, we

present the results using posterior mean estimates for γ in Appendix.D-B.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 13: Bayesian inference: estimating the distance between speaker and receiver using MAP of γ.

B. Bayesian inference: results using posterior mean estimates

Probabilistic models are implemented in *PyMC* [60]; there are other probabilistic programming packages available, e.g. *Stan* [57] and *Turing.jl* [29]. The NUTS sampler [42], which is a variant of Hamiltonian Monte Carlo (HMC) sampler [58], is used to perform inference on (α, κ, σ). We sample 2 chains for each parameter, with chain length 1000; the initial 600 warm-up samples are discarded. Sampling is performed using 2 processors, which typically takes less than 5 minutes (sampling time is largely affected by the volume of data used, e.g. sampling rate, single signal or aggregated signals). Here we present the predictions of RIR (Fig.14), time domain waveforms (Fig.15) and distance (Fig.16) using mean posterior estimates from forward mode inference.

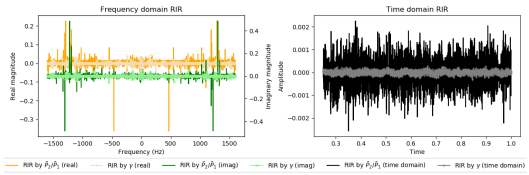
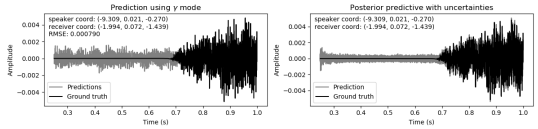
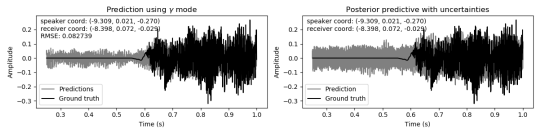


Fig. 14: Mean posterior estimated RIR for the training signal pair.

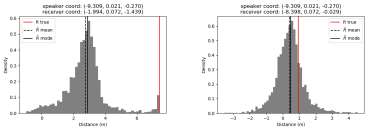


(a) Predicting the training wave profile.



(b) Predicting a test wave profile.

Fig. 15: Predictions using posterior mean and samples from MCMC inference results.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 16: Bayesian inference: estimating the distance between speaker and receiver using mean of γ.

C. Neural parameter estimation results

When training the neural-physical model, we have available the speaker and receiver positions, as well as their wave profiles. A fully connected, multi-layer perceptron (MLP) network, with *ReLU* activations and layer sizes (4800, 128, 256, 256, 128, 4800), is constructed to consume the speaker spectral waveform and outputs the propagation coefficient. Eight wave profiles are used as training set and one remains as the test instance. The loss converges after 500 epochs with a learning rate of 1e-4. Fig.17 shows the model performance, over one of the training sample, in frequency domain; we observe reasonably good match between the actual and predicted spectral signals. Fig.18 further compares the training and test performances in time domain. We observe that the test performance, quantified by the RMSE metric, is slightly worse than the Bayesian method presented in Fig.12b. Using the neural estimated wave propagation coefficient, we are able to derive the distance between any two speaker-receiver pair. A comparison of the distance estimates are made in Fig.19 (neural estimation), Fig.20 (least squares) and Table.I in which we also include results from Bayesian mean estimation (Appendix.D-B) and neural autoencoder estimation (Appendix.D-D). Among these three distance estimation methods, Bayesian inference wins in both cases, although the other two methods also show comparable performances.

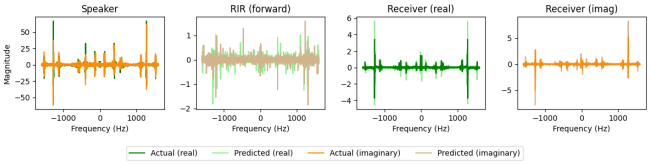


Fig. 17: Neural-physical model performance in frequency domain on the training sample with coordinate (-3.987,0.072,-1.271).

