Audio Engineering Society
Convention Express Paper

Presented at the 156th Convention
2024 June 15-17, Madrid, Spain

Interpolation of loudspeaker level balloons from polar measurements by using deep learning

Víctor Manuel Catalá Iborra[1,2]

[1] *DAS Audio Group, Fuente del Jarro, Spain*

[2] *Universitat de València, València, Spain*

Correspondence should be addressed to Víctor Catalá (vcatala@dasaudio.com)

ABSTRACT

Complete radiation balloons are needed to analyze the performance of loudspeakers and to perform accurate tunning and electroacoustic predictions of loudspeaker systems. A method is proposed to obtain full radiation balloons from horizontal and vertical polar measurements by using U-Net, a deep learning architecture widely applied for image processing. Mean absolute errors lower than 4 dB were obtained on test data.

1 Introduction

Radiation patterns are fundamental in the design process of loudspeakers, both as targets and as real measurements to verify their performance and to tune them properly. Electroacoustic simulation software based on loudspeaker models relies on radiation data to perform calculations of SPL and other acoustic parameters. These data should include complete radiation balloons to provide accurate results. A 5º resolution on Phi, Theta angles in spherical coordinates is proposed by the Loudspeaker polar radiation measurements AES standard [1] to fully characterize a loudspeaker radiation.

While the measurement of horizontal and vertical polars of a loudspeaker can be run with a single turntable in a relatively short time, the measurement of full radiation balloons requires complex acquisition devices to rotate the loudspeaker or the microphones, with long set up time, providing a large set of measurements on the surface of a sufficient distant sphere. The number of measurements and the time needed is significantly longer than for only horizontal-vertical polar measurements. Another alternative is the use of a near field scanner [2] , [3] which is not affordable for many companies or institutions, and also produce large datasets with long acquisition times.

Having a function to map 3D SPL balloons from polar measurements could be convenient in terms of development time, equipment and involved data volume. Neural Networks, as universal approximators [4], are considered in this work for the task. While great efforts have been done to apply Neural Networks to predict or interpolate Head Related Transfer Functions (HRTF) from sparse measurements [5], [6], [7], [8], these methods have not been applied to loudspeaker radiation prediction, as far as this author knows. This work presents a method to approximate full 3D SPL balloons from horizontal and vertical measurements by using convolutional neural networks.

2 Method

In the proposed method, radiation balloons are represented as images showing SPL in colours as function of frequency (1/6 octave centre frequencies approx. starting at 125 Hz) in the x axis and the different measurement points in the y axis as rows. Measurements with only polar data show the response at vertical and horizontal corresponding polar points. This way, the problem has been reduced to an incomplete-image recovery problem.

A dataset of images has been created from full balloon measurements acquired as described in [9]. Input data images contain information only at the rows representing data at Phi= 0º, 90º, 180º, 270º (polar measurements), and Theta=0º, 180º (redundant on-axis and rear measurements), with any other rows set to zero. Output images contain data of all measurement points. Measurements are re-arranged in spiral order as described at [7] . Figure 1 shows an example of a full radiation balloon transformed to be used as image data by the proposed neural network.

