DOI: 10.61782/fa.2025.0245

FORUM ACUSTICUM
EURONOISE 2025

COMPLEX-VALUED NEURAL NETWORKS FOR THE REPRODUCTION OF LINEARLY SUPERPOSED SOUND FIELDS

Isaac. J. Lambert* **Vlad. S. Paul** **Philip. A. Nelson**
University of Southampton
Institute of Sound and Vibration Research
ijl1n21@soton.ac.uk

ABSTRACT

Previous numerical experiments have shown how a complex-valued neural network can be trained on a sampled recorded sound field to derive loudspeaker signals which reproduce the target sound field. Furthermore, some negative effects of spatial aliasing can be overcome by first training the network with a higher density of spatial samples than is used in the specification of the target sound field for reproduction. This amounts to increasing the size of the output layer of the network relative to the input. This work investigates whether such a neural network can also reproduce target sound fields consisting of the linear superposition of a number of plane waves. Numerical experiments are first carried out in which the neural network is trained on sound fields containing only single plane waves, and the reproduction of sound fields containing multiple plane waves is compared to the sum of reproductions of single plane waves. In a further numerical experiment, the neural network is trained using sound fields containing multiple plane waves. The reproductions produced by this network are again compared to the sum of reproductions of single plane waves. Finally, the effect of using different network structures on linear superposition is investigated.

Keywords: *machine learning, virtual audio, spatial aliasing*

**Corresponding author*: ijl1n21@soton.ac.uk.

1. INTRODUCTION

Sound field reproduction methods aim to accurately recreate a target acoustic pressure field using discrete loudspeakers, based on sampling of the target field at finite, discrete locations. A fundamental feature of acoustic pressure fields is that of linear superposition, which states that sound fields can be combined by summation. Effective sound field reproduction methods should also follow this principle, meaning that the reproduction of a sound field comprising the sum of multiple sound fields should be equal to the sum of the reproduction of the individual fields.

Wave field synthesis [1] is one early method of sound field reproduction. This method is based on Huygens's principle, which states that points on an incoming acoustic wave front can be modelled as secondary sources. Therefore, by measuring the pressure and pressure gradient at points along the wavefront, the propagation of the wave front can be reproduced by secondary sources. Furthermore, because the onward propagation of the wave front is known, the pressure and pressure gradient can also be extrapolated at other locations, allowing source placement at locations other than the measurement. A sound field can also be reproduced by deriving source strengths which result in the least squared error between the target and reproduced pressure at the location of sensors [2, 3]. This is an optimization process in which the least squares cost function is minimized, and because it involves matrix inversion a regularization parameter is usually included in this method. The least absolute shrinkage and selection operator (LASSO) cost function includes a sparsity term which minimises the number of

non-zero source strengths. This has been applied to sound field reproduction in order to reduce the number of active loudspeakers [4].

A feature of all aforementioned methods is spatial aliasing. This occurs for reproduction of frequencies above the spatial aliasing frequency, at which half a wavelength is equal to the distance between sensors. Therefore, increasing the density of sensors results in a higher spatial aliasing frequency.

Recently, neural networks have been applied to sound field reproduction problems. Recorded pressure signals have been used as input features to a convolutional neural network [5], which achieved reduced reproduction error when compared to the LASSO and least squares pressure matching method, particularly close to the spatial aliasing frequency. Furthermore, in cases with irregular source arrangements, the use of a convolutional neural network resulted in reduced reproduction error above the spatial aliasing when compared to the least squares pressure matching method [6].

Rather than processing the real and imaginary parts of complex acoustic pressure separately, a complex-valued multilayer perceptron (cMLP) [7] is capable of processing the real and imaginary parts of complex numbers together. This ensures that phase information is maintained throughout processing. Previous work [8] has shown how a complex-valued multilayer perceptron can effectively reproduce single-frequency plane-wave sound fields. However, the ability of this neural network to reproduce sound fields consisting of the linear superposition of multiple plane waves has not yet been investigated.

