DOI: 10.61782/fa.2025.0516

FORUM ACUSTICUM
EURONOISE 2025

LEARNING FROM DATA-DRIVEN SOUND FIELD ESTIMATION USING COMPLEX-VALUED NEURAL NETWORKS

Vlad S. Paul[1*] **Nara Hahn**[1] **Philip A. Nelson**[1]

[1] Institute of Sound and Vibration Research, University of Southampton, United Kingdom

ABSTRACT

This paper explores the use of a complex-valued neural network for virtual sensing applications. The aim is to estimate the pressures from single frequency plane waves from various directions at control points where physical microphone measurements are not feasible. Making use of measurements from a microphone array arranged on an open sphere, the proposed network is trained to infer the spatial properties of sound fields, predicting the pressure at designated virtual sensor locations. A key contribution of this work is the analysis of the network's internal operations via singular value decomposition (SVD) of its weight matrices. This analysis reveals how the captured sound fields are spatially encoded by the hidden layer, which can be considered as a pre-processing step. Different network configurations and training scenarios will be investigated, focusing on examining the spatial filtering performed by the hidden layer. The results not only demonstrate the potential of complex-valued neural networks in the context of virtual acoustic sensing but also provide valuable insights into its decision-making process.

Keywords: *complex-valued neural networks, sound field reconstruction, virtual sensing*

**Corresponding author*: *vsp1u21@soton.ac.uk.*

1. INTRODUCTION

Sound field reconstruction is crucial for applications including spatial audio rendering, immersive virtual reality, and active noise control [1, 2]. Very often, it is necessary to estimate the acoustic pressure at positions where physical microphone placement is impractical, an approach known as virtual sensing [3]. Virtual sensors allow estimation of the sound field at arbitrary positions by making use of data captured from limited or irregularly placed microphone arrays. Traditional methods such as spherical harmonic expansions (SHE) or interpolation techniques often suffer from spatial aliasing, limited frequency range, and modeling inaccuracies particularly when measurements are sparse [4].

Recently, data-driven approaches, particularly those using neural networks, have shown potential to overcome these limitations by directly learning spatial relationships within acoustic fields from measured data [5, 6]. However, common neural networks typically handle only real-valued data, neglecting the inherently complex-valued nature of frequency-domain signals (magnitude and phase). The phase information, critical for accurate spatial reconstruction of acoustic signals, is thus often poorly represented [7].

Complex-valued neural networks (CVNNs) offer a promising solution by naturally processing complex data and preserving the crucial magnitude-phase relationship [8, 9]. CVNNs have demonstrated superior performance in some acoustic applications such as near-field acoustic holography [10], classification of acoustic spectra [11] or sound source localisation [12]. Moreover, because wave propagation is modeled using

complex exponentials that capture both amplitude and phase, CVNNs naturally operate in this domain, potentially leading to more accurate sound field reconstructions.

This paper explores the application of CVNNs for estimating single-frequency sound fields captured from an open spherical microphone array, focusing on their use in virtual sensing scenarios. A complex-valued multilayer perceptron (cMLP) is used for this task. By performing singular value decomposition (SVD) on the network weights relating the input and the hidden layer, the spatial filtering mechanisms employed by the network are investigated, revealing insights into how CVNNs encode and interpret acoustic data. Through various network configurations, we demonstrate that CVNNs can robustly reconstruct virtual sensor signals with improved spatial accuracy compared to benchmark techniques, especially at frequencies above the spatial aliasing frequency.

2. METHODOLOGY

2.1 Problem description

Consider a homogeneous, free-space environment, where the sound field is modeled as a single-frequency plane wave. The pressure p measured at a microphone located at position $\mathbf{r}_\mathrm{m}$ is given by

$$p(k, \mathbf{r}_\mathrm{m}) = Ae^{-j(k\mathbf{n}_\mathrm{s}.\mathbf{r}_\mathrm{m}+\alpha)}, \tag{1}$$

where A denotes the unit amplitude, $k = 2\pi f/c$ is the wavenumber of frequency f and speed of sound c. The dot product $\mathbf{n}_\mathrm{s}.\mathbf{r}_\mathrm{m}$ determines the phase shift of the plane wave as it arrives at the microphone at position vector $\mathbf{r}_\mathrm{m}$, where the unit vector $\mathbf{n}_\mathrm{s}$ indicates the direction of the plane wave propagation. The additional phase term α is introduced to apply a random offset to each plane wave when training the neural network model. This ensures that the arrival time at the microphone array varies for each wave, effectively removing the phase shift from the plane wave starting point to the array.

