JASA ARTICLE

Sound absorption estimation of finite porous samples with deep residual learning[a)]

Elias Zea,[1,b)] (iD) Eric Brandão,[2] (iD) Mélanie Nolan,[3] Jacques Cuenca,[4] Joakim Andén,[5] and U. Peter Svensson[6]

[1]*The Marcus Wallenberg Laboratory for Sound and Vibration Research, Department of Engineering Mechanics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

[2]*Acoustical Engineering & Civil Engineering Graduate Program, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, 97050-140, Brazil*

[3]*School of Architecture, Rensselaer Polytechnic Institute, Troy, New York 12180, USA*

[4]*Siemens Industry Software, Interleuvenlaan 68, BE-3001 Leuven, Belgium*

[5]*Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

[6]*Department of Electronic Systems, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

ABSTRACT:

This work proposes a method to predict the sound absorption coefficient of finite porous absorbers using a residual neural network and a single-layer microphone array. The goal is to mitigate the discrepancies between predicted and measured data due to the finite-size effect for a wide range of rectangular absorbers with varying dimensions and flow resistivity and for various source-receiver locations. Data for training, validation, and testing are generated with a boundary element model consisting of a baffled porous layer on a rigid backing using the Delany–Bazley–Miki model. In effect, the network learns relevant features from the array pressure amplitude to predict the sound absorption as if the porous material were infinite. The method's performance is quantified with the error between the predicted and theoretical sound absorption coefficients and compared with the two-microphone method. For array distances close to the porous sample, the proposed method performs at least as well as the two-microphone method and significantly better than it for frequencies below 400 Hz and small absorber sizes (e.g., 20×20 cm^2). The significance of the study lies in the possibility of measuring sound absorption on-site in the presence of strong edge diffraction. © *2023 Acoustical Society of America.* https://doi.org/10.1121/10.0021333

I. INTRODUCTION

Free-field or *in situ* methods of measuring the sound absorption of acoustic materials aim at inferring the absorption properties (e.g., surface impedance and reflection and absorption coefficients) from measurements of the sound field in the vicinity of the measurement sample (Alkmim *et al.*, 2021; Allard and Sieben, 1985; Brandão *et al.*, 2015; Li and Hodgson, 1997; Mommertz, 1995). Although not standardized, the attractiveness of these methods lies in the fact that they provide angle-dependent absorption data [which cannot be measured with standardized methods (ISO, 1998, 2003)] and are applicable for materials mounted for their intended application. *In situ* methods rely on a mathematical model of the sound field above the material and generally assume the sample is much larger than the wavelength, i.e., the reflected field is predominantly specular. In the low- to the mid-frequency range, such an assumption is violated due to the sample's finiteness, and

the so-called "edge-diffraction effect" (or "finite-size effect") leads to bias errors between prediction and experimental data (de Bruijn, 1973; Thomasson, 1980). This bias is negligible at high frequencies with a much smaller wavelength than the sample size.

Considerable effort has been put into the problem of measuring the sound absorption of finite-size samples *in situ*. In particular, previous studies have compared experimental data with boundary element method (BEM) simulations to describe and account for the edge-diffraction effect (see, e.g., Brandão *et al.*, 2012; Hirosawa *et al.*, 2009; Luo *et al.*, 2020; Otsuru *et al.*, 2009). Recent studies have attempted to characterize the edge-diffraction effect experimentally (Ottink *et al.*, 2016; Richard *et al.*, 2017) and numerically (Brandão and Fernandez-Grande, 2022). These studies rely on a model of the sound field/absorber (or hypotheses on the measurement scenario). Fewer studies have used data to learn the edge-diffraction effect (Müller-Giebeler and Vorländer, 2021; Zea *et al.*, 2021).

Machine learning (ML) has been successfully applied in several acoustical problems (Bianco *et al.*, 2019), as demonstrated in the recent special issue of *The Journal of The Acoustical Society of America* (Michalopoulou *et al.*, 2021).

Related works have studied the prediction of room-acoustical parameters (Götz *et al.*, 2022; Yu and Kleijn, 2021) and mean room sound absorption (Foy *et al.*, 2021) from room impulse responses, as well as the prediction of porous material parameters from ultrasonic images (Lähivaara *et al.*, 2018), and transport parameters of fibrous materials from x-ray images (Jeon *et al.*, 2021). This paper proposes an ML-based approach to predict the sound absorption coefficient of finite porous samples. The strength of the proposed method is that it removes the finite-size effect without additional hypotheses on the sound field or the absorber.

This study uses a residual neural network to map the sound pressure field above the finite absorber to its sound absorption coefficient. The main goal is to suppress the edge-diffraction effect from the measurement by learning features from the pressure field that are relevant to predict the sound absorption coefficient as if the sample were infinite. The network consists of a simplified ResNet architecture (He *et al.*, 2016a,b), which allows testing well-studied ML models with a reasonable input size (here, the sound pressure spectra in 12×12 microphone positions). The model is trained and tested with data generated with a parameterized BEM model. The BEM simulation predicts the sound pressure emitted by a monopole above baffled, rectangular porous samples with uniform surface impedance over the sample. The parameters of the BEM model include the size and absorption characteristics of the sample, the sound source, and the receiver's distances and orientations relative to the sample. Furthermore, the trained model sensitivity is investigated regarding Gaussian noise and data outside the training set. The performance of the proposed network is compared with the two-microphone method as a benchmark (Allard and Sieben, 1985).

The outline of the paper is as follows. Section II describes the generation of the training, validation, and testing datasets via BEM. Section III describes the proposed method. Section IV shows the results of training and testing the proposed method, including a comparison with the two-microphone method, and Sec. V includes a sensitivity analysis. Conclusions are drawn in Sec. VI.

II. SOUND FIELD MODEL

A. Boundary element model

A simplified BEM model is considered, shown in Fig. 1, in which a point source is located at $\mathbf{r}_q = (x_q, y_q, z_q)$ and a receiver is located at $\mathbf{r} = (x, y, z)$. The receiver positions correspond to the microphone array, used as data for training and testing, and the two-microphone method, used as benchmark data. A finite rectangular absorber sample of dimensions $L_x \times L_y \times d$ is flush mounted to an infinite hard baffle at $z = 0$. The boundary condition in the upper half-space is free-field. The time-harmonic factor $e^{j\omega t}$ is omitted throughout.

The sound pressure, $p(\mathbf{r})$, can be written as the Helmholtz/Huygens integral
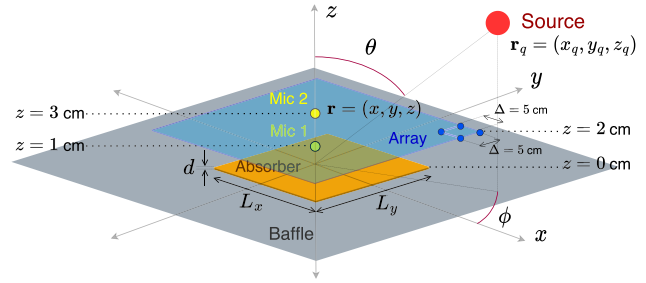

FIG. 1. (Color online) Schematic of the BEM simulation. The point source (red) at $\mathbf{r}_q$ excites the sound field at incidence angle (ϕ, θ). The absorber (orange) is baffled at $z = 0$ cm, with dimensions $L_x \times L_y \times d$. The microphone array (blue) is centered at $(0, 0, 2)$ cm and has 12×12 microphones spaced by $\Delta = 5$ cm. The receivers of the two-microphone method (yellow) are placed at $(0, 0, 1)$ cm and $(0, 0, 3)$ cm. In this example, the receiver $\mathbf{r}$ for the BEM calculation is Mic 2.

$$
c(\mathbf{r})p(\mathbf{r}) = \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_q\|}}{\|\mathbf{r}-\mathbf{r}_q\|} + \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_q'\|}}{\|\mathbf{r}-\mathbf{r}_q'\|} \\
- \frac{jk_0}{Z_s}\int_S p(\mathbf{r}_s) \times \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_s\|}}{4\pi\|\mathbf{r}-\mathbf{r}_s\|}\,dS,
$$

with

$$
c(\mathbf{r}) = \begin{cases} 0.5, & z = 0, \\ 1, & z \geq 0, \end{cases} \tag{1}
$$

where $k_0 = 2\pi f/c_0$ is the acoustic wavenumber, f is the frequency, c_0 the speed of sound in air, $j = \sqrt{-1}$, $\|\cdot\| = \sqrt{\sum|\cdot|^2}$ is the ℓ_2-norm of the vector, S is the surface boundary of the absorber, and Z_s is the surface impedance of the finite sample. The receiver point $\mathbf{r}$ can be located above or on the sample surface, and $\mathbf{r}_s$ is any point at the surface of the finite sample.