This paper investigates the ability of a cMLP to reproduce sound fields consisting of two plane waves over a range of angles of incidence. A neural network will first be trained on single frequency single plane waves, and another network on single frequency sound fields consisting of two plane waves where a large number of combinations of angles of incidence are used. The reproduction of the second network will be compared to the linear summation of reproductions produced by the first network. Furthermore, the ability of the first network to reproduce double plane waves is tested. This procedure is then repeated with networks which are trained using a greater number of sensors than are used to record the sound field for reproduction. This corresponds to a larger output layer size of the neural network than input. The details of the numerical experiments and neural networks used are given in the Methods section. The Results and Discussion section outlines the reproduction errors of the different methods across frequency and angle of incidence. Finally, the paper concludes by outlining the main findings and their implications for future work.

2. METHODS

2.1 Experiment 1

The complex pressure of a single plane wave $p_{single}(\omega, \theta, x, y)$ with wavenumber vector $\mathbf{k}_\theta(\omega, \theta)$ and phase Φ, at location $\mathbf{x}(x, y)$ is calculated as

$$p_{single}(\omega, \theta, x, y) = \mathrm{e}^{-\mathrm{j}(\mathbf{k}_\theta(\omega,\theta)\cdot\mathbf{x}(x,y)+\Phi)} \tag{1}$$

where θ is the angle of incidence of the plane wave and the wavenumber vector $\mathbf{k}_\theta(\omega, \theta)$ is calculated from the wavenumber $k(\omega)$ as

$$\mathbf{k}_\theta(\omega, \theta) = k(\omega)\cos(\theta)\mathbf{i} + k(\omega)\sin(\theta)\mathbf{j} \tag{2}$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors in the x and y direction. The complex pressure of a sound field p_{double} consisting of two plane waves with angles of incidence θ_1 and θ_2, where one plane wave has phase Φ_1 and the other has phase Φ_2 is

$$\begin{aligned} p_{double}(\omega, \theta, x, y) = \mathrm{e}^{-\mathrm{j}(\mathbf{k}_{\theta_1}(\omega,\theta_1)\cdot\mathbf{x}(x,y)+\Phi_1)} \\ +\mathrm{e}^{-\mathrm{j}(\mathbf{k}_{\theta_2}(\omega,\theta_2)\cdot\mathbf{x}(x,y)+\Phi_2)}. \end{aligned} \tag{3}$$

In the first experiment, a complex-valued neural network was trained using single-frequency sound fields calculated as in Eqn. (1) consisting of single plane waves with angles of incidence evenly spaced between $-180°$ and $180°$. A set of 120 plane waves was used, giving an angular resolution of $3°$. The phase of each plane wave in the data set was random. $20\,\%$ of the sound fields were randomly selected as validation data. The source and receiver layout used for the simulated reproduction was as follows. Four monopole sources were arranged to describe the corners of a $4\,\mathrm{m}$ x $4\,\mathrm{m}$ square. Within this region, 16 monopole receivers were arranged evenly across the central $1\,\mathrm{m}$ x $1\,\mathrm{m}$ square.

The input layer of the neural network consisted of 16 neurons, corresponding to the 16 complex pressures for each plane wave sampled at the 16 receivers. A single hidden layer of size 200 neurons with a complex cardioid