The measurements of the single-frequency plane waves are computed at a microphone array arranged on the surface of an open sphere. The sensors are placed nearly uniformly on the sphere at locations $\mathbf{r}_\mathrm{m}$ that replicate the angular positions of those of an Eigenmike em64 [13]. The radius of the array is set to $R_\mathrm{mic} = 0.4$ m. Due to the finite spatial sampling of the open sphere, the microphone array has an inherent spatial aliasing frequency f_alias, which limits the effective frequency range for accurate sound field reconstruction using traditional spherical harmonic (SH) methods. This frequency can be computed for the array used here as [4]

$$f_\mathrm{alias} = \frac{cN}{2\pi R_\mathrm{mic}} = 955.33 \text{ Hz}, \tag{2}$$

where $N = 7$ is the maximum SH order determined by the sampling of the open sphere microphone array and the speed of sound is assumed to be $c = 343$ m/s.

For the reconstruction task, the virtual sensors denoted here as control points are placed within a smaller sphere of radius $R_\mathrm{c} = 0.2$ m. A total of 2945 control points were uniformly distributed on a 3-dimensional Cartesian grid throughout the volume of this smaller sphere, ensuring complete and even spatial coverage. The spacing is determined based on the smallest wavelength of interest to ensure that spatial aliasing is avoided at the control points. This guarantees that the reconstruction at these locations remains accurate even though the microphone array itself may experience aliasing at higher frequencies.

2.2 Dataset Generation

To evaluate the performance of the cMLP for this sound field reconstruction task, three different datasets were generated, each corresponding to a specific plane wave frequency relative to the spatial aliasing threshold of the open-sphere microphone array (see Eq. (2)). The three plane wave frequencies are defined as

- One octave below aliasing: $f_\mathrm{low} = 477.66$ Hz
- At the aliasing frequency: $f_\mathrm{alias} = 955.33$ Hz
- One octave above aliasing: $f_\mathrm{high} = 1910.66$ Hz

This selection of frequencies enables the investigation of the network's robustness to aliasing effects and its ability to reconstruct sound fields at and beyond the aliasing threshold.

For each frequency, plane waves originating from 2000 different directions $\mathbf{n}_\mathrm{s}$ quasi-uniformly distributed over the sphere are generated with a random phase. The distribution of plane waves ensures comprehensive coverage of possible incident angles and facilitates a thorough assessment of the network's generalization to unseen directions. Each dataset was partitioned into 80% of the plane wave directions for training and 20%

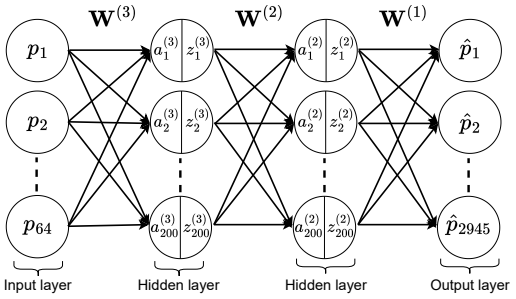


Figure 1: cMLP model with two hidden layers used for the sound field reconstruction task.

of the directions for validation purposes. By structuring the datasets in this manner, the aim is to rigorously evaluate the CVNN's ability to generalize to unseen plane wave directions across a range of frequencies, particularly in scenarios where spatial aliasing may impact reconstruction quality.

2.3 Network parameters

A cMLP with two hidden layers was used to solve the proposed sound field reconstruction task and is illustrated in Figure 1. The measured sound field at the $M = 64$ microphones on the open sphere were used as input features to the network. The target outputs were the pressure values at the 2945 control points within the smaller sphere. The network's objective was to accurately predict these values, effectively reconstructing the sound field within the control region. The two hidden layers contained 200 neurons and had the complex cardioid activation function [14], which reduces to the ReLU function if the data is real. The output layer contained no activation function. A learning rate of 0.001 was used initially and after every 50 epochs this value was halved. The model was trained for a total of 300 iterations using the ADAM optimizer to update the weights and with a batch size of 16.

2.4 Benchmark techniques

The proposed cMLP model is compared with two different signal processing based approaches in terms of sound field reconstruction performance. These benchmark methods are the spherical harmonic expansion (SHE) and least squares (LS) solution. All three techniques can be represented within a common framework that transforms microphone pressures into pressures at designated control points. The cMLP network is defined as a nonlinear approach due to the nonlinear activation functions while the other two benchmark methods are defined as linear transformations.