The first term on the right-hand side of Eq. (1) is the Green's function between the sound source and the receiver. The second term is the Green's function between the image sound source, at $\mathbf{r}_q' = (x_q, y_q, -z_q)$, and the receiver. Together they form the sound field with only the infinite, hard baffle as if the sample itself were absent. The last term carries information about the absorption and diffraction of the finite absorber, formulated as an integral over the finite absorber area. Discretizing the surface of the sample into N square elements and assuming that $p(\mathbf{r}_s)$ is constant over each element, Eq. (1) becomes

$$
c(\mathbf{r})p(\mathbf{r}) = \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_q\|}}{\|\mathbf{r}-\mathbf{r}_q\|} + \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_q'\|}}{\|\mathbf{r}-\mathbf{r}_q'\|} \\
- \frac{jk_0}{Z_s}\sum_{n=1}^{N} p(\mathbf{r}_{s_n}) \times \int_{S_n} \frac{e^{-jk_0\|\mathbf{r}-\mathbf{r}_{s_n}\|}}{4\pi\|\mathbf{r}-\mathbf{r}_{s_n}\|}\,dS_n, \tag{2}
$$

where the collocation method is used by placing $\mathbf{r} = \mathbf{r}_{s_n}$ at $n = 1, 2, \ldots, N$ on the surface of the sample, with $c(\mathbf{r}) = 0.5$. Thus, a system of equations is formed, and the surface pressure, $p(\mathbf{r}_{s_n})$, at each element, can be found.

2322   J. Acoust. Soc. Am. **154** (4), October 2023

Zea *et al.*

Once the surface pressure is known, it can be re-inserted into Eq. (2) to calculate the pressure at any receiver point for $z > 0$, with $c(\mathbf{r}) = 1.0$.

The absorption term is modeled by a prescribed surface impedance, Z_s, constant across the finite sample's surface. The characteristic impedance and wavenumber for the porous material are computed with the Delany–Bazley–Miki (DBM) model (Miki, 1990) as

$$Z_p(f) = \rho_0 c_0 \big(1 + 5.50\varsigma^{-0.632} - j8.43\varsigma^{-0.632}\big), \qquad (3)$$

$$k_p(f) = k_0 \big(1 + 7.81\varsigma^{-0.618} - j11.41\varsigma^{-0.618}\big), \qquad (4)$$

where $\varsigma \equiv \varsigma(f, \sigma) = 10^3 f/\sigma$, ρ_0 is the air density, and σ is the flow resistivity of the sample.

In the following, let us assume a hard-backed porous layer of thickness d and a uniform surface impedance over the sample. Thus, for a given angle of incidence θ, the surface impedance and absorption spectra can be calculated according to Allard and Atalla (2009), as

$$Z_s(f) = -j \frac{Z_p(f)}{\cos(\theta_t)} \cot\big[k_p(f)d\cos(\theta_t)\big], \qquad (5)$$

$$\alpha(f) = 1 - \left| \frac{Z_s(f)\cos(\theta) - \rho_0 c_0}{Z_s(f)\cos(\theta) + \rho_0 c_0} \right|^2, \qquad (6)$$

where $\theta_t = \arcsin[(k_0/k_p)\sin(\theta)]$.

The values of $\alpha(f)$ in Eq. (6) constitute the reference absorption spectra, which are used as the labels (outputs) of the ML model during training. Thus, the training set comprises sound fields of flushed porous absorbers following the DBM model. This does not guarantee that an ML approach trained with such data generalizes well to other material models.

An array of 12×12 receivers is considered in this study. The array aperture is 0.6×0.6 m, and the receivers are spaced by $\Delta = 5$ cm and placed at a distance of 2 cm above the sample surface. The receiver spacing prevents spatial aliasing up to the frequency $f_{max} = c_0/(2\Delta)$.[1] The highest simulated frequency is 2 kHz, with six boundary elements per wavelength. The integrals in Eq. (2) are calculated with linear interpolation and Gauss–Legendre quadrature with 36 points on each element (Atalla and Sgard, 2015; Wu, 2000). With the implemented configuration, the quadrature points do not coincide with the element center, which avoids singularities. Experimental validation for such BEM simulations using single point estimates can be found in Brandão et al. (2012).

In the present paper, the values of d and σ used in the BEM model are drawn from a log-uniform distribution, so the sound absorption function is sampled smoothly. For the range of material parameters found in practical scenarios, the sound absorption function of the DBM material model can have sharp changes for low values of thickness d and flow resistivity σ, especially at frequencies above 1600 Hz. If the samples in the training set correspond to values of d
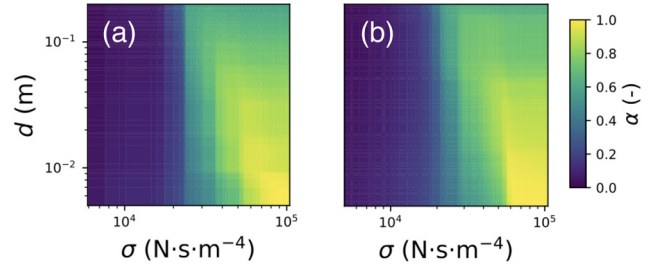

FIG. 2. (Color online) Influence of the sampling distribution of absorber thickness d and flow resistivity σ on the sound absorption coefficient function of the Delany–Bazley–Miki model (Miki, 1990) at 500 Hz. (a) Realization with uniform sampling. (b) Realization with log-uniform sampling.

and σ drawn from a uniform distribution, there is a risk that the absorption function is not sufficiently sampled when d and σ are small. This can cause difficulties for the network to generalize to cases in that (d, σ) region, as concluded from (Zea et al., 2021). A comparison between realizations of the two distributions is shown in Fig. 2. It can be observed that the sound absorption function is sampled more smoothly using the log-uniform distribution.

B. Datasets

A total of 330 000 sound field simulations above finite absorbers[2] have been computed with the BEM model from Sec. II A. Out of these, 300 000 simulations correspond to the training (T) and validation (V) datasets, 15 000 to the interpolation (I) dataset, and the remaining 15 000 to the extrapolation (X) dataset. The I and X datasets are test sets; thus, they assess the network's generalizability. A summary of the BEM model parameters is found in Table I. All datasets (T, V, I, and X) share the same sampling distributions of the sample size, thickness, flow resistivity, source azimuth/ elevation angles, number of microphones, and microphone spacing. Each sound field simulation is run at the 14 center frequencies of 1/3 octave bands from 100 Hz to 2000 Hz.

TABLE I. Parameters of the BEM model used to generate the training (T), validation (V), interpolation (I), and extrapolation (X) datasets.

Parameter	Value	Sampling	Datasets
Sample side L_x [cm]	[20, 100]	Uniform	TVIX
Sample side L_y [cm]	[20, 100]	Uniform	TVIX
Thickness d [mm]	[5, 200]	Log-uniform	TVIX
Flow res. σ [kNs/m^4]	[5, 100]	Log-uniform	TVIX
Source distance $\|\mathbf{r}_q\|$ [m]	[1.2, 1.8]	Uniform	TVI
	[0.7, 2.3]	Uniform	X
Source azimuth ϕ [deg.]	[0, 360)	Uniform	TVIX
Source elevation θ [deg.]	[0, 80]	Uniform	TVIX
No. microphones [—]	12×12	Grid	TVIX
Mic spacing Δ [cm]	5	—	TVIX
Array distance z [cm]	2	–	TVI
	[2, 20]	Uniform	X
Array rotation [deg.]	0	—	TVI
	[−90, 90]	Uniform	X

J. Acoust. Soc. Am. **154** (4), October 2023

Zea et al.   2323

1. Training and validation datasets

The training and validation datasets have 240 000 and 60 000 instances, respectively, and are obtained using the BEM parameters in Table I following the datasets T and V. The BEM computations are done in two main steps: (i) assembling the BEM matrix and (ii) computing the surface impedance, the absorption coefficient, and the pressure field. The first step is independent of the material properties and the source and microphone array configuration. This significantly accelerates the computations by choosing 3000 base cases with fixed (but random) absorber sizes and pre-assembling their corresponding BEM matrices. Then, each base case is run for 100 random combinations of the rest of the BEM parameters.