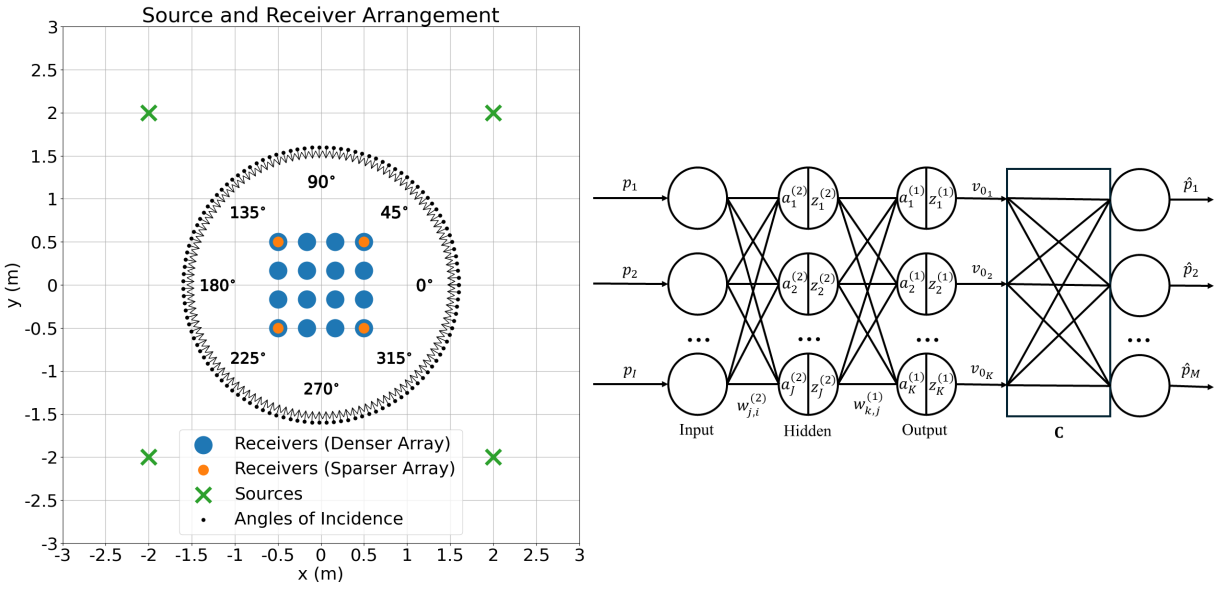

Figure 1. Geometry of sources, receivers and incident plane waves used in numerical experiments and architecture of neural network. Input data is target complex pressure $\mathbf{p}$ at sensors, corresponding to dense array in experiments 1 and 2 and sparse array in experiment 3. Output data corresponds to reproduced complex pressure at sensors $\hat{\mathbf{p}}$, taken from dense array in all cases. Cost function is calculated from $\mathbf{p}$ and $\hat{\mathbf{p}}$, taken from dense array of sensors in all cases.

activation function was used. Finally, the output layer consisted of 4 neurons and no activation function. The output size corresponded to the 4 monopole sources, and the outputs were multiplied by the matrix $\mathbf{C}$ of transfer functions between the sources and receivers. The cost function between the target and reproduced pressures could then be calculated and back-propagated through the network. The network was trained over 50 epochs using a batch size of 4. The mathematics specifying the cMLP followed the process outlined by Paul & Nelson [7] and the process to extend the network to include acoustic propagation is the same as in the previous paper [8]. Figure 1 illustrates the source and receiver arrangement as well as the neural network architecture used. Here, $\mathbf{p}$ is the complex pressures at the I input sensors and $\hat{\mathbf{p}}$ is the complex pressures at the M output sensors. In the case of the first experiment, both I and M were 16, meaning all 16 sensors were used in both the training and the specification of the sound field for reproduction. $\mathbf{w}^{(2)}$, $\mathbf{a}^{(2)}$, $\mathbf{z}^{(2)}$ are the complex weights, activation function and output of the hidden layer and $\mathbf{w}^{(1)}$, $\mathbf{a}^{(1)}$, $\mathbf{z}^{(1)}$ are the complex weights, activation function and output of the output layer. Figure 1 also shows how the angular coordinate system used to define the angle of incidence of plane waves relates to the source and receiver geometry used. For example, an angle of incidence of $0°$ refers to a plane wave travelling from the right of the sensor array to the left. The angles of incidence used in the numerical experiments are shown.