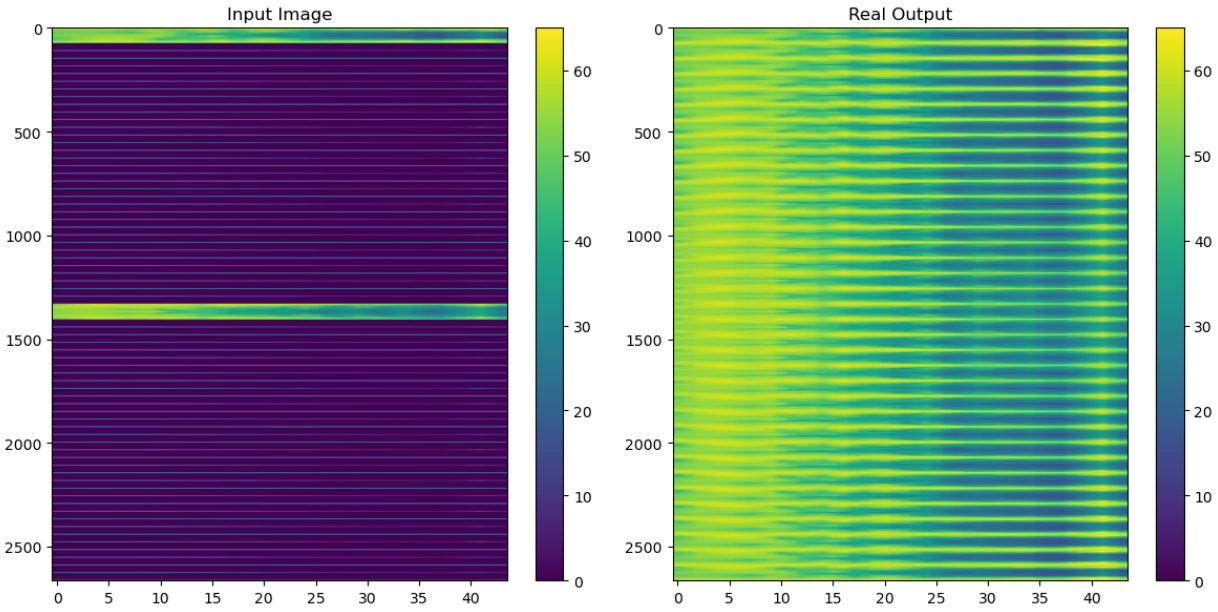

Figure 1. Example of input image (only polar data) and real output (full 3D balloon data).

In order to recover full images from only polar images, a convolutional neural network with an architecture similar to U-net [10] has been trained by using balloon data from 7 loudspeakers: 5 were use as training data, 1 as validation data, and 1 as test data (unseen by the neural network in the training process). U-net deep learning architecture has proved its efficiency at image problems with short data sets. Basically, it consists of an encoder path followed by a decoder one, with direct connections between them (skip connections) to pass features from encoders to decoders.

The architecture of the proposed neural network can be seen at Figure 2.

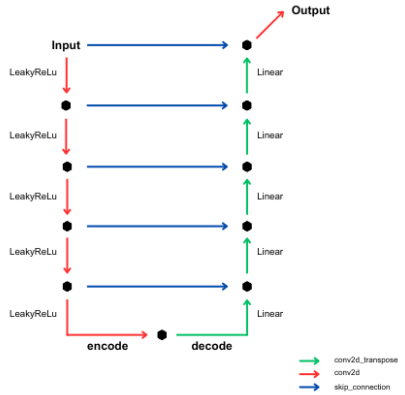

Figure 2. Proposed U-net architecture.

The parameters of the different layers of the proposed neural network are shown in Table 1.

Data were normalized between 0 and 1 for training. Activation functions where Leaky-Relu for encoder layers, Linear for decoder layers and Sigmoid for the output layer.

As input data already show level variation in dB, mean squared error (MSE) was used as loss to penalize large deviations.

Layer	Input size	Stride	Padding	Kernel size	Output size
conv2d_44	(2664, 44, 1)	(2, 1)	valid	(75, 3)	(1295, 42, 16)
conv2d_45	(1295, 42, 16)	(2, 2)	valid	(75, 3)	(611, 20, 32)
conv2d_46	(611, 20, 32)	(2, 2)	valid	(75, 3)	(269, 9, 64)
conv2d_47	(269, 9, 64)	(2, 2)	valid	(75, 3)	(98, 4, 128)
conv2d_48	(98, 4, 128)	(2, 2)	valid	(75, 3)	(12, 1, 256)
conv2d_transpose_32	(12, 1, 256)	(2, 1)	valid	(76, 4)	(98, 4, 128)
conv2d_transpose_33	(98, 4, 256)	(2, 2)	valid	(75, 3)	(269, 9, 64)
conv2d_transpose_34	(269, 9, 128)	(2, 2)	valid	(75, 4)	(611, 20, 32)
conv2d_transpose_35	(611, 20, 64)	(2, 2)	valid	(75, 4)	(1295, 42, 16)
conv2d_transpose_36	(1295, 42, 32)	(2, 1)	valid	(76, 3)	(2664, 44, 16)
conv2d_49	(2664, 44, 17)	NA	same	(75, 4)	(2664, 44, 1)

Table 1. Parameters of the proposed U-net.

3 Results

Figures 3-4 show training data with true and predicted outputs.