2. Interpolation dataset

The interpolation test set contains 15 000 instances generated with the same BEM model parameters as those in the training set (following the dataset I in Table I). The only difference is that the absorber sizes do not correspond to the 3000 base cases from the training set but rather an additional 15 000 random sizes and parameters. This dataset assesses the models' performance against absorber sizes not seen in the training set.

3. Extrapolation dataset

As with the interpolation set, the extrapolation set contains 15 000 additional instances, generated with the BEM model parameters chosen for the training set (following the dataset X in Table I) with three differences:

- Expanded range of source distances: from $\|\mathbf{r}_q\| \in [1.2, 1.8]$ m to $[0.7, 2.3]$ m.
- Varying range of array distances: from $z = 2$ cm to $z \in [2, 20]$ cm.
- Varying range of array rotations: from no rotation to $[-90°, 90°]$ rotations.

This dataset assesses the model's performance against unseen absorber sizes, source distances, array distances, and array rotations.

III. CONVOLUTIONAL NEURAL NETWORKS

In supervised learning, neural networks have been widely applied to establish nonlinear mappings between acoustical quantities. In particular, convolutional neural networks (CNNs) (LeCun *et al.*, 1989) aim at learning local features from the data at multiple resolution levels and exploit these features for classification or regression (LeCun *et al.*, 2015). In contrast to fully connected neural networks (e.g., multi-layer perceptron), CNNs enable shift-invariant processing of high-dimensional data with fewer operations (Liu *et al.*, 2017). The cost reduction is achieved by sharing weights and using shift-invariant filters (Goodfellow *et al.*, 2016). Indeed, it has been shown in the literature that learning shift-invariant features is highly relevant for sound field

reconstruction with microphone arrays (Hahmann *et al.*, 2021).

The building block of a CNN, the convolutional layer, consists of a convolution between input feature maps and a weighted kernel (or filter), followed by an activation function. In mathematical terms, we have

$$\mathbf{g}_j^{(l)} = \varphi \left(\sum_{i=1}^{\mathcal{I}_{l-1}} \mathbf{g}_i^{(l-1)} \circledast \mathbf{w}_{ji}^{(l)} + b_j^{(l)} \right), \quad j = 1, ..., \mathcal{J}_l, \quad (7)$$

where $\mathbf{g}_i^{(l-1)}$ is the ith input feature map to the lth layer, with spatial dimensions $W_{l-1} \times H_{l-1}$. The symbol $\circledast$ denotes two-dimensional convolution, and $b_j^{(l)}$ contains the bias added in the jth feature map of the lth layer. The convolution kernel at the lth layer, $\mathbf{w}_{ji}^{(l)}$, has size $U_l \times K_l$ (typically smaller than the spatial dimensions of the input, $W_{l-1} \times H_{l-1}$), and it is applied to the input feature maps $\mathbf{g}_i^{(l-1)}$, $i = 1, ..., \mathcal{I}_{l-1}$. The $\mathcal{J}_l$ output feature maps corresponding to the lth layer are obtained by computing Eq. (7) with the filter kernels $\mathbf{w}_{ji}^{(l)}$, $j = 1, ..., \mathcal{J}_l$. The activation function $\varphi(\cdot)$ makes the map non-linear; an example is the rectified linear unit (ReLU), $\varphi(\cdot) = \max\{\cdot, 0\}$ (Nair and Hinton, 2010).

A. Deep residual learning

The justification for adopting residual learning in this study is to extract multiple features at different resolution levels—relevant to map the edge-diffracted sound field to the sound absorption coefficient while optimizing the number of computations and preventing the vanishing gradient problem (Hochreiter, 1991). Empirical evidence shows that deeper networks perform worse beyond a certain number of layers than their shallower counterparts (He *et al.*, 2016a). He *et al.* introduced residual learning to counteract this issue without sacrificing computational cost and performance. The idea is to introduce residual blocks, which add a "shortcut" (skip connection) from the input to the output of the block. In essence, residual learning allows the network to learn the identity function without adding trainable parameters, so the performance of a $(D + 1)$-depth network is at least as good as that of a D-depth network.

A typical residual block consists of two branches. A skip connection branch entails an identity map, and a feature extraction branch entails a sequence of a point-wise 1×1 convolution,[3] a convolutional layer, and another point-wise convolution (He *et al.*, 2016a). In mathematical terms, the $j = 1, ..., \mathcal{J}_l$ feature maps at the output of the residual block are defined as

$$\mathbf{g}_j^{(l)} = \varphi \left(\sum_{i=1}^{\mathcal{I}_{l-1}} \mathbf{g}_i^{(l-1)} \circledast \mathbf{w}_{ji}^{(l)} + b_j^{(l)} \right) + \mathbf{g}_j^{(l-3)}, \quad (8)$$

where $\mathbf{g}_j^{(l-3)}$ is the jth feature map of the input to the residual block. Note that Eq. (8) assumes that the skip connection and feature extraction branches have the same number of output maps, i.e., $\mathcal{I}_{l-2} = \mathcal{J}_l$. When this is not the case, a
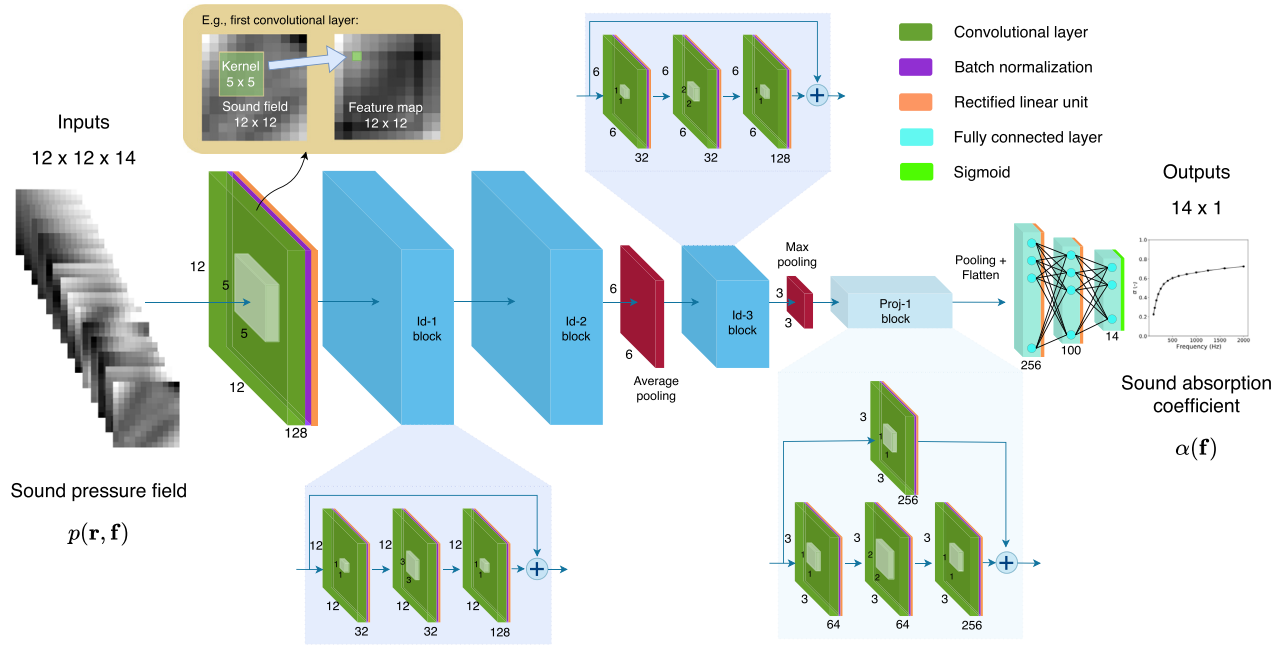
FIG. 3. (Color online) Schematic of the neural network proposed in this paper. Input layer: 12×12 sound pressure field at center frequencies of 1/3 octave bands from 100 to 2000 Hz. Output layer: sound absorption coefficient at said frequencies. Top-left: the convolution between the input pressure field and the kernel yields the $(3, 3)$th element of the feature map.

point-wise convolution is applied to the input feature maps (e.g., Proj -1 block in Fig. 3).

B. Proposed model

The proposed architecture, shown in Fig. 3, consists of a deep CNN with four residual blocks and 17 layers, similar in spirit to the ResNet model (He *et al.*, 2016a). As shown in Fig. 3, three identity residual blocks and one linear projection block are used, with batch normalization to bound the statistical moments of the feature maps across layers (Ioffe and Szegedy, 2015).