The proposed cMLP approach can be defined as a nonlinear transformation of the pressure recorded at the microphone array $\mathbf{p}_\mathrm{m}$ to the pressure at the control points $\mathbf{p}_\mathrm{c}$ and is given by

$$\mathbf{p}_\mathrm{c} = \mathcal{N}(\mathbf{p}_\mathrm{m}), \tag{3}$$

where $\mathcal{N}$ represents the nonlinearity of the network model.

2.4.1 *Spherical Harmonic Expansion (SHE)*

The SHE approach exploits the physical properties of the acoustic wave propagation. Using the common framework, the technique can be defined as $\mathbf{p}_\mathrm{c} = \mathbf{W}_\mathrm{she}\mathbf{p}_\mathrm{m}$. The computation of $\mathbf{W}_\mathrm{she}$ involves three main steps.

First, the measured pressure $\mathbf{p}_\mathrm{m}$ can be expressed in terms of SH using matrix form notation as

$$\mathbf{p}_\mathrm{m} = \mathbf{Y}_\mathrm{m}\mathbf{D}_\mathrm{m}\tilde{\mathbf{p}}, \tag{4}$$

where $\mathbf{Y}_\mathrm{m}$ is the matrix of SHs evaluated at the $M = 64$ microphones and $\tilde{\mathbf{p}}$ is the vector of SH coefficients [4]. Due to the number of microphones on the open sphere, the maximum SH order is $N = 7$, which yields $(N+1)^2$ SH functions. It follows that $\mathbf{Y}_\mathrm{m}$ is a 64×64 matrix. The matrix $\mathbf{D}_\mathrm{m}$ is a 64×64 diagonal matrix representing the radial dependency at R_mic, defined as $\mathbf{D}_\mathrm{m} = \mathrm{diag}\{j_\mathrm{n}(kR_\mathrm{mic})\}$ with $j_\mathrm{n}(kR_\mathrm{mic})$ being the spherical Bessel function of order n. The vector of SH coefficients $\tilde{\mathbf{p}}$ can thus be computed from $\tilde{\mathbf{p}} = \mathbf{D}_\mathrm{m}^{-1}\mathbf{Y}_\mathrm{m}^{\dagger}\mathbf{p}_\mathrm{m}$, where $\mathbf{Y}_\mathrm{m}^{\dagger}$ is the Moore-Penrose pseudoinverse of $\mathbf{Y}_\mathrm{m}$.

Using a similar approach, the pressure at the control points can be represented in terms of SH as

$$\mathbf{p}_\mathrm{c} = \mathbf{Y}_\mathrm{c}\mathbf{D}_\mathrm{c}\tilde{\mathbf{p}}, \tag{5}$$

where $\mathbf{Y}_\mathrm{c}$ is the matrix of SH evaluated at the 2945 control points and $\mathbf{D}_\mathrm{c}$ is the diagonal matrix of radial dependencies at R_cp. Replacing the vector of SH coefficients $\tilde{\mathbf{p}}$ yields

$$\mathbf{p}_\mathrm{c} = \mathbf{Y}_\mathrm{c}\mathbf{D}_\mathrm{c}\mathbf{D}_\mathrm{m}^{-1}\mathbf{Y}_\mathrm{m}^{\dagger}\mathbf{p}_\mathrm{m} \tag{6}$$

and a transformation matrix $\mathbf{W}_\mathrm{she}$ can be computed from $\mathbf{W}_\mathrm{she} = \mathbf{Y}_\mathrm{c}\mathbf{D}_\mathrm{c}\mathbf{D}_\mathrm{m}^{-1}\mathbf{Y}_\mathrm{m}^{\dagger}$.

This approach provides a systematic framework for reconstructing the sound field at the control points, however it suffers from certain limitations. For example, the matrix $\mathbf{D}_\mathrm{m}^{-1}$ contains terms of the form $1/j_\mathrm{n}(kR_\mathrm{mic})$, where the spherical Bessel function $j_\mathrm{n}(kR_\mathrm{mic})$ may approach 0. For a given radius, the spherical Bessel functions at multiple frequencies are zero and can lead to numerical instabilities which will impact the reconstruction of the pressure. The second limitation is related to the spatial aliasing. If the measured microphone pressure $\mathbf{p}_\mathrm{m}$ contains errors due to undersampling at high frequencies, these errors will affect the computation of $\mathbf{p}_\mathrm{c}$, further degrading the reconstruction quality [4].

