Audio Engineering Society



Convention Express Paper

Presented at the 156th Convention

2024 June 15-17, Madrid, Spain

*This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author’s advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library* [*(http://www.aes.org/e-lib)*](http://www.aes.org/e-lib)) *all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

Interpolation of loudspeaker level balloons from polar measurements using U-net convolutional neural networks

Víctor Manuel Catalá Iborra1,2

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 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 convolutional neural network architecture used for image processing.

# Introduction

Radiation patterns are fundamental in the design process of loudspeakers, both as targets and as real measurements to verify their performance. 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 in order 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.

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 turn 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 scanners [2], which are not affordable for many companies or institutions, and also produce large datasets with long measurement times.

While great efforts have been done to apply Neural Networks to predict or interpolate Head Related Transfer Functions (HRTF) from sparse measurements [3], [4], [5], [6], 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.

# Method

In the proposed method, radiation balloons are represented as images showing SPL in colours as function of frequency 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. 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 [5] . Figure 1 shows an example of a full radiation balloon transformed to be used as image data by the proposed neural network.

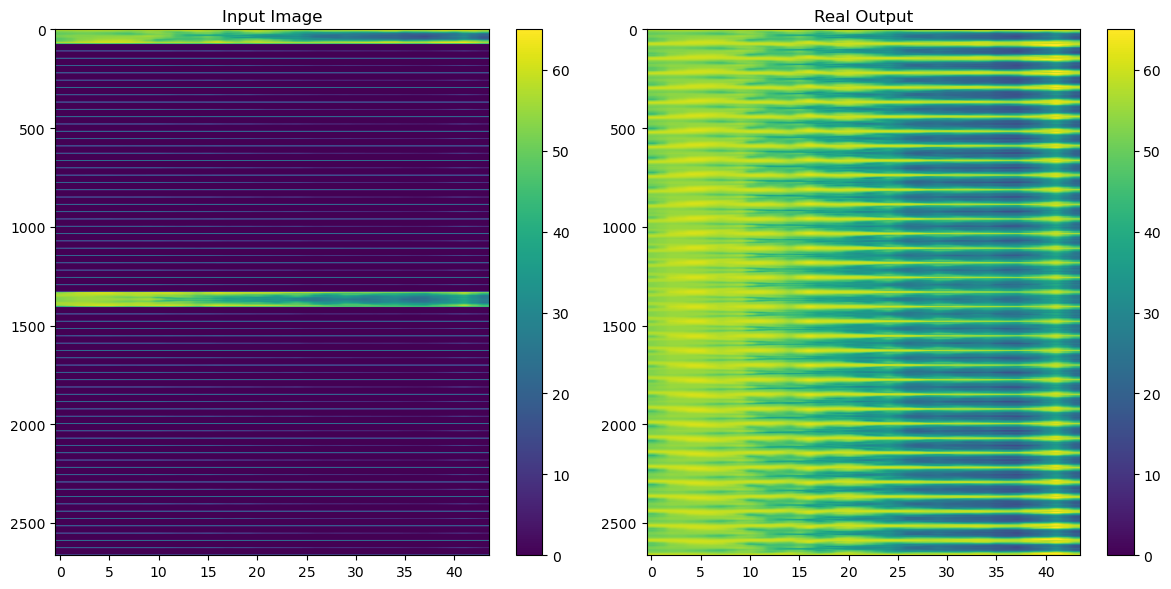


Figure 1. Example of input image (only polar data) and real output (full 3D balloon).

In order to recover full images from only polar images, a convolutional neural network with U-net architecture [7] 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 architecture has proved its efficiency with short data sets. Hablar un poco más sobre u-net.

The details of the proposed u-net can be seen at figure

Meter aquí arquitectura

Data were normalized between 0 and 1 for training. Activation functions where … and sigmoid for the output layer.

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

# Results

Figure 2 shows training data with actual and predicted output. A first look shows that the neural network can approximately fill the incomplete images.