Following training, the reproductions by the neural network were compared to the target sound fields at 225 positions evenly distributed across the target $1\,\mathrm{m}$ x $1\,\mathrm{m}$ square. Microphone noise was simulated by adding a random value drawn from a Gaussian distribution with zero mean and a standard deviation of 0.1 to the real and

imaginary parts of the input data. This testing process was repeated over 100 trials and the average error across trials taken. In each trial, a reproduction of double plane wave sound fields was calculated from the reproduction of single plane wave sound fields by linear summation. For example, the reproduction of a target sound field containing two plane waves, one with angle of incidence 0° and the other with angle incidence 45° would be calculated by adding together the reproduction of these two plane waves individually. For each value of θ_1, a particular random phase ϕ_1 was applied, and a particular random ϕ_2 was applied for each θ_2. To generate the necessary single plane wave reproductions with angles of incidence θ_2 and phase ϕ_2, a second network was trained on a data set with these phase values. This was because the network was not assumed to generalise to reproduction of plane waves with phase different to the training data. The mean-squared error of these reproductions was then calculated across the 225 test positions. The mean-squared error E_g was calculated from

$$E_g = \frac{1}{2}\frac{(\mathbf{p}_g - \hat{\mathbf{p}}_g)^{\mathrm{H}}(\mathbf{p}_g - \hat{\mathbf{p}}_g)}{\mathbf{p}_g^{\mathrm{H}}\mathbf{p}_g} \tag{4}$$

where $\mathbf{p}_g$ and $\hat{\mathbf{p}}_g$ are vectors of the target and reproduced pressure at the 225 points. The subscript g indicates that this error is calculated from the ground truth at all 225 test positions. The error was normalised by the squared magnitude of the target pressure $\mathbf{p}_g^{\mathrm{H}}\mathbf{p}_g$. In the results, this is referred to as the "sum of singles" case.

The ability of this network to reproduce sound fields consisting of two plane waves was also tested. In this case, the input data consisted of the complex pressure at the 16 sensors, calculated according to Eqn. (3). For each of the 120 evenly spaced values of θ_1, a sound field was generated for values of $\theta_2 \geq \theta_1$. The result of this being that the input data did not include pairs of sound fields for which θ_1 in one was equal to θ_2 in the other. For each value of θ_1 and θ_2 a particular random phase ϕ_1 and ϕ_2 was applied. The double plane wave sound fields were corrupted with noise as before and propagated through the network trained on single plane waves to derive an estimate of the complex pressure at the 225 target locations. The reproduction error was calculated at these locations and the testing process repeated 100 times. This is referred to as the "trained on single" case in the results.

As a benchmark, optimal source strengths were also calculated according to the least squared error method. The source strengths were calculated using the 16 receiver signals. The optimal source strengths $\mathbf{v}_0$ were calculated from

$$\mathbf{v}_0 = [\mathbf{C}^{\mathrm{H}}(\omega)\mathbf{C}(\omega) + \beta\mathbf{I}]^{-1}\mathbf{C}^{\mathrm{H}}(\omega)\mathbf{p}(\omega) \tag{5}$$

where β is a regularisation parameter with a value of 0.001. In the results for experiments 1 and 2, this is referred to as the "least squared error" case.

2.2 Experiment 2

This procedure made use of the same source and receiver arrangement as the previous numerical experiment, as well as the same neural network structure. In this instance a neural network was trained using input data consisting of double plane wave sound fields. The double plane wave sound fields were calculated in the same way as in Experiment 1. As in the previous method, the average reproduction error was tested over 100 trials of double plane wave data sets corrupted with noise. Again, the results were compared to the least squared error solution using the 16 receiver signals. This is referred to as the "trained on double" case in the results section.