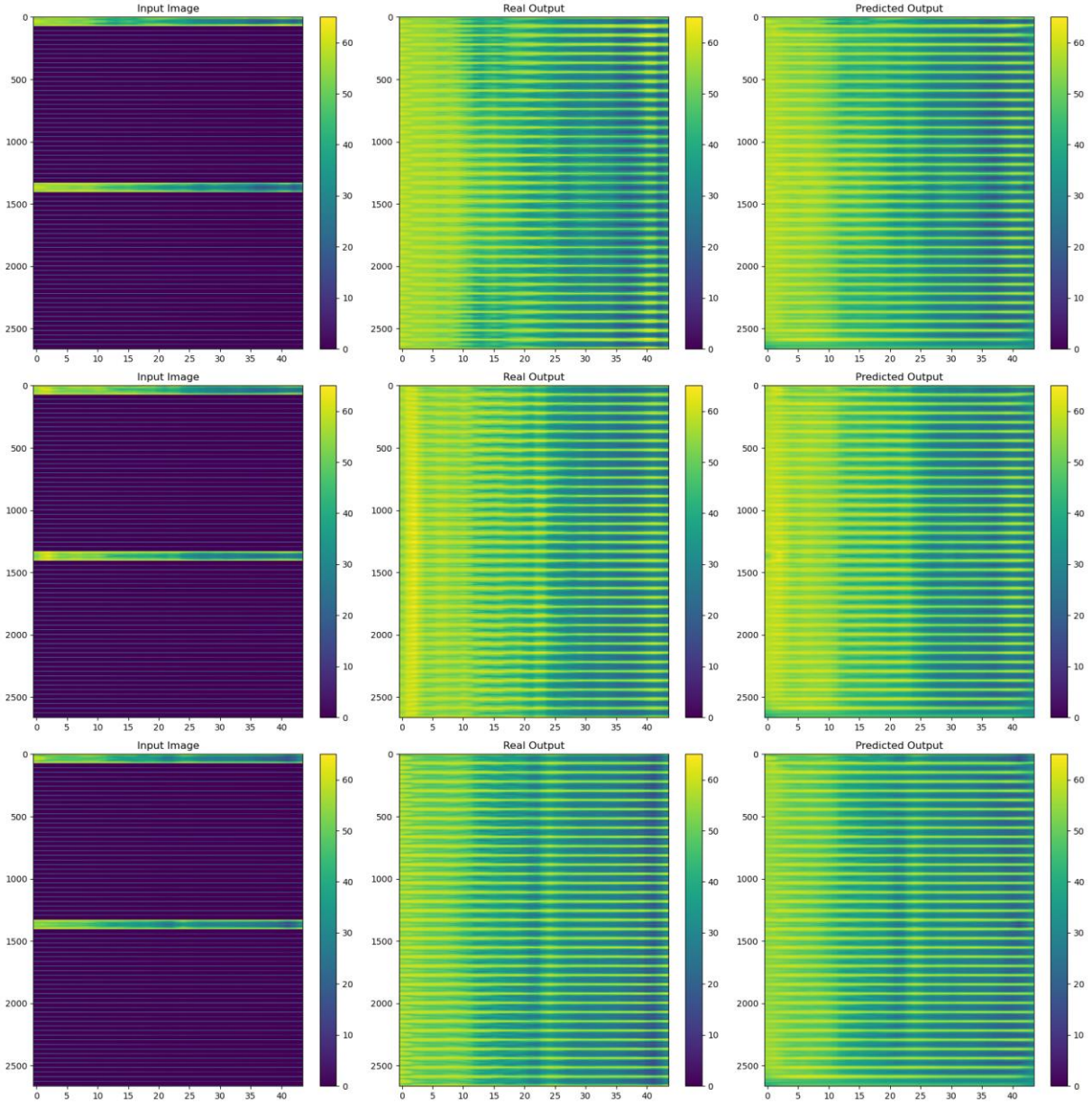

Figure 3. Training data images 1-2-3. Left column: only polar data. Central column: ground truth. Right: U-net prediction.