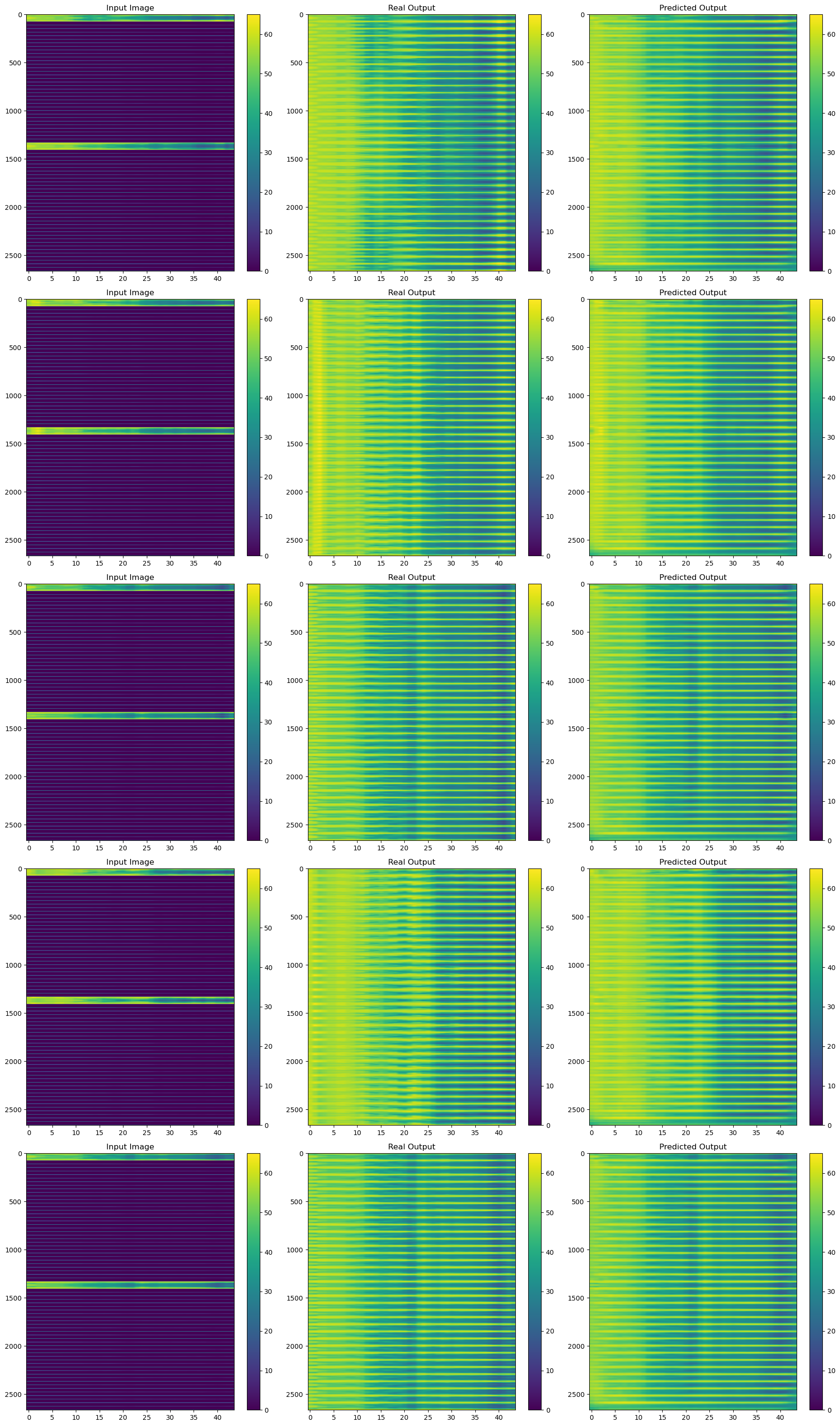


Figure 2. Training data images. Left column: only polar data. Central column: ground truth. Right: U-net prediction.