2.4.2 *Least squares (LS) solution*

This method computes a simple linear mapping between the pressures at the measurement microphone array and that at the control points. The aim is to minimise the error in estimating the control point pressures based on the available microphone data. First, two matrices are generated with the available training data at the microphones and control points. $\breve{\mathbf{P}}_\mathrm{m} \in \mathbb{C}^{64\times1600}$ contains the 64 pressures at the microphone positions for all 1600 training plane waves. $\breve{\mathbf{P}}_\mathrm{c} \in \mathbb{C}^{2945\times1600}$ contains the pressure values at the control points for the same set of plane waves.

We assume that the relationship between the measured microphone pressures in the training dataset $\breve{\mathbf{P}}_\mathrm{m}$ and the pressures at the control points $\breve{\mathbf{P}}_\mathrm{c}$ can be approximated by a linear mapping represented by the transformation matrix $\mathbf{W}_\mathrm{ls}$. This can be formulated as $\breve{\mathbf{P}}_\mathrm{c} \approx \mathbf{W}_\mathrm{ls}\breve{\mathbf{P}}_\mathrm{m}$. To determine the optimal matrix $\mathbf{W}_\mathrm{ls}$, the aim is to minimize the reconstruction error in the least-squares sense

$$\min_{\mathbf{W}_\mathrm{ls}} \|\breve{\mathbf{P}}_\mathrm{c} - \mathbf{W}_\mathrm{ls}\breve{\mathbf{P}}_\mathrm{m}\|_F^2, \tag{7}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. After some algebraic steps including a derivative with respect to $\mathbf{W}_\mathrm{ls}$, it follows that the general solution for $\mathbf{W}_\mathrm{ls}$ is expressed as

$$\mathbf{W}_\mathrm{ls} = \breve{\mathbf{P}}_\mathrm{c}\,\breve{\mathbf{P}}_\mathrm{m}^\dagger, \tag{8}$$

where $\breve{\mathbf{P}}_\mathrm{m}^\dagger$ denotes the Moore-Penrose pseudoinverse of $\breve{\mathbf{P}}_\mathrm{m}$ [15]. For the validation set, the pressures at the control points are estimated from

$$\mathbf{p}_\mathrm{c} = \mathbf{W}_\mathrm{ls}\mathbf{p}_\mathrm{m}. \tag{9}$$

Table 1: Normalised mean squared errors of the cMLP and the benchmark techniques for the validation dataset, expressed as an averaged error over all control points in dB.

	f_low	f_alias	f_high
cMLP	-43 dB	-29 dB	-3 dB
SHE	-80 dB	0 dB	18 dB
LS	-81 dB	-9 dB	-1 dB

It should be noted that both the cMLP and the LS methods are data-driven approaches, using the same training data $\breve{\mathbf{P}}_\mathrm{c}, \breve{\mathbf{P}}_\mathrm{m}$. In contrast, the SHE method relies on a physical (analytical) model rather than empirical data.

3. RESULTS

This section will present a comparative analysis between the cMLP and the benchmark techniques across various scenarios. First, the sound field reconstruction performance is assessed for the general task as described above. Next, an analysis of the internal operations of the three techniques will be presented, focusing on the singular vectors of the matrices. In the end, based on previous observations discussed in [16], the performance of the cMLP with microphone pressures as input will be compared to that of the cMLP with other features as input.

3.1 Normalised mean squared errors

Table 1 shows the reconstruction performance of the cMLP and the two benchmark techniques (SH and LS) expressed as an averaged error in dB for all validation plane waves. The error is averaged over all 2945 control points. It can be observed that for the f_low, both benchmark techniques outperform the cMLP by a significant amount. This is not necessarily surprising, since at frequencies significantly below the aliasing frequency, these traditional methods are highly accurate due to their solid mathematical foundation and the ideal conditions, such as sufficient spatial sampling relative to the wavelength. At low frequencies, the acoustic field varies smoothly over space and the pressure can be accurately captured using a relatively small number of SH modes. Due to the simple wave patterns, simple linear mappings such as SHE and LS approaches can handle these spatial patterns, while it is likely that the nonlinear cMLP model introduces unnecessary complexity.

When the frequency approaches or exceeds the spatial aliasing frequency, traditional methods like SH and LS degrade significantly, as observed in Table 1. This is mainly due to the errors in the measured sound field introduced by aliasing. Especially for the SHE approach, higher-order spherical modes are not properly captured and the missing information leads to a lower reconstruction performance. The cMLP model outperforms both benchmark techniques at f_{alias} and f_{high} due to the fact that it can learn complex patterns from the data, including aliasing artifacts. Essentially, the network is able to compensate for the undersampling in the input data and can reconstruct an aliasing free sound field at the control points. Similar promising observations have been made by [5, 7, 17] and suggest that one can use neural networks to improve the reconstruction performance around and above spatial aliasing.