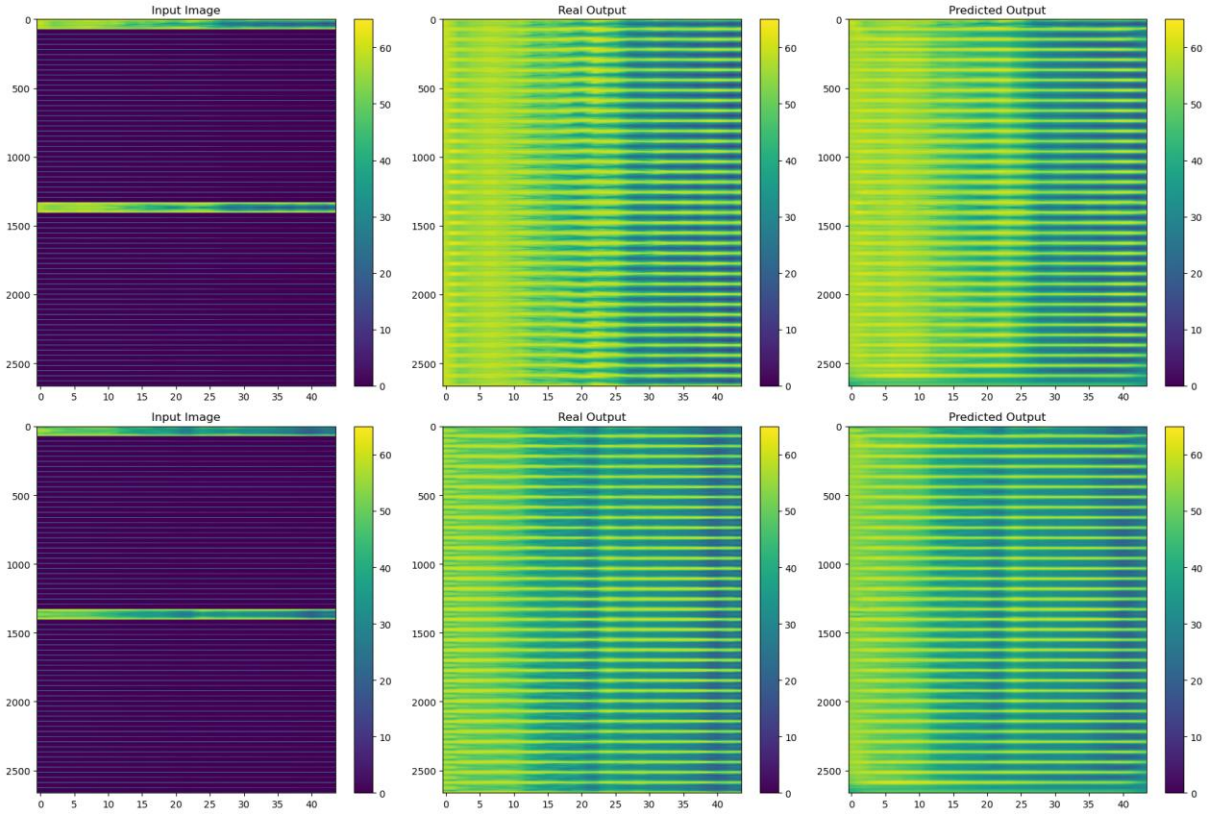

Figure 4. Training data images 4-5. Left column: only polar data. Central column: ground truth. Right: U-net prediction.

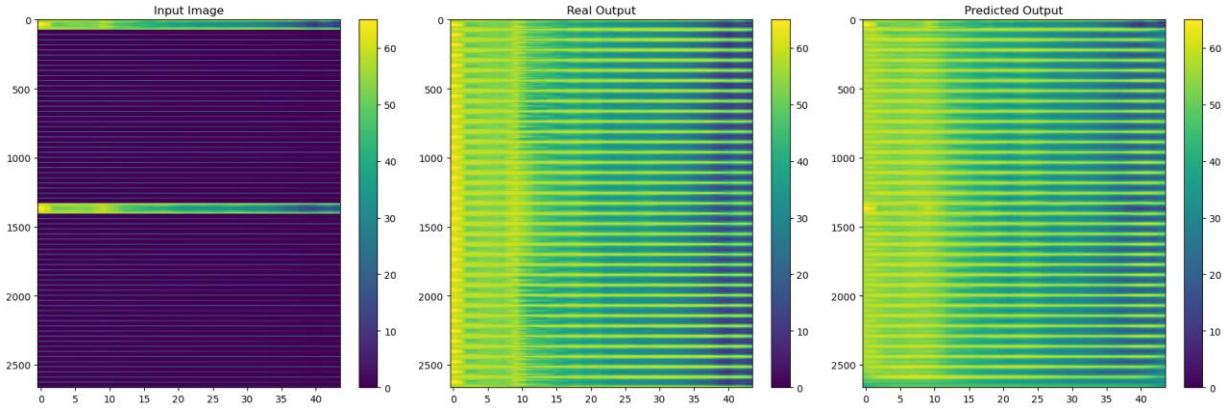

Figure 5. Validation data. Left column: only polar data. Central column: ground truth. Right: U-net prediction.