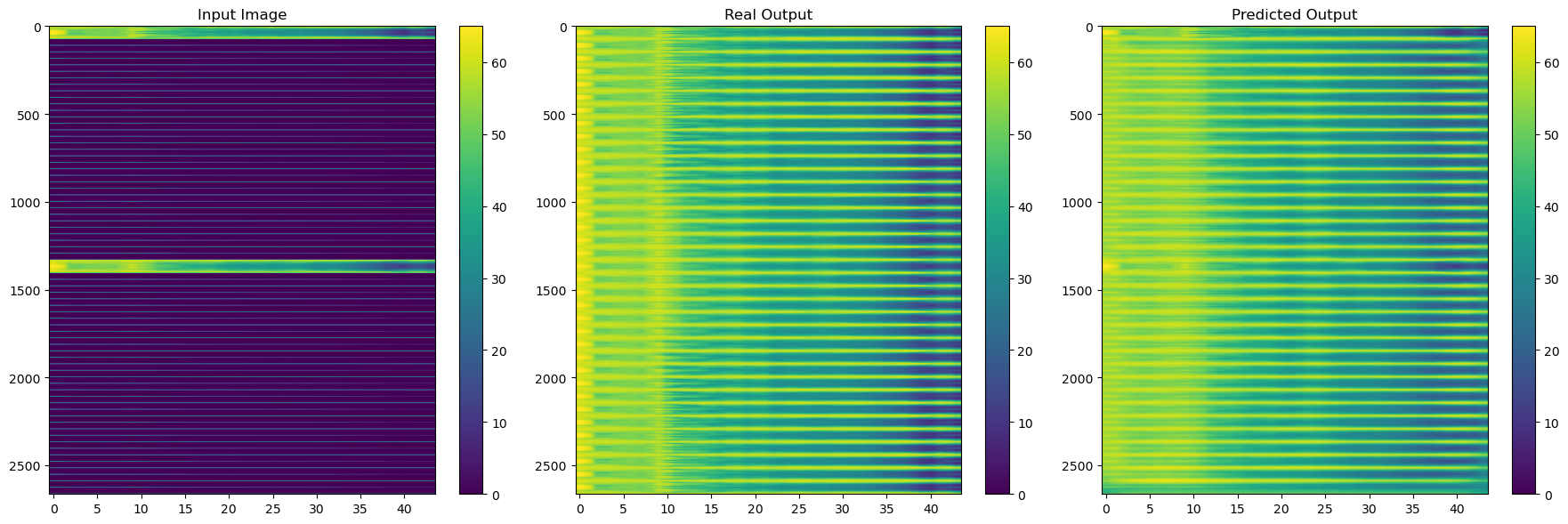


Figure 3. Validation data.