As shown on the left of Fig. 3, the inputs contain the magnitude of the frequency-dependent pressure fields in $F = 14$ two-dimensional images, $|\mathbf{p}(\mathbf{r}, \mathbf{f})| \in \mathbb{R}^{12 \times 12 \times 14}$, and the example on the top-left illustrates one of the output feature maps in the first convolutional layer. The right of Fig. 3 shows the network output with the frequency-dependent sound absorption coefficients in a vector, $\boldsymbol{\alpha}(\mathbf{f}) \in \mathbb{R}^{14}$. Essentially, the model is trained to learn the map: $\boldsymbol{\alpha}(\mathbf{f}) = \mathcal{NN}\{|\mathbf{p}(\mathbf{r}, \mathbf{f})|\}$.

Note that a single-layer microphone array is used, in contrast to acoustic holography-based methods, e.g., Hirosawa *et al.* (2009), Nolan (2020), Ottink *et al.* (2016), and Tamura (1990), that use multiple layers to separate incident from reflected fields. The absolute pressure field is also used instead of the complex-valued pressure field. The motivation behind these choices is to reduce the size of the input (number of sensor positions and input channels) and, consequently, the number of trainable parameters and training time of the network. It is worth mentioning that CNNs are discretization-variant methods; thus, their accuracy is generally proportional to the image dimensions used for training.

In the development of this work, it has been observed that models with residual learning perform better than those without [e.g., LeNet-like (Zea *et al.*, 2021)]. Other interesting approaches include residual learning blocks from the Inception (Szegedy *et al.*, 2015), Xception (Chollet, 2017), and Inception-ResNet (Szegedy *et al.*, 2017) models. It has been observed from the data in the present study that differences in performance are marginal across models with different residual blocks. A comparison of the performance across these models is out of the scope of the present paper.

The proposed neural network has 239 274 trainable parameters and is trained with a batch size of $B = 128$. The loss function is the mean-squared error (MSE),

$$\text{MSE} = \frac{1}{B \times F} \sum_{i=1}^{B} \sum_{j=1}^{F} \left(\alpha_{i,j} - \alpha_{i,j}^{\star} \right)^2, \tag{9}$$

where $\alpha_{i,j}$ is the reference sound absorption coefficient at the jth frequency of the ith instance of the batch (see Sec. II B 1); and $\alpha_{i,j}^{\star}$ is the corresponding absorption coefficient predicted with the proposed model. The MSE is chosen in this study as it is a customary choice in statistical regression problems (Goodfellow *et al.*, 2016) and indicates the variance between the predicted and reference absorption coefficients.

The network weights are optimized with the training set using the Adam optimizer (Kingma and Ba, 2015) as its performance is superior to other optimizers tested in this work. The initial learning rate is set to 10^{-3}, and no learning schedule is used. For validation and testing with unseen data, the metrics computed are the MSE and the mean absolute error (MAE),

J. Acoust. Soc. Am. **154** (4), October 2023
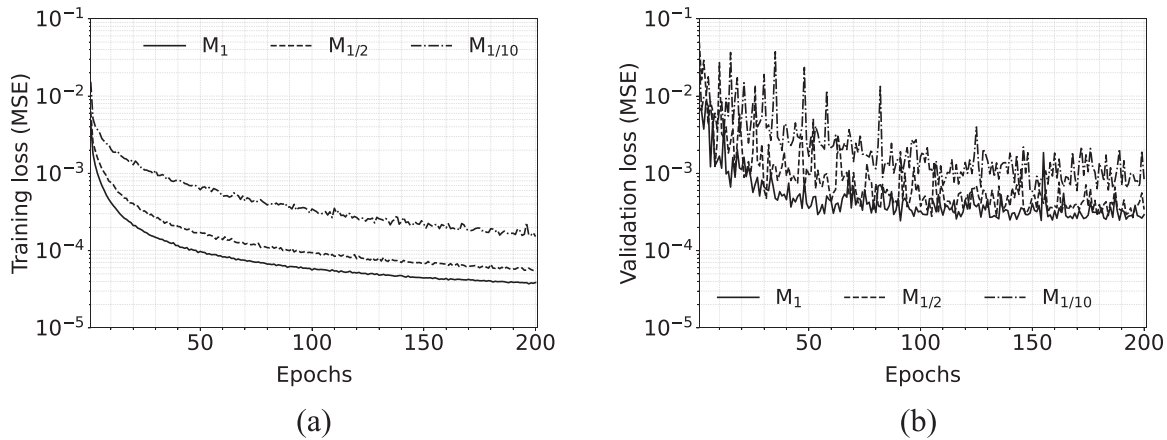
Zea *et al.* 2325

FIG. 4. Mean-squared error (MSE) versus epochs for varying sizes of the training set (100%, 50%, and 10%) with the proposed method. (a) Training loss. (b) Validation loss.

$$\text{MAE} = \frac{1}{B \times F} \sum_{i=1}^{B} \sum_{j=1}^{F} |\alpha_{i,j} - \alpha_{i,j}^{\star}|. \qquad (10)$$

IV. RESULTS

The network's development, training, and testing are done with the KERAS deep learning library (Chollet, 2015). The code is run in Google Colab, setting 32 GB RAM and GPU hardware acceleration (NVIDIA T4). With these settings, it takes approximately 26 s per epoch to train with the training dataset of 240 000 absorbers. The network's weights are initialized with KERAS's default uniform initializer XAVIER, and training is stopped at 200 epochs as the validation loss converges at this point.

A. Training and validation results

To assess the impact of the size of the training set on the learning, the proposed network is trained with 100%, 50%, and 10% of the training set, resulting in models M_1, $M_{1/2}$, and $M_{1/10}$, respectively. Figure 4 shows the corresponding learning curves.

As shown in Fig. 4(a), the training loss converges more slowly and to a larger value as the training dataset becomes smaller. It can be seen that the three models reach a training loss of 2×10^{-4} at approximately 20, 40, and 160 epochs; in close correspondence with the dataset sizes $(1 : 2 : 10)$. A similar observation can be drawn from the validation loss in Fig. 4(b). Nevertheless, the differences between the models M_1 and $M_{1/2}$ are relatively marginal, suggesting that the training performance might reach a limit of no substantial improvement beyond half the training dataset. The validation loss suggests that, after 200 epochs, the network will likely learn noise from the data, eventually leading to an overfitted model.

The dataset size also influences the training speed: fewer batches per epoch imply quicker training. Model $M_{1/2}$ takes about 13 s per epoch, while model $M_{1/10}$ takes about 3 s per epoch using the GPU resources in Google Colab.

Recall that model M_1 takes 26 s per epoch with the complete training set.

B. Performance with interpolation and extrapolation sets

Table II shows the performance metrics for the proposed model and the two test datasets. The extrapolation test set is more challenging for the network since it represents parameters drawn from a wider distribution than those used for training (see Secs. II B 2 and II B 3). This is likely why the MSE values are two orders of magnitude larger with the extrapolation test set than the interpolation test set.

When tested with the interpolation dataset, it can be seen that the MSE of the models M_1 and $M_{1/2}$ are marginally close, within a 5×10^{-5} difference. The MSE of the model $M_{1/10}$ differs by approximately 5×10^{-4} from those of M_1 and $M_{1/2}$. A similar observation can be drawn from the MAE of $M_{1/10}$, which is about twice that of the M_1 and $M_{1/2}$. Regarding the extrapolation dataset, the MSE and MAE values for the three models are rather uniform, within 5×10^{-2} and 2×10^{-1}, respectively, and do not vary as much as with the interpolation dataset. It can be argued that the unseen data in the extrapolation dataset, in contrast to the seen data in the training dataset (cf. Table I), causes the error metrics in Table II to be similar despite the different size of the training sets. Another result that can be pointed

TABLE II. Mean-squared error (MSE) and mean-absolute error (MAE) for the proposed model trained with 100% (M_1), 50% ($M_{1/2}$), and 10% ($M_{1/10}$) of the training set, and tested with the interpolation (I) and extrapolation (X) datasets. The MSE values at the 200th epoch of the training (T) and validation (V) loss are shown for comparison.

Dataset	Metric	M_1	$M_{1/2}$	$M_{1/10}$
T	MSE	3.9×10^{-5}	5.6×10^{-5}	1.5×10^{-4}
V	MSE	2.9×10^{-4}	3.4×10^{-4}	8.4×10^{-4}
I	MSE	2.8×10^{-4}	3.3×10^{-4}	8.3×10^{-4}
	MAE	1×10^{-2}	1.2×10^{-2}	1.9×10^{-2}
X	MSE	4.9×10^{-2}	5.3×10^{-2}	5.6×10^{-2}
	MAE	1.6×10^{-1}	1.7×10^{-1}	1.8×10^{-1}

out here is that the models M_1 and $M_{1/2}$ have a relatively similar statistical performance, which suggests that far less training data is sufficient.