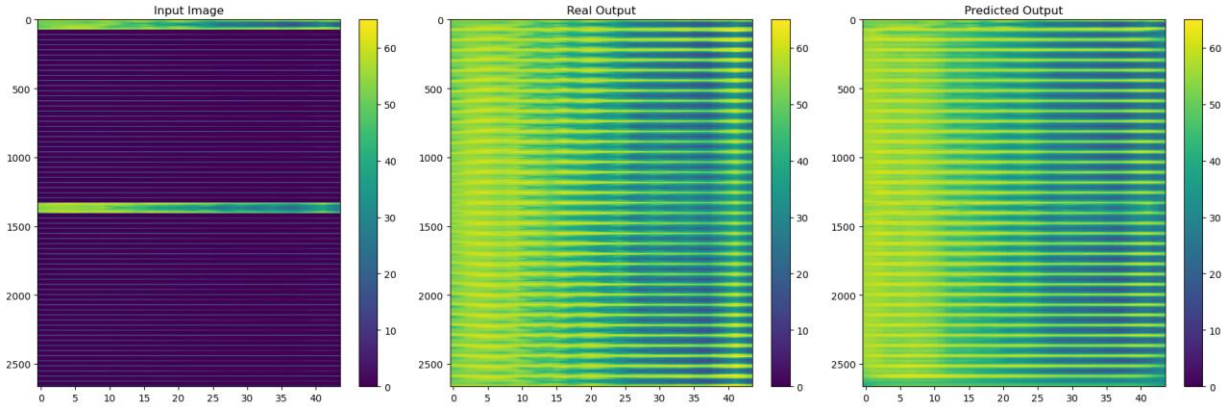

Figure 6. Test data. Left column: only polar data. Central column: ground truth. Right: U-net prediction.

Figures 5-6 present validation and test data predictions.

A first look shows that the neural network can approximately fill the incomplete images. To quantify the errors between true and predicted images, mean absolute errors (MAE) from MSE are listed below in Table 2:

Training data	2.8 dB
Validation data	4.0 dB
Test data	3.8 dB

Table 2. MAE, in dB, between predicted and ground truth images.

With the proposed method, a MAE lower than 4 dB was achieved on test data (unseen by the NN during training process). In this case, test data was a line array unit with strong differences between horizontal and vertical coverages at high frequencies.

For a more convenient understanding of results, polar plots at the plane defined by Phi=45º are shown at Figures 7-8-9-10-11-12-13.

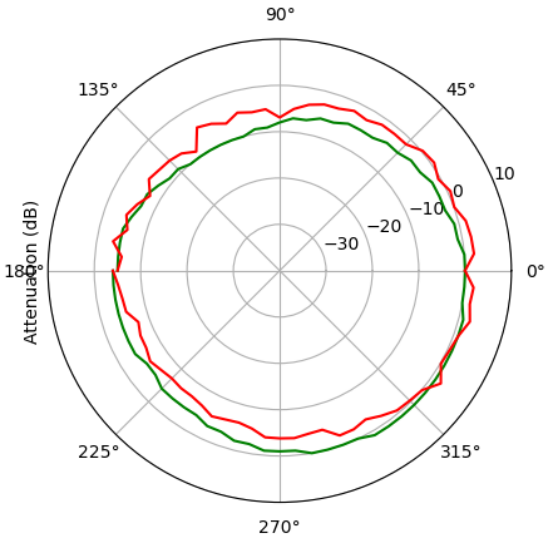

Figure 7. Polar plot at Phi=45º - 250 Hz.

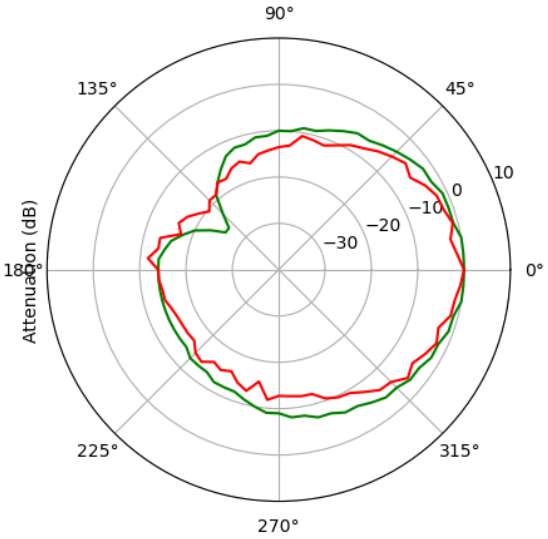

Figure 8. Polar plot at Phi=45º - 500 Hz.