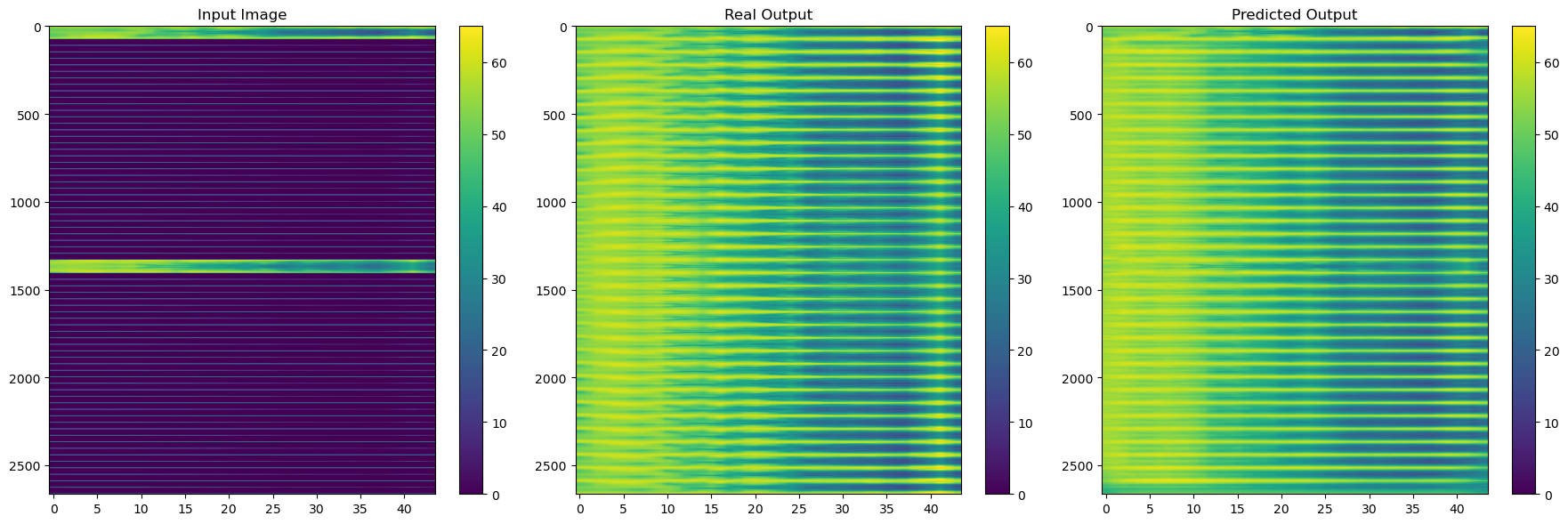
****

Figure 4. Test data (line array unit).

Figures 3-4 present validation and test data predictions.

The actual errors between true and predicted images are listed below, mean absolute error (MAE) from MSE:

Training data: 2.8388824 dB

Validation data: 4.012279 dB

Test data: 3.7689905 dB

With the proposed method a MAE lower than 4 dB can be achieved on test data (unseen by the NN during training process) for a line array unit (with strong differences between horizontal and vertical coverages at high frequencies). With more conventional loudspeakers the MAE is even better. Falta hacerlo

For a more convenient understanding of results, polar plots at Phi=45º are shown at Figures 5-11.

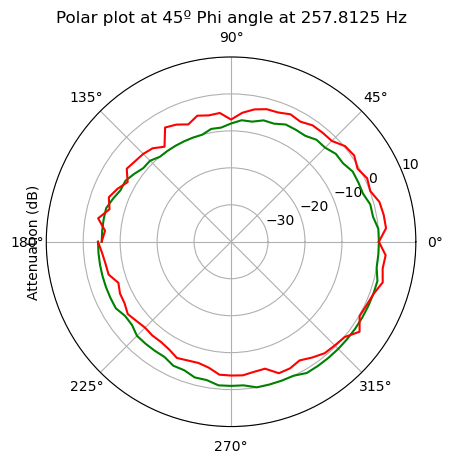


Figure 5. Polar plot at 250 Hz.

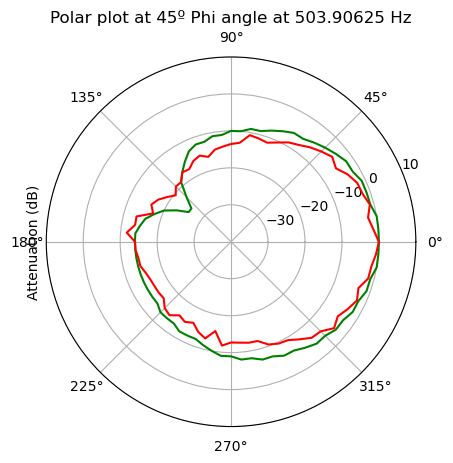


Figure 6. Polar plot at 500 Hz.

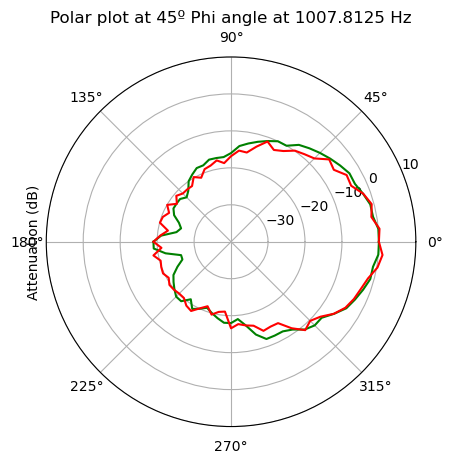


Figure 7. Polar plot at 1000 Hz.

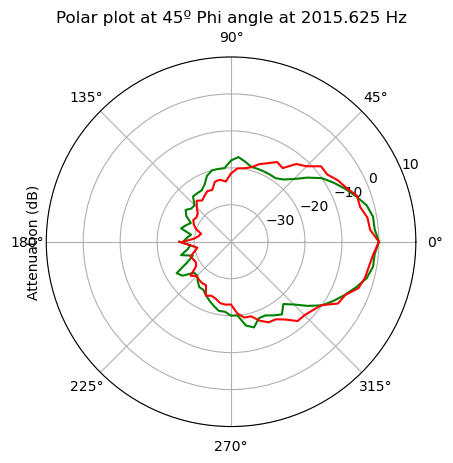


Figure 8. Polar plot at 2 kHz.