Figure 2 shows an example of the sound field reconstruction performance at the control points for a plane wave at $f_{\text{alias}} = 955.33$ Hz using the trained cMLP model.

3.2 Spatial filtering

In previous work [16], it has been shown that for a similar task of sound field reconstruction, the first hidden layer in the cMLP is decomposing the input pressure field into spatial basis functions that resemble SH modes. Specifically, by evaluating the SVD of $\mathbf{W}^{(3)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\text{H}}$, where $\mathbf{W}^{(3)}$ is the matrix relating the input and the first hidden layer (see Figure 1), the resulting right singular vectors in $\mathbf{V}$ exhibit clear physical interpretations. For low frequencies below f_{alias}, the first mode closely resembles an omnidirectional pattern, while subsequent modes reflect dipole-like patterns oriented along different spatial axes.

Following this, it is interesting to compare the spatial filtering operations of the cMLP model with the SHE and LS methods. This can be achieved by computing the SVD of $\mathbf{W}_{\text{she}}$ and $\mathbf{W}_{\text{ls}}$ and examining their right singular vector matrices $\mathbf{V}$. For all methods, the columns of $\mathbf{V}$ represent spatial patterns at the microphone positions. To investigate the relation of these patterns to traditional SH modes, we first define a so-called beamformer output for a plane wave arriving from direction $\mathbf{n}_{\text{l}}$ as

$$z_{\text{j}}(\mathbf{n}_{\text{l}}) = \mathbf{v}_{\text{j}}^{\text{H}}\mathbf{p}(k, \mathbf{n}_{\text{l}}), \tag{10}$$

where $j = 1...J$. By evaluating $z_{\text{j}}(\mathbf{n}_{\text{l}})$ at L directions, we form a beamformer output matrix $\mathbf{Z} = \mathbf{V}^{\text{H}}\mathbf{P}$, where $\mathbf{P} = [\mathbf{p}(\mathbf{n}_1), \mathbf{p}(\mathbf{n}_1), ..., \mathbf{p}(\mathbf{n}_L)]$. For the three different methods, we define three matrices of beamformer outputs $\mathbf{Z}_{\text{mlp}}, \mathbf{Z}_{\text{ls}}, \mathbf{Z}_{\text{she}}$. Using a general beamformer output matrix $\mathbf{Z}$, each of the above matrices can be expanded in terms of SH as

$$\mathbf{Z}^{\text{T}} = \mathbf{Y}_{\text{l}}\widetilde{\mathbf{Z}}, \tag{11}$$

where $\mathbf{Y}_{\text{l}}$ is an $L \times (N+1)^2$ matrix containing SH basis functions evaluated at L directions and $\widetilde{\mathbf{Z}}$ is an $(N+1)^2 \times J$ matrix of SH expansion coefficients [4]. Note that $N = 7$ is the maximum SH order. Similarly, the plane wave sound field matrix $\mathbf{P}$ can be expanded in terms of SH as $\mathbf{P} = \mathbf{Y}_{\text{m}}\widetilde{\mathbf{P}}$, where $\mathbf{Y}_{\text{m}}$ is a $M \times (N+1)^2$ matrix containing SH basis functions evaluated at $M = 64$ microphones. The SH coefficient matrix $\widetilde{\mathbf{P}}$ can be further decomposed into $\widetilde{\mathbf{P}} = \mathbf{D}_{\text{l}}\mathbf{Y}_{\text{l}}^{\text{T}}$, with $\mathbf{D}_{\text{l}}$ a diagonal matrix containing the spherical Bessel functions $j_{\text{n}}(kR_{\text{mic}})$ and phase terms j^{-n} originating from the SH expansion of plane waves. Substituting the above terms into the beamformer expression yields