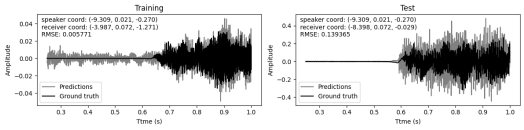
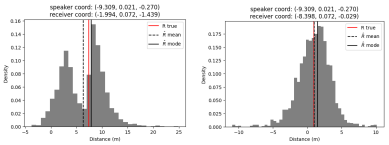
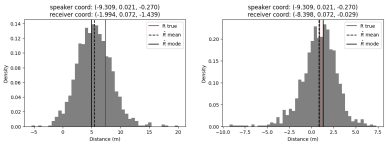


Fig. 18: Neural-physical model performances in time domain on training and test samples.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 19: Estimating the distance between speaker and receiver using neural estimated γ (posterior MAP).



(a) Test receiver 1. (b) Test receiver 2.

Fig. 20: Estimating the distance between speaker and receiver using least squares estimated γ.

Finally, we compare the estimated propagation coefficient in Fig.21, also presented are the least squares estimations obtained using a single pair of speaker-receiver signals. It is observed that, neural estimation and least squares strictly preserve the symmetries in α and κ (α is symmetric *w.r.t* y-axis, while κ is symmetric about the origin.); Bayesian inferred coefficients, as observed in Fig.10, loosely satisfy the symmetry. The neural network learns the symmetry from the inputs as well as by the loss constraint. Also, neural estimation gives estimations on a similar scale as that of least squares, while Bayesian inference gives different values depending on whether mode or mean sample values are chosen. The Bayesian MAP values coincide with neural estimation and least squares, while mean posterior values deviate from others. These frequency-dependent values, i.e. $0 \leq \alpha < 5$ and $|\kappa| < 100$, are sensible for wave propagation in air; they are dependent on some intrinsic properties of the propagating medium such as temperature, humidity, impedance and resistance.

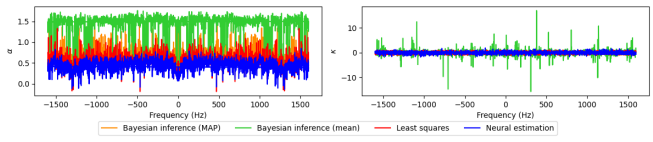


Fig. 21: Comparison of the estimated propagation coefficient using three methods. Fully connected MLP architecture is used for neural estimation. Least squares estimations are derived using the receiver signal at coordinate (-3.987,0.072,-1.271).

D. Neural parameter estimation: results using auto-encoder architecture

In parallel to the MLP architecture, here we employ the autoencoder architecture for neural parameter estimation. The autoencoder sequentially consists of an encoder and a decoder with symmetric, fully connected layers of sizes (4800, 128, 256, 256, 128, 128, 256, 256, 128, 4800) and *ReLU* activations. An conceptual diagram of the autoencoder architecture is shown in Fig.22. The input is the speaker spectral waveform and output the propagation coefficient. Eight wave profiles are used as training data and one remains as the test instance. *RMSProp* [59], a variant of the gradient descent method, is used to update the network weights. In total 500 epochs with a learning rate of 1e-4 are used.

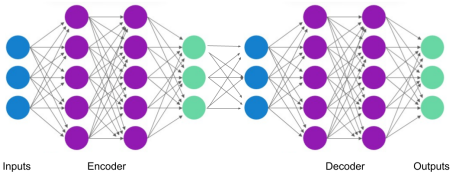
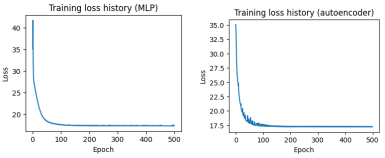


Fig. 22: An illustrative diagram of the autoencoder architecture.



(a) MLP. (b) Autoencoder.

Fig. 23: Neural network training loss history.

Fig.23 shows the training history, it is observed that the empirical losses stabilize after 200 epochs for both MLP and autoencoder scenarios. Fig.24 and Fig.25 confirm the match between actual and predicted signals. Similar to MLP, the test RMSE of autoencoder is observed to be slightly worse than the Bayesian method. The estimated distance, obtained using the mean posterior γ, is shown in Fig.26. A comparison of the estimated wave propagation coefficient is shown in Fig.27, and the comparison of the estimated distances in Table.I. It is seen that, the Bayesian MAP estimates give cloest results to the ground truth.