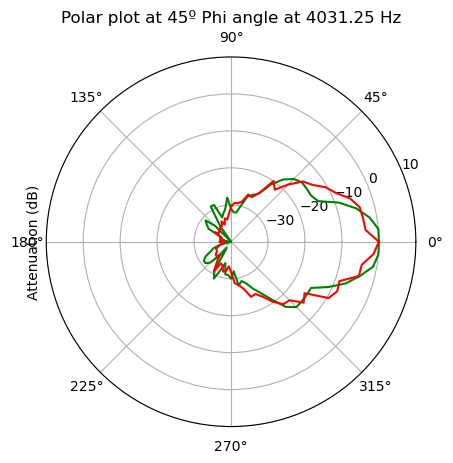


Figure 9. Polar plot at 4 kHz.

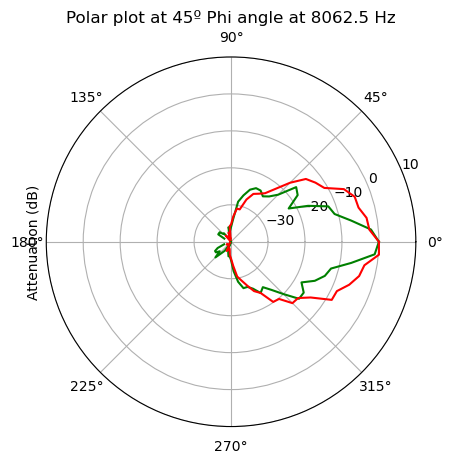


Figure 10. Polar plot at 8 kHz.

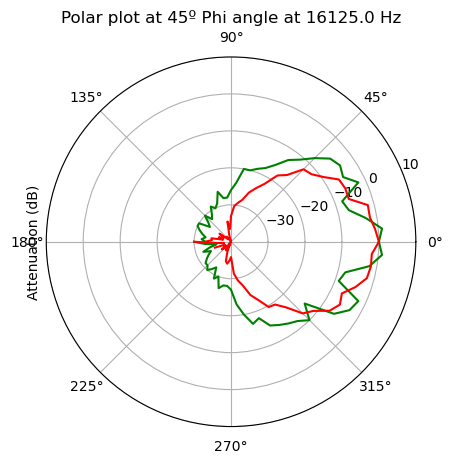


Figure 11. Polar plot at 16 kHz.

Comentar polares. Phi=45º es justo en el medio de polar horizonal y vertical, lo más alejado de los datos de entrada y por tanto el peor caso.

Si hay tiempo prepararé un dataset que incluya medición en otro plano (theta=90), que se puede medir igual de fácil que las polares. Es simplemente poner el altavoz mirando hacia arriba sobre la mesa giratoria. Con estos datos se conseguirán mejores resultados.

# Conclusions

La potencia de las redes neuronales puede utilizarse para generar globos completos de radiación a partir de medidas polares.

Añadiendo medidas a Theta=90º sale mejor?

En alta frecuencia la predicción es peor debido a que las variaciones de SPL con el ángulo son mayores. Modelos más refinados se requerirían para conseguir mayor precisión aquí.

Se plantea aplicar métodos similares para la predicción de la respuesta de fase y así obtener la respuesta completa del altavoz.

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. Accessed: Feb. 20, 2019. [Online]. Available: http://www.aes.org/e-lib.

[3] 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.

[4] F. Ma, T. D. Abhayapala, P. N. Samarasinghe, and X. Chen, “Physics Informed Neural Network for Head-Related Transfer Function Upsampling”.

[5] 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.

[6] S. S. Alotaibi, “Modeling of Individual Head-Related Transfer Functions (HRTFs) Based on Spatiotemporal and Anthropometric Features Using Deep Neural Networks”, doi: 10.1109/ACCESS.2024.3358202.

[7] 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.