$$\mathbf{Z} = \mathbf{V}^{\text{H}}\mathbf{P} \tag{12}$$

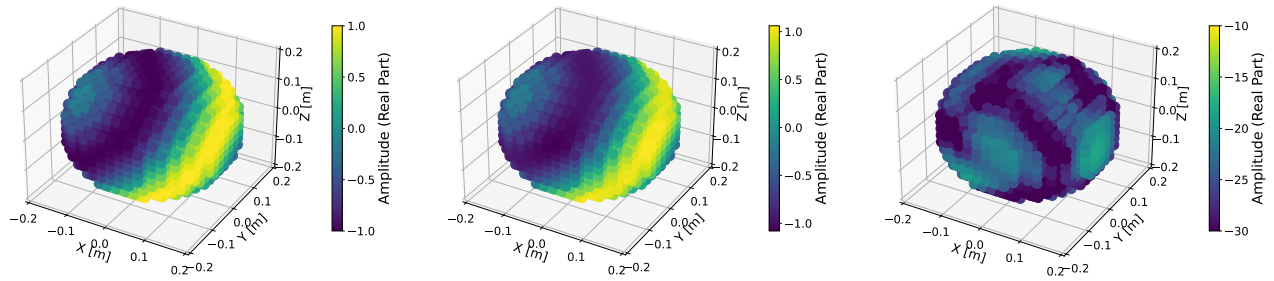
$$(\mathbf{Y}_{\text{l}}\widetilde{\mathbf{Z}})^{\text{T}} = \mathbf{V}^{\text{H}}\mathbf{Y}_{\text{m}}\widetilde{\mathbf{P}} \tag{13}$$

and so the SH coefficients of the beamformer becomes

$$\widetilde{\mathbf{Z}} = (\mathbf{V}^{\text{H}}\mathbf{Y}_{\text{m}}\mathbf{D}_{\text{l}})^{\text{T}}. \tag{14}$$

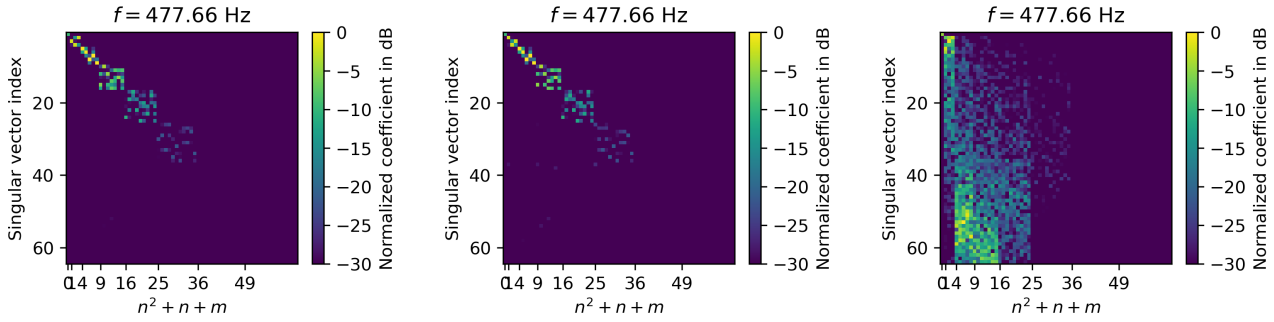
Figure 3 shows three plots corresponding to the magnitude of the elements of the matrices $\widetilde{\mathbf{Z}}_{\text{mlp}}, \widetilde{\mathbf{Z}}_{\text{ls}}, \widetilde{\mathbf{Z}}_{\text{she}}$ for the dataset of plane waves at $f_{\text{low}} = 477.66$ Hz. These plots show how strongly each SH mode is excited by each singular vector. The SHs corresponding to the first singular vectors are considered to be the most important when decomposing the pressures at the microphones. Essentially, the plots show how each of the techniques processes microphone pressures generated by plane waves arriving from directions uniformly distributed around the sphere.

It is interesting to observe that the LS and SHE approaches behave very similarly at f_{low} and this behaviour is emphasised by the almost identical reconstruction performance from Table 1 at f_{low}. Even if the LS method does not use any SH decomposition of the pressure, it concentrates its strength in the lower-order SH and requires minimal contributions from higher-order SHs. The magnitude plot of the cMLP model is different,



(a) Target (b) Prediction (c) Error

Figure 2: Example of reconstruction performance of one plane wave at $f_{\text{alias}} = 955.33$ Hz for the cMLP model.