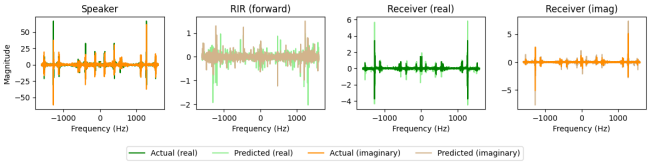


Fig. 24: Neural-physical model performance in frequency domain on the training sample with coordinate (-3.987,0.072,-1.271).

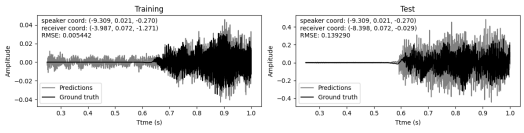
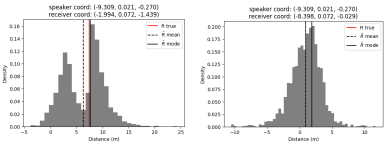


Fig. 25: Neural-physical model performances in time domain on training and test samples.



(a) Test receiver 1. (b) Test receiver 2.

Fig. 26: Estimating the distance between the speaker and receiver using neural (autoencoder) estimation of γ.

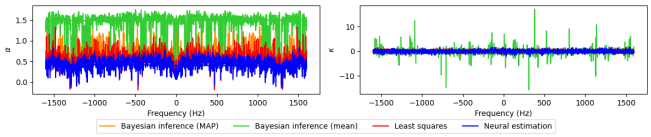


Fig. 27: Comparison of the estimated propagation coefficient using three methods. Autoencoder architecture is used for neural estimation. Least squares estimations are derived using the receiver signal at coordinate (-3.987,0.072,-1.271).

E. Discussions

The Bayesian method provides a principled framework for learning and inference. It learns from small amount of data and achieve impressive performance in time domain prediction (Fig.12) and distance estimation (Fig.13), this is attributed to the embedding of prior knowledge (i.e. priors representing the physical constraints) and the contribution from data-evidenced likelihood. Beyond priors and sample efficiency, it also offers uncertainty quantification, i.e. parameter uncertainty propagates to prediction time (e.g. Fig.12), which is advantageous particularly in safety-critical scenarios (e.g. fire alarming, self-driving) where the posterior can supply better risk assessment over all possible outcomes compatible with observations and thus more informed decisions [77]. The neural-physical method considers the forward and backward physical loss by design (see Fig.1 and Eq.4), which might be of advantage compared to black-box deep learning approaches. Learning with physical constraints is efficient as it makes most use of data and may avoid trajectories which otherwise be a waste; uncertainty can be yielded by emsembling multiple networks which are trained on subsets of data. We have also applied weights as hyper-parameters to the loss components; weighting the loss components changes the landscape of the loss function. With the log transform trick, least squares for the wave propagation problem enjoys an intrinsic solution compatible with physical intuition, it's fast and usually gives point estimation (the uncertainty in this case comes naturally from physics). All three methods give wave propagation estimates, as well as the resulting distance estimation, with different degrees of accuracy (Fig.21); the inference quality is assessed by the match between the recovered signals and measurements, as well as their physical meanings.

Limitations and future work: more work can be done in the future to improve the existing approaches. First, the environment is assumed to be homogeneous, and wave solution in this work doesn't involve wave inference (e.g. overlapping); we can increase the complexity of the wave propagation equation to account for multi-direction wave propagation, reflection and absorption, with the expectation of achieving better accuracy. Second, scalable Bayesian methods are demanded for large data setting. One can aggregate the training data for Bayesian inference, expecting better generalized capacity and robustness; the same can be achieved with Bayesian online inference when data is collected in a streaming manner. Continuous Bayesian model updates is computationally advantageous compared to least squares which requires retraining on the entire dataset. Further, Bayesian inference results can tell when the probabilistic model is mis-specified, i.e. when the data is inconsistent with the model assumptions. Therefore, apart from model calibration, we can use existing model to detect data heterogeneity and environment changes. For deep spectral learning, accuracy may be improved by introducing complex-valued neural network which utilizes complex derivative (e.g. Wirtinger calculus [50]) to correlate the real and imaginary parts. One can also weight the training instances with prior beliefs about data quality and noise levels to improve modelling. Using symmetric architectures (e.g. auto-encoder) may be advantageous in some specific problem setting (e.g. temporally and spatially reversible processes). Besides, future work can explore sequential learning methods such as deep learning based sequence-to-sequence models [38] for benchmarking.