C. Prediction of sound absorption coefficients

This section compares the proposed method's predictive performance against the two-microphone method for a few absorbers from the interpolation and extrapolation datasets. As shown in Fig. 1 and mentioned in the Appendix, the positions of the receivers in the two-microphone method are kept fixed for all the cases, at 1 and 3 cm above the center of the sample.

Figure 5 compares the sound absorption coefficients for one sample from the interpolation dataset. A 3D view of the BEM model and the pressure field inputs are also shown in Figs. 5(a) and 5(b), respectively. The sound absorption curves in Fig. 5(c) are obtained with the analytical model according to (Allard and Atalla, 2009) [see Eq. (6)]; with the two-microphone method (Allard and Sieben, 1985); and with the proposed network trained with 100% of the training dataset (M_1). An excellent agreement can be observed between the sound absorption coefficients predicted by the network and the analytical model, indicating that the proposed method can effectively mitigate the finite-size effect. This is also expected from the results in Table II. However, more significant discrepancies are observed with the two-microphone method, attributed to the edge diffraction effects captured by the receivers.

Figure 6 shows the corresponding results for two absorbers from the extrapolation dataset. The sound absorption coefficient of the first case, shown in Fig. 6(c), is underestimated by the two-microphone method. Indeed, this is a case where strong finite-size effects are present at the microphones due to the oblique incidence angle of the sound source, which also challenges the plane wave assumption of the two-microphone method. In addition, a tendency of the two-microphone method to underestimate the sound absorption of highly reflecting samples has been observed, e.g., Alkmim et al. (2021). In contrast, the proposed method demonstrates an excellent agreement with the analytical

model. This can be attributed to the array distance being 4.4 cm, which is not far from the distance of 2 cm chosen in the training set. The network generalizes well to unseen source distance and array rotations for this particular absorber.

The scenario of the second absorber, shown in the bottom row of Fig. 6, is particularly challenging for the proposed method because the array distance from the sample is 17.7 cm, almost nine times further than the distance chosen in the training set. Additionally, the source distance, $\|\mathbf{r}_q\| = 1.1$ m, lies outside the range of the distances chosen in the training set (between 1.2 and 1.8 m). This likely causes the proposed method to overestimate the sound absorption obtained with the analytical model at all frequencies, implying that the network cannot generalize well for this measurement scenario. The sound absorption coefficients predicted by the two-microphone method agree reasonably well with those predicted by the analytical model. Nevertheless, more substantial variations in the performance could arise if the two microphones were placed further away from the sample (Brandão et al., 2012). Section V B includes a detailed analysis of the sensitivity of the proposed method against unseen data, which further supports these results.

V. SENSITIVITY ANALYSIS

This section examines the sensitivity of the proposed model to measurement noise and unseen data. The model chosen for the analysis is M_1.

A. Measurement noise

To examine the predictive performance of the proposed model against measurement noise, the sound pressure field $\mathbf{p}(f) \in \mathbb{C}^M$ at the M microphone positions of the array is artificially contaminated with noise at frequency f,

$$\mathbf{p}_e(f) = \mathbf{p}(f) + \mathbf{e}(f), \tag{11}$$

where the noise vector $\mathbf{e}(f) \in \mathbb{C}^M$ is complex Gaussian with zero mean and variance $E(f)$, according to a given signal-to-noise ratio (SNR). The SNR is defined as
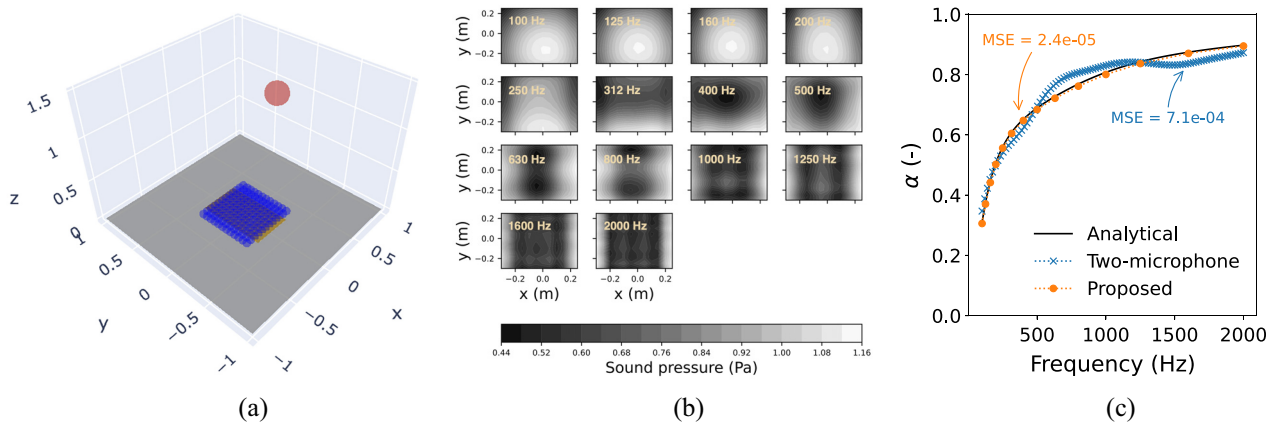

(a) (b) (c)

FIG. 5. (Color online) (a) 3D view of the BEM model, (b) input pressure fields, and (c) sound absorption coefficient spectrum of an absorber from the interpolation set: $L_x = 43$ cm, $L_y = 73$ cm, $\sigma = 38.9$ kN s/m^4, $d = 10.3$ cm, $\|\mathbf{r}_q\| = 1.5$ m, $\theta = 10.9°$, $\phi = 297.6°$, and $z = 2$ cm.
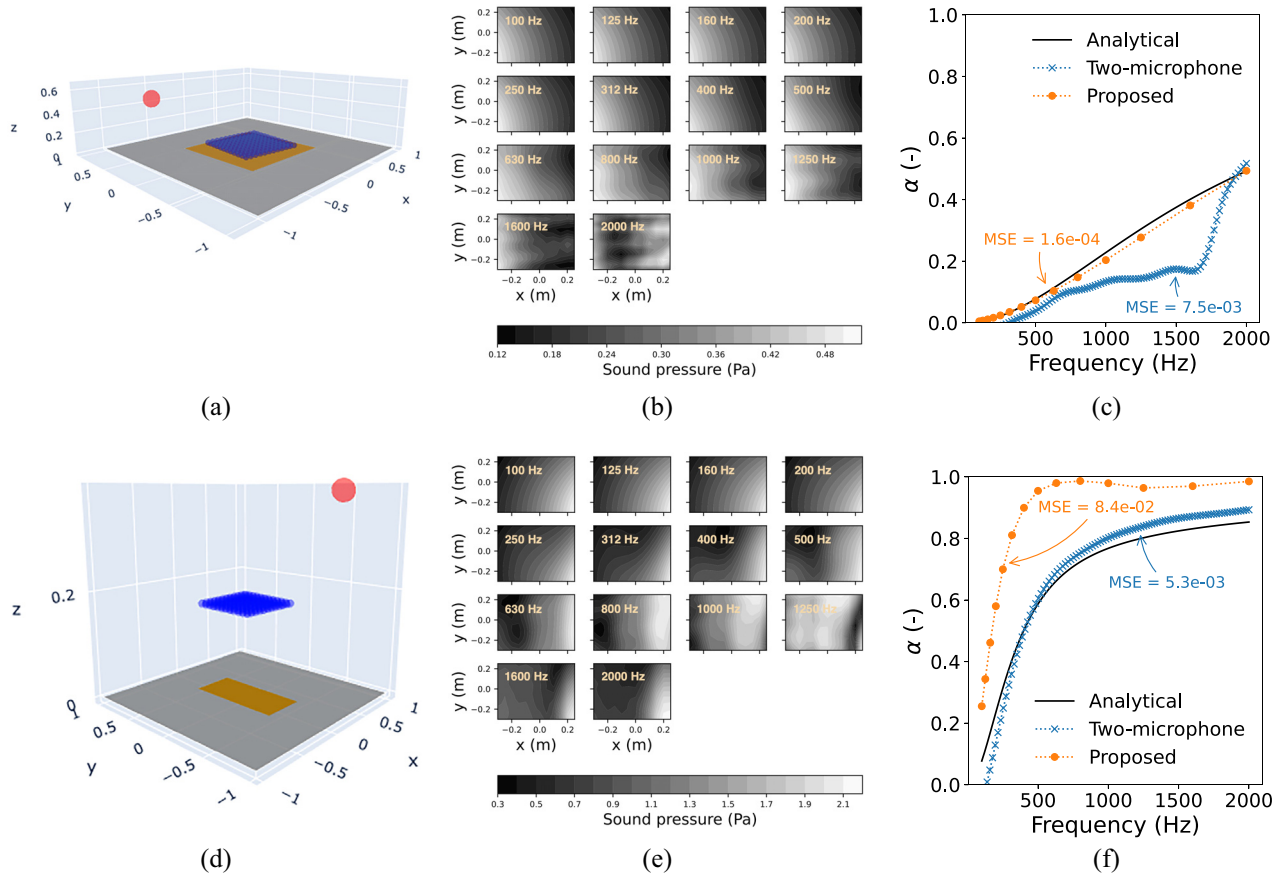
FIG. 6. (Color online) (a), (d) 3D view of the BEM model, (b), (e) input pressure fields, and (c), (f) sound absorption coefficients of two absorbers from the extrapolation set. (Top row) $L_x = 92$ cm, $L_y = 90.7$ cm, $\sigma = 78.7$ kN s/m^4, $d = 7$ mm, $\|\mathbf{r}_q\| = 1.38$ m, $\theta = 61.6°$, $\phi = 188.8°$, $z = 4.4$ cm, and the array is rotated by $7.1°$. (Bottom row) $L_x = 43.2$ cm, $L_y = 95.1$ cm, $\sigma = 47.3$ kN s/m^4, $d = 2.8$ cm, $\|\mathbf{r}_q\| = 1.1$ m, $\theta = 71.1°$, $\phi = 338.8°$, $z = 17.7$ cm, and the array is rotated by $1.7°$.