(a) LS approach: $\widetilde{\mathbf{Z}}_{\text{ls}}$ (b) SHE approach: $\widetilde{\mathbf{Z}}_{\text{she}}$ (c) cMLP model: $\widetilde{\mathbf{Z}}_{\text{mlp}}$

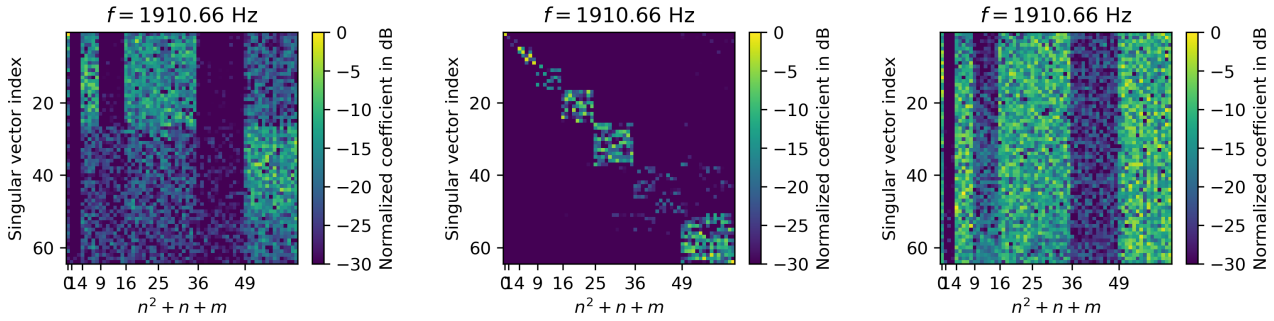
Figure 3: Magnitude of the SH expansion coefficients $\widetilde{\mathbf{Z}}$ for the dataset of plane waves at $f_{\text{low}} = 477.66$ Hz. Magnitude is normalized for each plot individually. The horizontal axis indicates the spherical harmonic index $n^2 + n + m$ where n and m are the SH order and degree, respectively.

showing that the network uses the information from more SHs in each singular vector and there is no clear pattern that an increased singular vector corresponds to SHs at higher orders. Even though the cMLP is using additional information compared to the other two methods, this information is only increasing the complexity of the learning, without really adding any benefit to the reconstruction performance (see Table 1). A pattern emerges though for all three techniques. For the f_{low} dataset, all three techniques use SHs just up to order $N = 5$.

Above the aliasing frequency at f_{high}, the magnitudes of the SH coefficients look different and are shown in Figure 4. The SHE approach has an expected behaviour with the magnitude of the coefficients split over all SHs, including those of higher order. The LS approach and the cMLP model have very similar behaviours and each singular vector uses a linear combination of most of the SHs. There are some obvious gaps of strength in all three plots at the same SHs and these gaps correspond to the zeros in the Bessel functions at specific orders. Interestingly, even if both the LS and the cMLP models do not use any Bessel functions in their approach, the methods cannot make use of the information contained in those SHs.

3.3 Performance analysis using different input features

Based on the fact that the weight matrix $\mathbf{W}_{\text{mlp}}$ between the input and the hidden layer is learning a modal decom-



(a) LS approach: $\widetilde{\mathbf{Z}}_{\mathrm{ls}}$ (b) SHE approach: $\widetilde{\mathbf{Z}}_{\mathrm{she}}$ (c) cMLP model: $\widetilde{\mathbf{Z}}_{\mathrm{mlp}}$

Figure 4: Magnitude of the SH expansion coefficients $\widetilde{\mathbf{Z}}$ for the dataset of plane waves at $f_{\mathrm{high}} = 1910.66$ Hz. Magnitude is normalized for each plot individually. The horizontal axis indicates the spherical harmonic index $n^2 + n + m$ where n and m are the SH order and degree, respectively.

Table 2: Normalised mean squared error of the cMLP using microphone pressures as input (cMLP-P) with one or two hidden layers, compared to the cMLP using SH coefficients in the input layer (cMLP-SH) with one or two hidden layers.

	f_{low}	f_{alias}	f_{high}
cMLP-P 1 layer	-19 dB	-14 dB	-1 dB
cMLP-P 2 layers	-43 dB	-29 dB	-3 dB
cMLP-SH 1 layer	-44 dB	-26 dB	-1 dB
cMLP-SH 2 layers	-41 dB	-29 dB	-2 dB

position of the sound field, similar to SHs, an interesting comparison of reconstruction performance is shown in Table 2. Here, the cMLP with microphone pressures as input layer (cMLP-P) is compared to a cMLP with SH coefficients as input layer (cMLP-SH). It is expected that by adding this pre-processing stage of SH decomposition of microphone pressures, the new cMLP model might need less time to converge and will need only one hidden layer to achieve a similar performance to that of the original cMLP. It can be observed that if the microphone signal is used as input into the network, using just one hidden layer decreases the reconstruction performance substantially at all three frequencies. However, if the preprocessing step of SH decomposition is added, the cMLP-SH model achieves a similar performance with one hidden layer as the cMLP-P with two hidden layers. This suggests that when the first hidden layer is removed from the cMLP-P approach, the network loses its ability to perform the critical basis transformation step. Providing this preprocessing step simplifies the learning task and reduces the network complexity needed. These observations further confirm that the learned internal spatial basis significantly contributes to the network's reconstruction capabilities.