2.3 Experiment 3

In the final numerical experiment, a different neural network structure was used. The input layer consisted of four neurons, which corresponded to the complex pressure at the four corner receivers for each sound field. The rest of the network structure was unchanged so that all 16 receivers were used in calculating the cost function, meaning that $I = 4$ and $M = 16$. Both the procedures outlined in experiment 1 and experiment 2 were carried out again using this network structure. In this case, the benchmark methods were the least squared error method using the 4 receiver signals and the least squared error method using the 16 receiver signals.

3. RESULTS AND DISCUSSION

3.1 Experiments 1 and 2

The overall reproduction error of each target sound field is taken as the mean-average squared error across 225 points, evenly distributed across the $1\,\mathrm{m}$ x $1\,\mathrm{m}$ target reproduction area. The errors MSE are given in dB, calcu-

lated from E such that

$$MSE = 10\log_{10}(E) \tag{6}$$

Averaging the reproduction errors across all sound fields for each frequency gives an overall measure of the performance of the methods as a function of frequency. Figure 2 shows the mean squared error of each method as a function of frequency. The general form of the results demonstrate spatial aliasing. As frequency increases, the mean squared error of the methods increases. Above the spatial aliasing frequency (515 Hz), the error remains high and consistent. Comparing the error of the four methods as a function of frequency, there is little variation between the four. Above spatial aliasing, the error produced by network trained on single plane waves is marginally lower, as is the least squared error method.

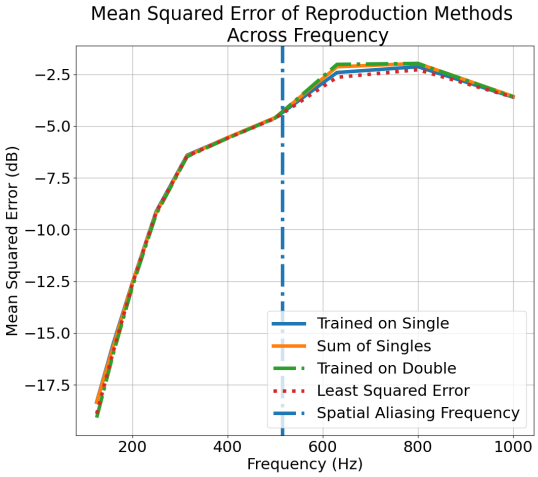

Figure 2. Squared error of reproduction methods, averaged across angles of incidence of plane waves, as a function of frequency

The effect of angle of incidence can be investigated by plotting the reproduction error across angle of incidence. Because each sound field consists of two plane waves, for each plot the angle of incidence of the "first" plane wave is held constant whilst the "second" varies between $-180°$ and $180°$ at 120 angles of incidence. Figure 3 shows the reproduction error of 160 Hz (left) and 1000 Hz (right) plane waves with the first angle of incidence being $45°$.

In the 160 Hz case, the reproduction error of all methods is low for angles of incidence of $-135°$, $-45°$, $45°$ and $135°$. This is because at these angles the direction of the plane waves are close to the location of a source. In the same way, the error is high when the angle of incidence is $-180°$, $-90°$, $0°$ and $90°$, when the direction is between sources. When the angle of incidence of the two plane waves is similar, between around $0°$ and $90°$, the reproduction error becomes very variable. This suggests that the particular phase relation between the two plane waves has a particularly strong effect in this case. The reproduction error of the methods is similar across angles of incidence. In the 1000 Hz case, the overall reproduction error is higher as a consequence of spatial aliasing. Furthermore, the regions over which the reproduction error is high (where the angle of incidence is between sources) is much wider in this case.
Figure 4 shows the reproduction error of 160 Hz (left) and 1000 Hz (right) plane waves with the first angle of incidence being $-180°$, which corresponds to a direction between sources. Once again, the performance of the different methods is similar. At 160 Hz, there is strong variation as a result of the phase relation between the two plane waves. In the 1000 Hz case, once again the reproduction error is generally higher due to spatial aliasing. As with the $45°$ case, the reproduction methods each had similar reproduction errors. The similarity in performance between the neural network trained on single plane waves and double plane waves is particularly significant. The network trained on single plane waves required a smaller training data set and was therefore substantially faster to train, but still produced acceptable results. Therefore, the network trained on single plane waves was able to generalise to the more complicated double plane wave case for the sound fields studied in this paper. As an illustrative example, Figure 5 shows the ground truth and reproduced sound fields for a 315 Hz double plane wave, which is below the spatial aliasing frequency. The reproduced sound fields are clearly similar between the three methods shown. The network trained on single plane waves is able to produce a sound field which is substantially different to any single plane wave, demonstrating how the network is able to generalise to cases which are substantially different to the training data.