F. Software packages and code availability

All experiments are implemented in *Python*. Fourier transforms are performed using *SciPy* [33], Bayesian inference is implemented in *PyMC3* [60], neural networks are constructed with *PyTorch* [48]. Random effects are minimized by setting seeds for reproducibility. Main packages are listed in Table.III. All software is open source, and codes available.

Package	**Version**
PyTorch	v2.0.1
Numpy	v1.22.4
Pandas	v1.5.3
PyMC3	v3.11.2
SciPy	v1.10.1

TABLE III: Versions of main packages (Conda environment)

References

[1] Antonio Alguacil, Michaël Bauerheim, Marc C. Jacob, and Stéphane Moreau. Predicting the propagation of acoustic waves using deep convolutional neural networks. *Journal of Sound and Vibration*, 512:116285, 2021.

[2] Shaikhah Alkhadhr and Mohamed Almekkawy. Modeling the wave equation using physics-informed neural networks enhanced with attention to loss weights. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023.

[3] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. In *The Journal of the Acoustical Society of America*, 1979.

[4] Yang Bai, Nakul Garg, and Nirupam Roy. Spidr: ultra-low-power acoustic spatial sensing for micro-robot navigation. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, MobiSys '22, page 99–113, New York, NY, USA, 2022. Association for Computing Machinery.

[5] Manmeet S. Bhabra, Wael H Ali, and Pierre FJ Lermusiaux. High frequency stochastic acoustic wavefront propagation and joint ocean-acoustic inference: The gmm-do wavefront. *The Journal of the Acoustical Society of America*, 2022.

[6] Michael J. Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A. Roch, Sharon Gannot, and Charles-Alban Deledalle. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628, 11 2019.

[7] Stefan Bilbao and Brian Hamilton. Wave-based room acoustics simulation: Explicit/implicit finite volume modeling of viscothermal losses and frequency-dependent boundaries. *Journal of the Audio Engineering Society*, 2017.

[8] D. Botteldoore. Finite-difference time-domain simulation of low-frequency room acoustic problems. *Journal of the Acoustical Society of America*, 1995.

[9] Aki Vehtari; Andrew Gelman; Daniel Simpson; Bob Carpenter;Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved rb for assessing convergence of mcmc. *https://arxiv.org/abs/1903.08008*, 2019.

[10] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *International Conference on 3D Vision (3DV)*, 2017.

[11] Changan Chen, Carl Schissler, Sanchit Garg, Philip Kobernik, Alexander Clegg, Paul Calamia, Dhruv Batra, Philip W Robinson, and Kristen Grauman. Soundspaces 2.0: A simulation platform for visual-acoustic learning. *arXiv*, 2022.

[12] Ziqi Chen, Ning Xiang, and Kirill V. Horoshenkov. Boundary admittance estimation for wave-based acoustic simulations using bayesian inference. *JASA Express Letters*, 2(8):081601, 08 2022.

[13] Ning Chu, Yue Ning, Liang Yu, Qian Huang, and Dazhuan Wu. A fast and robust localization method for low-frequency acoustic source: Variational bayesian inference based on nonsynchronous array measurements. *IEEE Transactions on Instrumentation and Measurement*, 70:1–18, 2021.

[14] Christophe Andrieu; Nando de Freitas; Arnaud Doucet; Michael I. Jordan. An introduction to mcmc for machine learning. *Machine Learning*, 50:5–43, 2003.

[15] Enzo De Sena, Hacıhabiboğlu Hüseyin, Zoran Zoran, Cvetković, and Julius O. Smith. Efficient synthesis of room acoustics via scattering delay networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)*, 2015.