4. CONCLUSION

In this study, the use of a complex-valued multilayer perceptron (cMLP) was investigated for a sound field reconstruction task, specifically addressing virtual sensing scenarios with spherical microphone arrays. While traditional methods (spherical harmonic expansions and least-squares solutions) excel at frequencies below spatial aliasing, the cMLP outperformed these benchmarks at and above the aliasing threshold, highlighting its robustness to spatial sampling limitations. Singular value decomposition of the network's input layer weights revealed that the network implicitly learns spatial basis functions closely resembling spherical harmonic modes, indicating a physically interpretable internal representation. Further validation showed that directly using spherical harmonic coefficients as inputs allowed for substantial simplification of the cMLP architecture, maintaining high reconstruction accuracy. Thus, complex-valued neural networks not only offer improved reconstruction at higher frequencies but also implicitly discover optimal spatial decompositions aligned with classical acoustics theory.

5. REFERENCES

[1] P. A. Nelson, "Active control of acoustic fields and the reproduction of sound," *J. Sound Vibration*, vol. 177, no. 4, pp. 447–477, 1994.

[2] E. Erdem, Z. Cvetković, and H. Hacıhabiboğlu, "3d perceptual soundfield reconstruction via virtual microphone synthesis," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 31, pp. 1305–1317, 2023.

[3] L. Liu, S. M. Kuo, and M. Zhou, "Virtual sensing techniques and their applications," in *Proc. IEEE Int. Conf. Netw. Sens. Control*, (Okayama, Japan), pp. 31–36, Mar. 2009.

[4] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015.

[5] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *J. Acoust. Soc. Am. (JASA)*, vol. 153, no. 2, pp. 1179–1190, 2023.

[6] S. Koyama, J. G. C. Ribeiro, T. Nakamura, N. Ueno, and M. Pezzoli, "Physics-informed machine learning for sound field estimation: Fundamentals, state of the art, and challenges," *IEEE Signal Process. Mag.*, vol. 41, no. 6, pp. 60–71, 2024.

[7] V. S. Paul, N. Hahn, and P. A. Nelson, "Sound field analysis using a complex-valued neural network," in *Proc. Institute. Acoust.*, (Bristol, UK), Nov. 2024.

[8] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," *arXiv preprint arXiv:1705.09792*, 2018.

[9] C. Lee, H. Hasegawa, and S. Gao, "Complex-valued neural networks: A comprehensive survey," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 8, pp. 1406–1426, 2022.

[10] X. Luan, M. Olivieri, M. Pezzoli, F. Antonacci, and A. Sarti, "Complex-valued physics-informed neural network for near-field acoustic holography," in *Proc. 32nd Eur. Signal Process. Conf. (EUSIPCO)*, (Lyon, France), pp. 126–130, Aug. 2024.

[11] V. S. Paul and P. A. Nelson, "Efficient design of complex-valued neural networks with application to the classification of transient acoustic signals," *J. Acoust. Soc. Am. (JASA)*, vol. 156, no. 2, pp. 1099–1110, 2024.

[12] G. Kim, D. K. Han, and H. Ko, "Sound source localization using complex-valued deep neural networks," in *Proc. IEEE Int. Conf. Consumer Electronics (ICCE)*, (Las Vegas, NV, USA), pp. 1–4, Jan. 2024.

[13] mh acoustics, *em64 Eigenmike release notes*. 25 Summit Ave Summit, NJ 07901, December 2023.

[14] P. Virtue, S. X. Yu, and M. Lustig, "Better than real: Complex-valued neural nets for mri fingerprinting," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, (Abu Dhabi, UAE), pp. 3953–3957, Oct. 2017.

[15] S. Elliott, *Signal processing for active control*. Elsevier, 2001.

[16] V. Paul, N. Hahn, and P. Nelson, "Representation learning in complex-valued neural networks for sound field reconstruction," in *Proc. 49th German Annu. Conf. Acoust. (DAGA)*, (Copenhagen, Denmark), Mar. 2025.

[17] X. Hong, B. Du, S. Yang, M. Lei, and X. Zeng, "End-to-end sound field reproduction based on deep learning," *J. Acoust. Soc. Am. (JASA)*, vol. 153, no. 5, pp. 3055–3055, 2023.