3.2 Experiment 3

Throughout this subsection, all neural networks have a structure consisting of an input size of 4 and an output

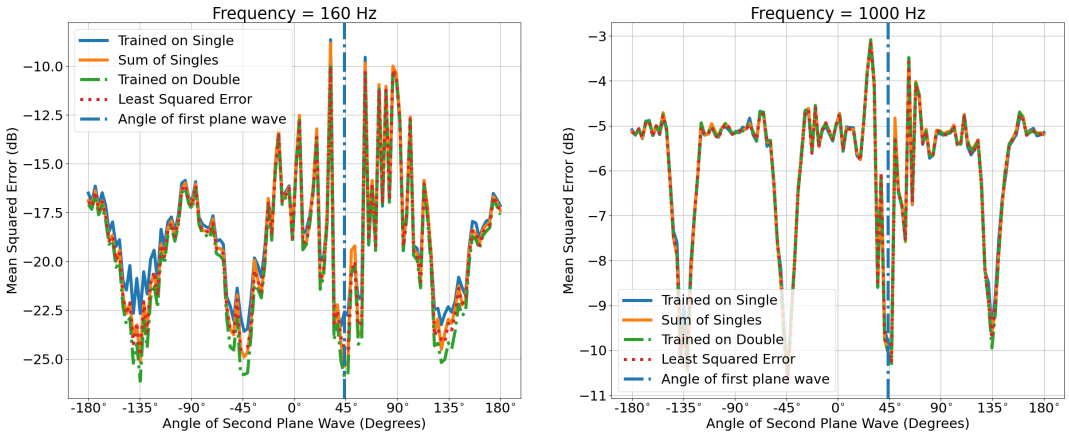

Figure 3. Squared error of reproduction methods for sound fields consisting of 160 Hz (left) and 1000 Hz (right) plane waves. First wave has angle of incidence 45°, second plane wave is plotted across angle of incidence.

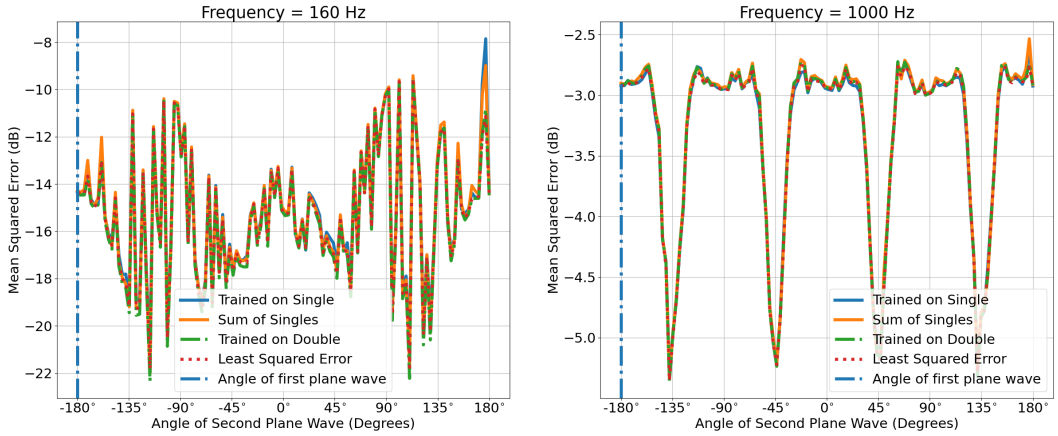

Figure 4. Squared error of reproduction methods for sound fields consisting of 160 Hz (left) and 1000 Hz (right) plane waves. First wave has angle of incidence -180°, second plane wave is plotted across angle of incidence.