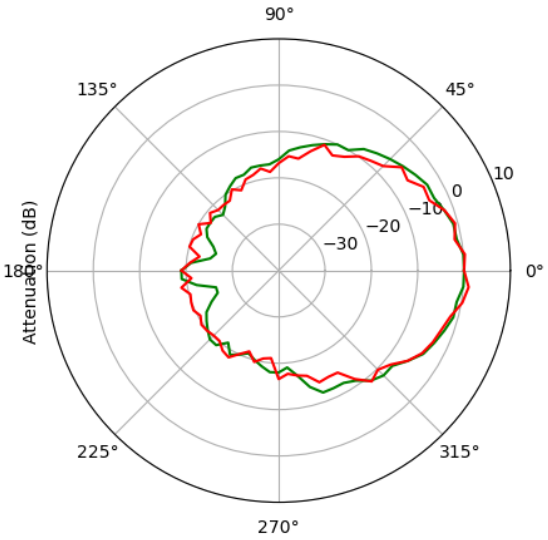

Figure 9. Polar plot at Phi=45º - 1 kHz.

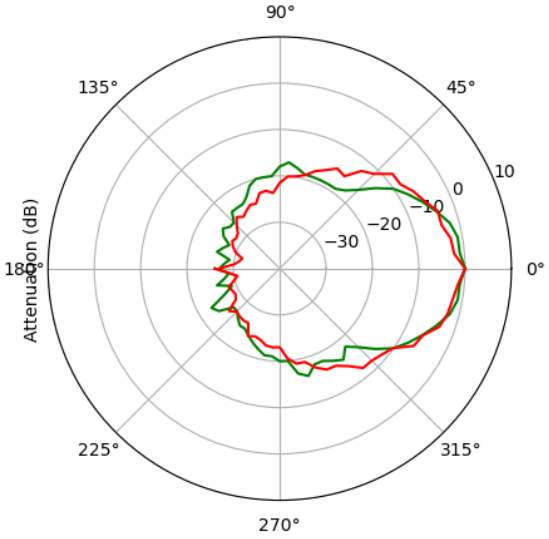

Figure 10. Polar plot at Phi=45º - 2 kHz.

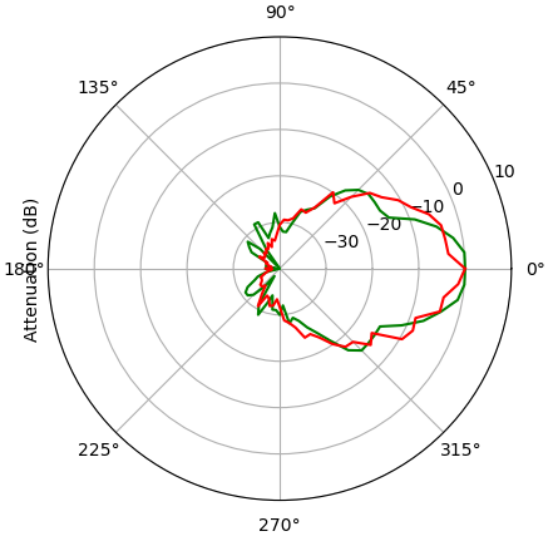

Figure 11. Polar plot at Phi=45º - 4 kHz.

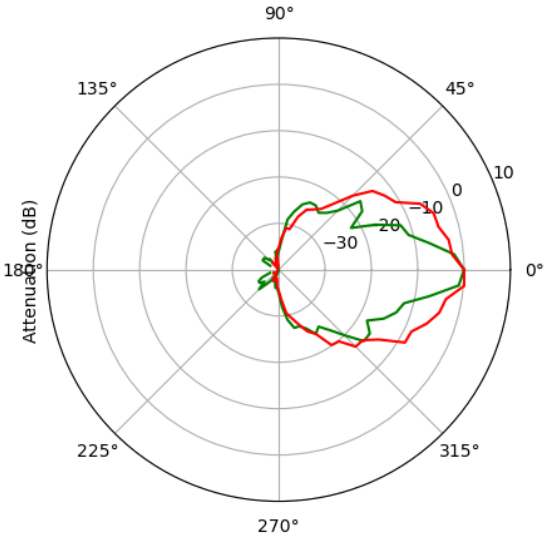

Figure 12. Polar plot at Phi=45º - 8 kHz.