[16] Eduard Deines, Martin Hering-Bertram, Jan Mohring, Jevgenijs Jegorovs, and Hans Hagen. Audio-visual Virtual Reality System for Room Acoustics. In Hans Hagen, editor, *Scientific Visualization: Advanced Concepts*, volume 1 of *Dagstuhl Follow-Ups*, pages 303–320. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2010.

[17] Andrei-Cristian Barbos; François Caron; Jean-François Giovannelli; Arnaud Doucet. Clone mcmc: Parallel high-dimensional gaussian gibbs sampling. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.

[18] Vincent Dumont, Verónica Rodríguez Tribaldos, Jonathan Ajo-Franklin, and Kesheng Wu. Deep learning for surface wave identification in distributed acoustic sensing data. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 1293–1300, 2020.

[19] Scott R. Eliason. *Maximum Likelihood Estimation : Logic and Practice*. Newbury Park: Sage, 1993.

[20] Gerald Enzner. Bayesian inference model for applications of time-varying acoustic system identification. In *2010 18th European Signal Processing Conference*, pages 2126–2130, 2010.

[21] Richard Feynman. *Sound. The wave equation*, volume 1, chapter 47. California Institute of Technology, Pasadena, CA, online edition, 2013 edition, 1963.

[22] Kai Fischer, Martin Simon, Stefan Milz, and Patrick M´ader. StickyLocalization: Robust End-To-End Relocalization on Point Clouds using Graph Neural Networks. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 307–316, 2022.

[23] Thomas Funkhouser, Nicolas Tsingos, Ingrid Carlbom, Gary Elko, Mohan Sondhi, James West, Gopal Pingali, Patrick Min, and Addy Ngan. A beam tracing method for interactive architectural acoustics. *Journal of the Acoustical Society of America*, 2003.

[24] Zoubin Ghahramani and Matthew Beal. Variational inference for bayesian mixtures of factor analysers. In S. Solla, T. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999.

[25] Yuhang He and Andrew Markham. SoundDoA: Learn sound source direction of arrival and semantics from sound raw waveforms. In *Interspeech*, 2022.

[26] Yuhang He, Niki Trigoni, and Andrew Markham. SoundDet: Polyphonic moving sound event detection and localization from raw waveform. In *International Conference on Machine Learning (ICML)*, 2021.

[27] Murray Hodgson and Eva-Marie Nosal. Experimental evaluation of radiosity for room sound-field prediction. *Journal of the Acoustical Society of America*, 2006.

[28] Rémi Bardenet; Arnaud Doucet; Chris Holmes. On markov chain monte carlo methods for tall data. *Journal of Machine Learning Research*, 18:1–43, 2017.

[29] Zoubin Ghahramani Hong Ge, Kai Xu. Turing: a language for flexible probabilistic inference. *Proceedings of Machine Learning Research*, 84:1682–1690, 2018.

[30] Yongchao Huang. *Determining material properties using impact wave propagation method*. Oxford DPhil thesis, 2017.

[31] M.J. Humara, W.H. Ali, A. Charous, M. Bhabra, and P.F.J. Lermusiaux. Stochastic acoustic ray tracing with dynamically orthogonal differential equations. In *OCEANS 2022, Hampton Roads*, pages 1–10, 2022.

[32] Yasuhide Hyodo, Chihiro Sugai, Junya Suzuki, Masafumi Takahashi, Masahiko Koizumi, Asako Tomura, Yuki Mitsufuji, and Yota Komoriya. Psychophysiological effect of immersive spatial audio experience enhanced using sound field synthesis. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2021.

[33] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001.

[34] Yun Yang; Martin J. Wainwright; Michael I. Jordan. On the computational complexity of high-dimensional bayesian variable selection. *Ann. Statist.*, 44:2497–2532, 2016.

[35] Adar Kahana, Eli Turkel, Shai Dekel, and Dan Givoli. Obstacle segmentation based on the wave equation and deep learning. *Journal of Computational Physics*, 413:109458, 2020.

[36] Mendel Kleiner, Bengtinge Dalenbäck, and Peter Svensson. Auralization-an overview. *Journal of the Audio Engineering Society*, 1993.