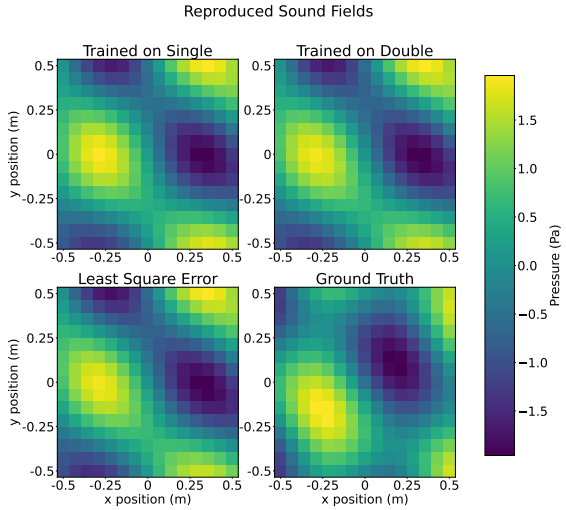

Figure 5. Ground truth and reproduced sound fields for an example 315 Hz double plane wave sound field. Reproduction methods shown are neural network trained on single and double plane waves, and least squared error method.

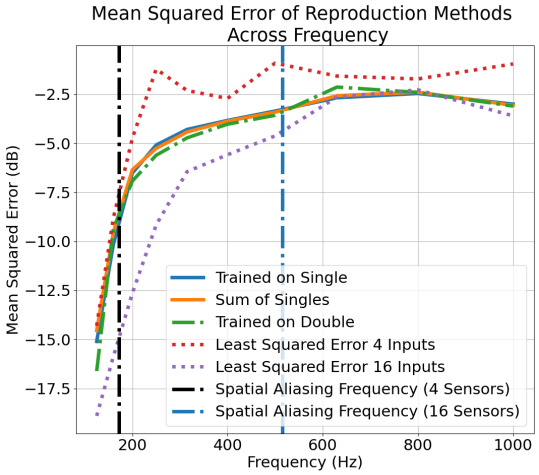

Figure 6. Squared error of reproduction methods, averaged across angles of incidence of plane waves, as a function of frequency

size of 16. In this way the networks were trained on all 16 sensor errors, but the estimations were based on only the 4 input sensors. Figure 6 shows the reproduction error at each frequency, averaged across all target sound fields. These results demonstrate how the reproduction error of the neural networks depends on the spatial aliasing frequency of the input and output data. At low frequency, the error of the neural networks is close to the least squared error method with 4 inputs. This means that here the performance is determined mostly by the input data. However, above the spatial aliasing frequency of the 4 sensor array (172 Hz), the error of the neural networks diverges from the least squared error method with 4 sensors and becomes closer to the least squared error method with 16 sensors. At these higher frequencies, the neural networks therefore achieve improved performance as a result of being trained using more microphone signals. Above the spatial aliasing frequency of the 16 sensor array (515 Hz), the error of the neural networks is very close to the error of the least squared error method with 16 sensors. At these frequencies, therefore, the performance is mostly determined by the output (training) data.

Figure 7 shows the error of the reproduction methods as a function of angle of incidence of the second plane wave, where the angle of the first 160 Hz (left) and 1000 Hz (right) plane wave is $45°$. In the 160 Hz case, the reproduction error of the neural networks is generally between the error of the two least squared error methods, but is closer to that of the least squared error method with 4 sensors. Once again, the error of the methods varies strongly depending on the phase relation between the two plane waves. Around $-45°$ and $135°$ the error of the neural network trained on double plane waves is lower compared to the other networks. At 1000 Hz, the reproduction error of the neural networks once again lies between the error of the two least squared error methods. These results demonstrate the same features of spatial aliasing seen in Figure 3 and Figure 4.