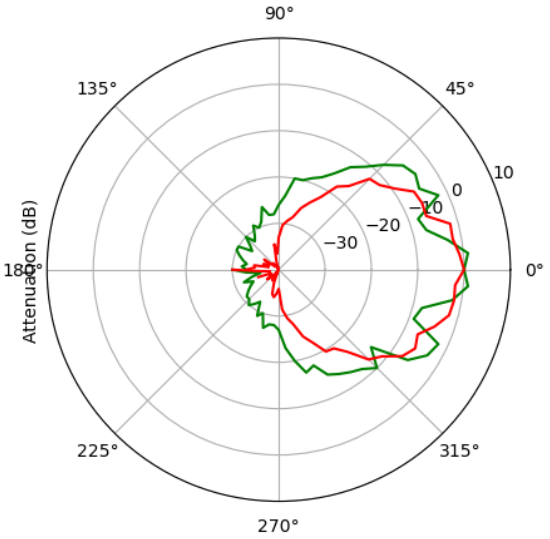

Figure 13. Polar plot at Phi=45º - 16 kHz.

The plane at Phi=45º (combined information of measurements at Phi=45º and Phi=225º) is just in the middle of the horizontal and vertical polar measurements and contains, therefore, the predictions at the farthest points from provided measurements.

4 Conclusions and further research

Neural Networks can be used to approximate level loudspeaker radiation balloons from polar measurements. In particular, U-net deep learning architectures can provide MAE lower that 4 dB over full balloons.

At high frequencies, with steeper variation of SPL against angles, the performance of the proposed network is lower. Refined architectures, more data and farther research are needed to improve accuracy.

Similar methods could be applied to phase interpolation in order to get complex responses.

More powerful models could be achieved by combining true polar measurements and full 3D simulations (from Finite Element Method software, for example) as input data, to obtain more accurate predictions. In this case, simple polar measurements would refine full 3D simulations.

Acknowledgement

Thanks to DAS Audio for the use of its facilities and its kind support.

Thanks to Emilio Soria for his suggestions.

References

[1] A.E.S., “Standard on acoustics - Sound source modeling - Loudspeaker polar radiation measurements,” *AES56-2008 (reaffirmed 2014)*, 2014.

[2] W. Klippel and C. Bellmann, “Holographic Nearfield Measurement of Loudspeaker Directivity,” in *141st Audio Engineering Society International Convention 2016*, Los Angeles, USA, 2016. Available: http://www.aes.org/e-lib.

[3] M. Bru, “Cylinder Measurement Method for Directivity Balloons,” in *149th Audio Engineering Society Convention 2020, AES 2020*, Audio Engineering Society, Oct. 2020. Available: http://www.aes.org/e-lib.

[4] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, 1989, doi: 10.1016/0893-6080(89)90020-8.

[5] B. Tsui, W. A. P. Smith, and G. Kearney, “Low-order spherical harmonic HRTF restoration using a neural network approach,” *Applied Sciences (Switzerland)*, vol. 10, no. 17, 2020, doi: 10.3390/APP10175764.

[6] F. Ma, T. D. Abhayapala, P. N. Samarasinghe, and X. Chen, “Physics

Informed Neural Network for Head-Related Transfer Function Upsampling," pp. 1–13, 2023. Available: http://arxiv.org/abs/2307.14650

[7] Z. Jiang, J. Sang, C. Zheng, A. Li, and X. Li, "Modeling individual head-related transfer functions from sparse measurements using a convolutional neural network," *J Acoust Soc Am*, vol. 153, no. 1, pp. 248–259, 2023, doi: 10.1121/10.0016854.

[8] S. S. Alotaibi and M. Wickert, "Modeling of Individual Head-Related Transfer Functions (HRTFs) Based on Spatiotemporal and Anthropometric Features Using Deep Neural Networks," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3358202.

[9] V. Catalá and F. Li, "DSP loudspeaker 3D complex correction," in *146th Convention of the Audio Engineering Society*, Dublin, Irland, 2019. Available: http://www.aes.org/e-lib.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. doi: 10.1007/978-3-319-24574-4_28.