[37] Roman Kuc and M. W. Siegel. Physically based simulation model for acoustic sensor robot navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(6):766–778, 1987.

[38] Vitaly Kuznetsov and Zelda E. Mariet. Foundations of sequence-to-sequence modeling for time series. In *International Conference on Artificial Intelligence and Statistics*, 2018.

[39] Andrew Luo, Yilun Du, Michael J Tarr, Joshua B Tenenbaum, Antonio Torralba, and Chuang Gan. Learning neural acoustic fields. *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[40] Sagnik Majumder, Changan Chen, Ziad Al-Halah, and Kristen Grauman. Few-shot audio-visual learning of environment acoustics. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[41] Sheri Martinelli. Computational modeling of acoustic wavefronts propagating in an underwater environment with uncertain parameters. *The Journal of the Acoustical Society of America*, 134(5):4114–4114, 11 2013.

[42] Andrew Gelman Matthew D. Hoffman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

[43] David M. Blei; Alp Kucukelbir; Jon D. McAuliffe. Variational inference: A review for statisticians. *https://arxiv.org/pdf/1601.00670.pdf*, 2016.

[44] Frank Michel. *Simulation and Visualization of In- and Outdoor Sound*. doctoralthesis, Technische Universität Kaiserslautern, 2008.

[45] J. Mourjopoulos. On the variation and invertibility of room impulse response functions. *Journal of Sound and Vibration*, 102(2):217–228, 1985.

[46] I. Nevliudov, V. Yevsieiev, S. Maksymova, N. Demska, K. Kolesnyk, and Olha Miliutina. Mobile robot navigation system based on ultrasonic sensors. In *2023 IEEE XXVIII International Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED)*, volume 1, pages 247–251, 2023.

[47] Eva-Marie Nosal, Murray Hodgson, and Ian Ashdown. Investigation of the validity of radiosity for sound-field prediction in cubic rooms. *Journal of the Acoustical Society of America*, 2004.

[48] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[49] Andrzej Pietrzyk. Computer modeling of the sound field in small rooms. *Journal of the Audio Engineering Society*, 1998.

[50] PyTorch. Autograd mechanics, 2022.

[51] Pratyaksh P. Rao and Abhra Roy Chowdhury. Learning to listen and move: An implementation of audio-aware mobile robot navigation in complex indoor environment. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 3699–3705, 2022.

[52] Majid Rasht-Behesht, Christian Huber, Khemraj Shukla, and George Em Karniadakis. Physics-informed neural networks (pinns) for wave propagation and full waveform inversions. *Journal of Geophysical Research: Solid Earth*, 127, 2021.

[53] Anton Ratnarajah, Zhenyu Tang, and Dinesh Manocha. TS-RIR: translated synthetic room impulse responses for speech augmentation. In *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021.

[54] Anton Ratnarajah, Shi-Xiong Zhang Zhang, Meng Yu, Zhenyu Tang, Dinesh Manocha, and Dong Yu. Fast-RIR: Fast neural diffuse room impulse response generator. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[55] Anton Jeran Ratnarajah, Zhenyu Tang, and Dinesh Manocha. IR-GAN: Room impulse response generator for far-field speech recognition. *Interspeech*, 2021.

[56] Alexander Richard, Peter Dodds, and Vamsi Krishna Ithapu. Deep impulse responses: Estimating and parameterizing filters with deep networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[57] Bob Carpenter; Andrew Gelman; Matthew D. Hoffman; Daniel Lee; Ben Goodrich; Michael Betancourt; Marcus Brubaker; Jiqiang Guo; Peter Li; Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76, 2017.

[58] Simon Duane; Anthony D.Kennedy; Brian J.Pendleton; Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

[59] Sebastian Ruder. An overview of gradient descent optimization algorithms. *ArXiv*, abs/1609.04747, 2016.

[60] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.

[61] Lauri Savioja and U. Peter Svensson. Overview of geometrical room acoustic modeling techniques. *Journal of the Acoustical Society of America*, 2015.

[62] Johannes Schmid. Physics-informed neural networks for solving the helmholtz equation. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 2023.