4. CONCLUSIONS

A complex-valued multilayer perceptron trained on sound fields consisting of the linear superposition of two plane waves gives results that are very similar to the linear superposition of the reproductions of single plane waves by a complex-valued multilayer perceptron trained on sound fields consisting of single plane waves. A neural network trained on single plane waves also results in

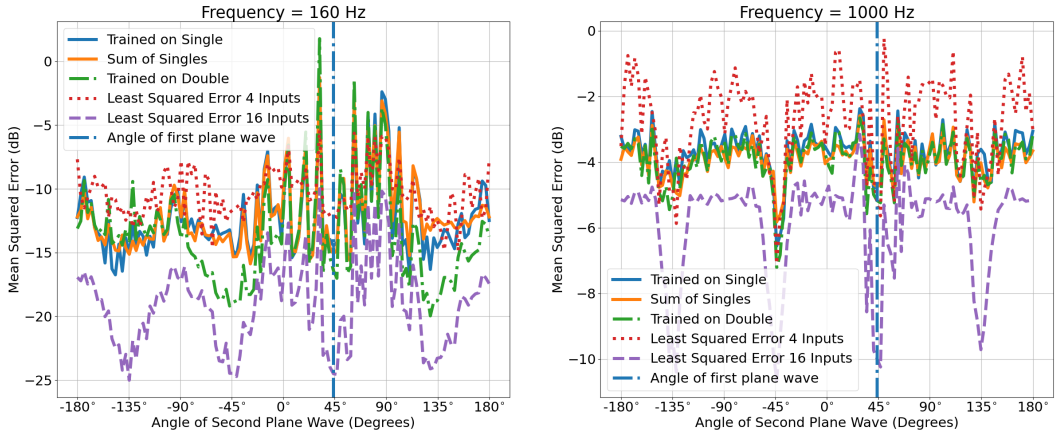

Figure 7. Squared error of reproduction methods for sound fields consisting of 160 Hz (left) and 1000 Hz (right) plane waves. First wave has angle of incidence 45°, second plane wave is plotted across angle of incidence. Neural networks trained using 4 inputs and 16 outputs

similar reproduction error when reproducing sound fields consisting of two plane waves. Therefore, a neural network trained on these simpler sound fields is able to generalise to some extent to more complicated sound fields.

A neural network can be trained with a larger output layer than input layer size. This means that a larger number of microphone signals are used in training than are used to record sound fields for reproduction. This network structure results in performance which is somewhere between the least squared error method using each of the sensor arrangements. In particular, between the spatial aliasing frequency of the two sensor arrangements, the neural network achieves performance which is notably better than the least squared error method using fewer inputs.

5. REFERENCES

[1] A. J. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, no. 5, p. 2764, 1993.

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

[3] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

[4] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, 2010.

[5] X. Hong, B. Du, S. Yang, M. Lei, and X. Zeng, "End-to-end sound field reproduction based on deep learning," *The Journal of the Acoustical Society of America*, vol. 153, no. 5, p. 3055, 2023.

[6] L. Comanducci, F. Antonacci, and A. Sarti, "Synthesis of soundfields through irregular loudspeaker arrays based on convolutional neural networks," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, 2024.

[7] V. S. Paul and P. A. Nelson, "Efficient design of complex-valued neural networks with application to the classification of transient acoustic signals," *The Journal of the Acoustical Society of America*, vol. 156, no. 2, pp. 1099–1110, 2024.

[8] I. J. Lambert, V. S. Paul, and P. A. Nelson, "Complex-valued neural networks for the reproduction of single frequency sound fields," in *Reproduced Sound, Institute of Acoustics Proc.*, vol. 46, (Bristol, UK), 2024.