$$\mathrm{SNR} = 10 \log_{10} \frac{P(f)}{E(f)}, \tag{12}$$

where $P(f) = \|\mathbf{p}(f)\|^2/M$ is the array signal power at frequency f. The noise variance is thus obtained via Eq. (12) as $E(f) = P(f) \times 10^{-\mathrm{SNR}/10}$. Note that the absolute sound pressure used as input for the CNN model is computed after adding the noise, i.e., taking the absolute value of Eq. (11).

The noise is added in a frequency-dependent manner, i.e., different random realizations per frequency, and in a sensor-independent way using the signal power of the pressure field per frequency. The SNR values range from 0 to 30 dB, with 5 dB increments.

Figure 7 summarizes the performance metrics (MSE and MAE) for both interpolation and extrapolation test sets and the various SNR values. The MSE with the interpolation dataset deteriorates much more strongly with decreasing SNR than with the extrapolation dataset. This suggests that noise in the interpolation set dominates the differences in MSE across SNR values. As will later be shown in Sec. V B, the proposed method is agnostic to unseen array distance up to around 4–5 cm, and beyond this distance, its performance deteriorates regardless of the SNR. In contrast, when the network is exposed to unseen information, as with the

extrapolation set, the MSE does not change as much across SNR values. A similar observation can be made with the MAE metric.

B. Noisy measurements for extrapolation set

This last analysis examines the performance of the proposed method in the presence of measurement noise, emphasizing unseen data during the training phase and a
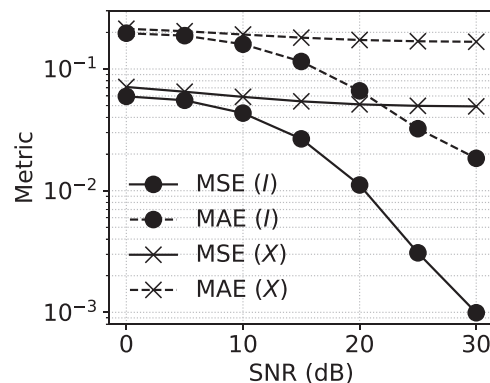

FIG. 7. Mean-squared error (MSE) and mean-absolute error (MAE) against signal-to-noise ratio (SNR) for the interpolation (I) and extrapolation (X) sets.

comparison with the two-microphone method. As presented in Sec. II B 3, the source distance, array distance, and array rotation in the extrapolation test set are varied outside the range of the parameters used to generate the training set. The extrapolation dataset is clustered into discrete intervals of source distances, array distances, and array rotations. The MSE is calculated for each cluster of parameters and SNR values considered in Sec. V A. Figure 8 summarizes the results.

It can be seen that the MSE increases with decreasing SNR in all cases. Nevertheless, the MSE difference in the absence and presence of noise with 25 dB SNR is relatively marginal. Interestingly, as shown in Figs. 8(a) and 8(c), the MSE curves are more or less constant across source distances and array rotations, respectively, for each of the SNR values. This means the proposed model can generalize well to unseen source distances and array rotations. It can be conjectured that the various incidence angles of the source in the training set help the network learn features similar to those in sound fields with rotated arrays, thus allowing the proposed method to predict sound absorption regardless of the array rotation accurately.

As shown in Fig. 8(b), however, there is a trend of increasing MSE (i.e., overfitting) with increasing array distance, noticeable for less noisy measurements (e.g., SNR ≥ 25 dB). This observation supports the noiseless results in Figs. 6(c) and 6(e), where the proposed model performs better when the array is closer to 2 cm (i.e., the distance used in the BEM model to generate the training set). In particular, the MSE performance in Fig. 8(b) is relatively similar up to array distances of 5 cm, with variations on the order of 5×10^{-3}. The increase in MSE of Fig. 8(b) is gentle and suggests that the model is not overfitting too much. At any rate, the proposed model is suspected of having more difficulties generalizing to a new range of array distances than an expanded set of source distances. Similar results have been observed with the MAE metric, not shown here for conciseness.

Additionally, the proposed and the two-microphone models are compared in Fig. 9, considering five frequencies and four clusters of array distances. The central learning is the interplay between frequencies and edge-diffraction and between array distance and overfitting, and how these

determine the statistical performance of the proposed method and the two-microphone method. At 125 and 250 Hz, the proposed method has more counts of smaller errors compared to the two-microphone method. For the array distance cluster of $z = 3$ cm (top row of Fig. 9), the neural network performs at least as well as the two-microphone method at all frequencies. However, similar to the results in Sec. IV C and Fig. 6(f), the bottom rows of Fig. 9 show that the performance of the proposed method deteriorates due to the pressure field at unseen array distances, in particular at frequencies beyond 500 Hz, rendering the two-microphone method more accurate. That the two-microphone method performs worse (better) at low (high) frequencies corroborates the stronger (weaker) contribution of edge diffraction to the sound field. It is worth noting that the error distributions of the two-microphone method differ slightly row-wise, although the array distance variable is meaningless for this method. This is likely because no combination of BEM parameters is identical in the extrapolation dataset.

Another way to understand this result is from the perspective of the area of the sample. It is known that as the absorber size exceeds the acoustic wavelength, the microphones hardly capture the diffracted sound field, and the two-microphone method approximates the sound absorption coefficient as if the sample were infinite. Conversely, as the absorber size is smaller than the wavelength, the diffracted sound field contaminates the predictions of the two-microphone method.

Plotting similar results to those in Fig. 9 by clustering sample area instead of frequencies is possible. The error distributions follow a similar behavior (not shown here for brevity): the neural network performs better than the two-microphone method for the smaller absorbers, i.e., as the acoustic wavelength becomes larger than the sample's size.