[63] Stefan Schoder, Étienne Spieser, Hugo Vincent, Christophe Bogey, and Christophe Bailly. Acoustic modeling using the aeroacoustic wave equation based on pierce's operator. *AIAA Journal*, 2023.

[64] Dirk Schröder. *Physically based real-time auralization of interactive virtual environments*. PhD thesis, RWTH Aachen University, 2011.

[65] Keegan Yi Hang Sim, Yijia Chen, Yuxuan Wan, and Kevin Chau. Advanced automobile crash detection by acoustic methods. *Proceedings of Meetings on Acoustics*, 39(1):055001, 01 2020.

[66] Asbjørn Krokstad; S. Strøm; Svein Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 1968.

[67] M.J. Box; D. Davies; W.H. Swann. *Non-Linear optimisation Techniques*. Oliver & Boyd, 1969.

[68] Z. Tang, L. Chen, B. Wu, D. Yu, and D. Manocha. Improving reverberant speech training using diffuse acoustic simulation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[69] Zhenyu Tang, Rohith Aralikatti, Anton Jeran Ratnarajah, and Dinesh Manocha. Gwa: A large high-quality acoustic dataset for audio processing. In *ACM SIGGRAPH 2022 Conference Proceedings*, SIGGRAPH '22, New York, NY, USA, 2022. Association for Computing Machinery.

[70] Oleksandr Terletskyi, Petro Venherskyi, Valeriy Trushevskyi, and Ostap Hrytsyshyn. Development of a ray tracing framework for simulating acoustic waves propagation enhanced by neural networks. In *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*, pages 124–126, 2023.

[71] Mehmet Özgür Türkoğlu, Eric Brachmann, Konrad Schindler, Gabriel Brostow, and Áron Monszpart. Visual Camera Re-Localization Using Graph Neural Networks and Relative Pose Supervision. In *International Conference on 3D Vision (3DV)*. IEEE, 2021.

[72] Jorge Vargas, Suleiman Alsweiss, Onur Toker, Rahul Razdan, and Joshua Santos. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*, 21(16), 2021.

[73] Charles Verron, Mitsuko Aramaki, Richard Kronland-Martinet, and Grégory Pallone. A 3-d immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech, and Language Processing (TASLP)*, 2010.

[74] Enhao Wang, Dewang Chen, Tianqi Fu, and Lei Ma. A Robot Relocalization Method Based on Laser and Visual Features. In *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)*, 2022.

[75] Hao Wang, Jian Li, Linfeng Wang, Lin Liang, Zhoumo Zeng, and Yang Liu. On acoustic fields of complex scatters based on physics-informed neural networks. *Ultrasonics*, 128:106872, 2023.

[76] D.V. Widder. *The Heat Equation (Pure and Applied Mathematics)*. Academic Press Inc, 1975.

[77] Evdoxia Taka; Sebastian Stein; John H. Williamson. Increasing interpretability of bayesian probabilistic programming models through interactive representations. *Front. Comput. Sci.*, 2, 2020.

[78] Christian P Robert; Victor Elvira; Nick Tawn; Changye Wu. Accelerating mcmc algorithms. *WIREs Computational Statistics*, 10:e1435, 2018.

[79] Yanqi Wu, Hossein S. Aghamiry, Stéphane Operto, and Jianwei Ma. Helmholtz equation solution in non-smooth media by physics-informed neural network with incorporating quadratic terms and a perfectly matching layer condition. *GEOPHYSICS*, 2023.

[80] Ning Xiang. Model-based bayesian analysis in acoustics—a tutoriala). *The Journal of the Acoustical Society of America*, 148(2):1101–1120, 08 2020.

[81] Ning Xiang and Christopher Landschoot. Bayesian inference for acoustic direction of arrival analysis using spherical harmonics. *Entropy*, 21(6), 2019.

[82] Zhipeng Yang, Xinping Gui, Ju Ming, and Guanghui Hu. Bayesian approach to inverse time-harmonic acoustic obstacle scattering with phaseless data generated by point source waves. *Computer Methods in Applied Mechanics and Engineering*, 386:114073, 2021.