VI. CONCLUSION

The results from this study illustrate the potential of machine learning to mitigate the finite-size effect across a wide frequency range and for various porous absorber dimensions, flow resistivity values, source distances, and array rotations. In particular, the network can generalize
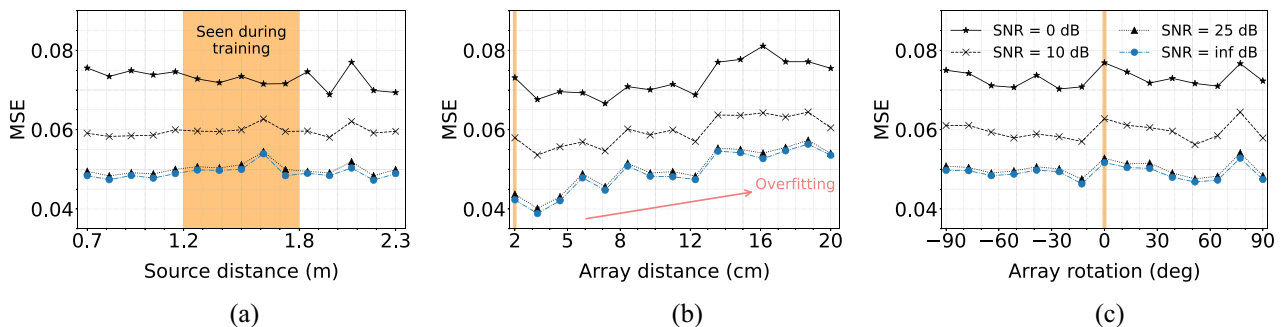

FIG. 8. (Color online) Sensitivity of the mean-squared error (MSE) to noisy pressure fields and unseen source distances, array distances, and array rotations. The orange regions indicate the parameters the network saw during training. Mean-squared error (MSE) for clustered sets of the extrapolation dataset and varying SNR. (a) MSE versus source distance. (b) MSE versus array distance. (c) MSE versus array rotation.
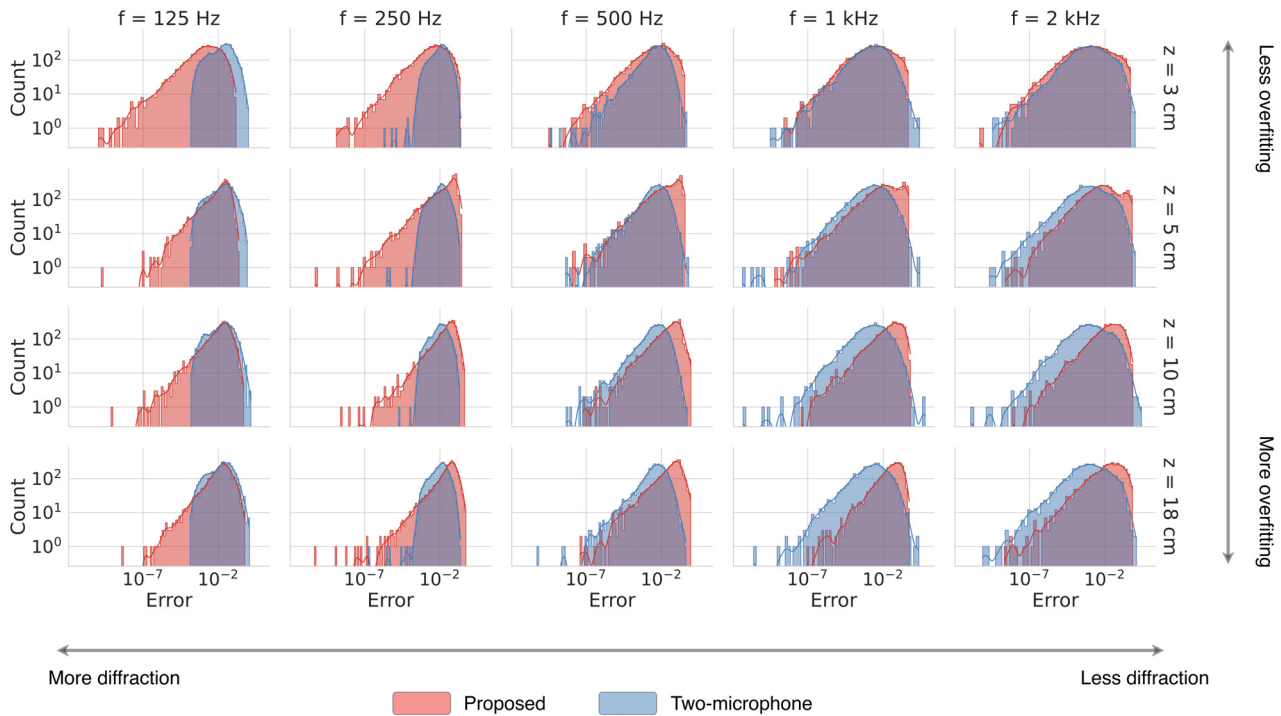
FIG. 9. (Color online) Error histograms obtained with the two-microphone method (blue/back) and the proposed method (red/front) with the noiseless extrapolation dataset at five frequencies (columns) and four clusters of array distances (rows). The error in the abscissae is the logarithmic squared error per frequency, $\log_{10}|\alpha(f) - \alpha^\star(f)|^2$, where α is the sound absorption coefficient produced with the analytical model, and $\alpha^\star(f)$ is the sound absorption coefficient obtained with the method of choice (two microphones or proposed).

well to source distances and array rotations not shown during the training phase. Although the proposed method is more complex than the two-microphone method regarding instrumentation and hardware, its accuracy for small absorbers at low frequencies is superior. The current bottleneck of the proposed method is the inability to generalize to unseen array distances. A way to overcome this is to fine-tune the model with a training set generated with various array distances.

A natural extension of the present work is to train the machine learning architecture with more realistic data. For instance, to deal with situations where the surface impedance is not constant along the sample's surface (e.g., spherical wave incidence). Alternatively, to deal with situations in which the sample is not flush mounted above the baffle, is contained in a frame, or has free edges. Furthermore, given the experimental validity of the BEM method (Brandão *et al.*, 2012), it is of great interest to validate the proposed model—trained with simulations—in an experimental scenario. This requires the design and deployment of a new measurement setup since the sampling parameters (e.g., microphone spacing and array dimensions) of experiments in the literature are incompatible with the proposed method. Training data in reverberant spaces can also be used to account for reflections from the environment. These topics are the subject of future research.

APPENDIX: TWO-MICROPHONE METHOD

The two-microphone method (Allard and Sieben, 1985) uses two microphones placed along the normal to the surface of interest to separate the incident and reflected components of the field. This is done under the assumption of specular reflection, such that the reflected sound field arises from the image source at $\mathbf{r}'_q$ as a spherical wave (Li and Hodgson, 1997). Furthermore, the reflection coefficient is assumed to be that of plane waves. The sound pressure at the microphone positions $\mathbf{r}_i$ and frequency f is thus

$$p(\mathbf{r}_i, f) = \frac{e^{-jk_0\|\mathbf{r}_i - \mathbf{r}_q\|}}{\|\mathbf{r}_i - \mathbf{r}_q\|} + R(f)\frac{e^{-jk_0\|\mathbf{r}_i - \mathbf{r}'_q\|}}{\|\mathbf{r}_i - \mathbf{r}'_q\|}, \tag{A1}$$

where $\mathbf{r}_i$, $i = 1, 2$ are the positions of the two microphones. The reflection coefficient of the sample follows:

$$R(f) = \frac{\dfrac{e^{-jk_0\|\mathbf{r}_2 - \mathbf{r}_q\|}}{\|\mathbf{r}_2 - \mathbf{r}_q\|} - \dfrac{p(\mathbf{r}_2)}{p(\mathbf{r}_1)}\dfrac{e^{-jk_0\|\mathbf{r}_1 - \mathbf{r}_q\|}}{\|\mathbf{r}_1 - \mathbf{r}_q\|}}{\dfrac{p(\mathbf{r}_2)}{p(\mathbf{r}_1)}\dfrac{e^{-jk_0\|\mathbf{r}_1 - \mathbf{r}'_q\|}}{\|\mathbf{r}_1 - \mathbf{r}'_q\|} - \dfrac{e^{-jk_0\|\mathbf{r}_2 - \mathbf{r}'_q\|}}{\|\mathbf{r}_2 - \mathbf{r}'_q\|}}, \tag{A2}$$

and the sound absorption coefficient is obtained via

2330   J. Acoust. Soc. Am. **154** (4), October 2023

Zea *et al.*

$$\alpha^{\star}(f) = 1 - |R(f)|^2. \tag{A3}$$

In an experimental setting, the distance between the two microphones must be large enough such that a phase difference is observed (Allard and Sieben, 1985), yet small enough to avoid spatial aliasing. The microphones are placed at 1 cm and 3 cm above the sample for numerical illustrations in the present paper.

[1]For air at room temperature, $f_{max} \approx 3.4$ kHz in this study.

[2]Datasets, models, and codes are available online (Zea, 2023).

[3]A physical interpretation of point-wise convolutions applied to pressure fields at multiple frequencies is that features can be learned from spatial and spectral domains. In the context of this study—taking the first convolutional layer as an example, a point-wise filter learns features from the sound pressure spectra.

Alkmim, M., Cuenca, J., De Ryck, L., and Desmet, W. (**2021**). "Angle-dependent sound absorption estimation using a compact microphone array," J. Acoust. Soc. Am. **150**(4), 2388–2400.

Allard, J. F., and Atalla, N. (**2009**). *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, 1st ed. (Wiley, Chichester, UK).

Allard, J. F., and Sieben, B. (**1985**). "Measurements of acoustic impedance in a free field with two microphones and a spectrum analyzer," J. Acoust. Soc. Am. **77**(4), 1617–1618.

Atalla, N., and Sgard, F. (**2015**). *Finite Element and Boundary Element Methods in Structural Acoustics and Vibration*, 1st ed. (CRC Press, Boca Raton, FL).

Bianco, M. J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M. A., Gannot, S., and Deledalle, C.-A. (**2019**). "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. **146**(5), 3590–3628.

Brandão, E., and Fernandez-Grande, E. (**2022**). "Analysis of the sound field above finite absorbers in the wave-number domain," J. Acoust. Soc. Am. **151**(5), 3019–3030.

Brandão, E., Lenzi, A., and Cordioli, J. (**2012**). "Estimation and minimization of errors caused by sample size effect in the measurement of the normal absorption coefficient of a locally reactive surface," Appl. Acoust. **73**(6-7), 543–556.

Brandão, E., Lenzi, A., and Paul, S. (**2015**). "A review of the *in situ* impedance and sound absorption measurement techniques," Acta Acust. Acust. **101**(3), 443–463.

Chollet, F. (**2015**). "Keras: Deep Learning for humans," https://github.com/fchollet/keras (Last viewed March 5, 2023).

Chollet, F. (**2017**). "Xception: Deep Learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807.

de Bruijn, A. (**1973**). "A mathematical analysis concerning the edge effect of sound absorbing materials," Acta Acust. Acust. **28**(1), 33–44.

Foy, C., Deleforge, A., and Di Carlo, D. (**2021**). "Mean absorption estimation from room impulse responses using virtually supervised learning," J. Acoust. Soc. Am. **150**(2), 1286–1299.

Goodfellow, I., Bengio, Y., and Courville, A. (**2016**). *Deep Learning* (MIT Press, Cambridge, MA).

Götz, G., Falcón Pérez, R., Schlecht, S. J., and Pulkki, V. (**2022**). "Neural network for multi-exponential sound energy decay analysis," J. Acoust. Soc. Am. **152**(2), 942–953.

Hahmann, M., Verburg, S. A., and Fernandez-Grande, E. (**2021**). "Spatial reconstruction of sound fields using local and data-driven functions," J. Acoust. Soc. Am. **150**(6), 4417–4428.

He, K., Zhang, X., Ren, S., and Sun, J. (**2016a**). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

He, K., Zhang, X., Ren, S., and Sun, J. (**2016b**). "Identity mappings in deep residual networks," in *Computer Vision—ECCV 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling (Springer, Cham, Switzerland), pp. 630–645.

Hirosawa, K., Takashima, K., Nakagawa, H., Kon, M., Yamamoto, A., and Lauriks, W. (**2009**). "Comparison of three measurement techniques for

the normal absorption coefficient of sound absorbing materials in the free field," J. Acoust. Soc. Am. **126**(6), 3020–3027.

Hochreiter, S. (**1991**). "Untersuchungen zu dynamischen neuronalen netzen," ("Studies on dynamic neural networks"), diploma thesis, Institut fur Informatik, Technische Universität Munich, Munich.

Ioffe, S., and Szegedy, C. (**2015**). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning, ICML'15*, Lille, France, Vol. 37, p. 448–456.

ISO (**1998**). ISO 10534-2:1998: "Acoustics – Determination of sound absorption coefficient and impedance in impedance tubes – Part 2: Transfer-function method" (International Organization for Standardization, Geneva, Switzerland).

ISO (**2003**). ISO 354:2003 "Acoustics – Measurement of sound absorption in a reverberation room" (International Organization for Standardization, Geneva, Switzerland).

Jeon, J. H., Chemali, E., Yang, S. S., and Kang, Y. J. (**2021**). "Convolutional neural networks for estimating transport parameters of fibrous materials based on micro-computerized tomography images," J. Acoust. Soc. Am. **149**(4), 2813–2828.

Kingma, D. P., and Ba, J. (**2015**). "Adam: A method for stochastic optimization," CoRR abs/1412.6980.

Lähivaara, T., Kärkkäinen, L., Huttunen, J. M. J., and Hesthaven, J. S. (**2018**). "Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography," J. Acoust. Soc. Am. **143**(2), 1148–1158.

LeCun, Y., Bengio, Y., and Hinton, G. (**2015**). "Deep learning," Nature **521**(7553), 436–444.

LeCun, Y., Jackel, L., Boser, B., Denker, J., Graf, H., Guyon, I., Henderson, D., Howard, R., and Hubbard, W. (**1989**). "Handwritten digit recognition: Applications of neural network chips and automatic learning," IEEE Commun. Mag. **27**(11), 41–46.

Li, J.-F., and Hodgson, M. (**1997**). "Use of pseudo-random sequences and a single microphone to measure surface impedance at oblique incidence," J. Acoust. Soc. Am. **102**(4), 2200–2210.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (**2017**). "A survey of deep neural network architectures and their applications," Neurocomputing **234**, 11–26.

Luo, Z.-W., Zheng, C.-J., Zhang, Y.-B., and Bi, C.-X. (**2020**). "Estimating the acoustical properties of locally reactive finite materials using the boundary element method," J. Acoust. Soc. Am. **147**(6), 3917–3931.

Michalopoulou, Z.-H., Gerstoft, P., Kostek, B., and Roch, M. A. (**2021**). "Introduction to the special issue on machine learning in acoustics," J. Acoust. Soc. Am. **150**(4), 3204–3210.

Miki, Y. (**1990**). "Acoustical properties of porous materials-modifications of Delany-Bazley models," J. Acoust. Soc. Jpn. (E) **11**(1), 19–24.

Mommertz, E. (**1995**). "Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique," Appl. Acoust. **46**(3), 251–263.

Müller-Giebeler, M., and Vorländer, M. (**2021**). "Modeling the edge effect for inverse determination of porous absorbers using feed forward neural networks," *Euronoise*, e-Congress.

Nair, V., and Hinton, G. E. (**2010**). "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Omnipress*, Madison, WI, p. 807–814.

Nolan, M. (**2020**). "Estimation of angle-dependent absorption coefficients from spatially distributed *in situ* measurements," J. Acoust. Soc. Am. **147**(2), EL119–EL124.

Otsuru, T., Tomiku, R., Din, N. B. C., Okamoto, N., and Murakami, M. (**2009**). "Ensemble averaged surface normal impedance of material using an *in-situ* technique: Preliminary study using boundary element method," J. Acoust. Soc. Am. **125**(6), 3784–3791.

Ottink, M., Brunskog, J., Jeong, C.-H., Fernandez-Grande, E., Trojgaard, P., and Tiana-Roig, E. (**2016**). "*In situ* measurements of the oblique incidence sound absorption coefficient for finite sized absorbers," J. Acoust. Soc. Am. **139**(1), 41–52.

Richard, A., Fernandez-Grande, E., Brunskog, J., and Jeong, C.-H. (**2017**). "Estimation of surface impedance at oblique incidence based on sparse array processing," J. Acoust. Soc. Am. **141**(6), 4115–4125.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (**2017**). "Inception-v4, Inception-ResNet and the impact of residual connections on learning,"

J. Acoust. Soc. Am. **154** (4), October 2023

Zea *et al.* 2331

in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17* (AAAI Press, Washington, DC), pp. 4278–4284.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (**2015**). "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9.

Tamura, M. (**1990**). "Spatial Fourier transform method of measuring reflection coefficients at oblique incidence. I: Theory and numerical examples," J. Acoust. Soc. Am. **88**(5), 2259–2264.

Thomasson, S.-I. (**1980**). "On the absorption coefficient," Acta Acust. Acust. **44**(4), 265–273.

Wu, T. (**2000**). *Boundary Element Acoustics: Fundamentals and Computer Codes*, 1st ed. (WIT Press, Southampton, UK).

Yu, W., and Kleijn, W. B. (**2021**). "Room acoustical parameter estimation from room impulse responses using deep neural networks," IEEE/ACM Trans. Audio. Speech. Lang. Process. **29**, 436–447.

Zea, E. (**2023**). "finite-absorber-ML," https://github.com/eliaszea/finite-absorber-ML (Last viewed July 13, 2023).

Zea, E., Brandão, E., Nolan, M., Andén, J., Cuenca, J., and Svensson, U. P. (**2021**). "Learning the finite size effect for *in-situ* absorption measurement," *Euronoise*, e-Congress.

2332   J. Acoust. Soc. Am. **154** (4), October 2023

Zea *et al.*