DOI: 10.61782/fa.2023.0653

COMPARISON OF THE ROOM'S DIMENSIONS AND ABSORPTION DISTRIBUTION ESTIMATION PERFORMANCE USING WAVE-BASED AND GEOMETRICAL ACOUSTICS DATASET

Yuanxin Xia[1] **Zhihan Guo**[1] **Cheol-Ho Jeong**[1]*

[1] Acoustic Technology, Department of Electrical and Photonics Engineering,
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ABSTRACT

Deep neural networks (DNNs) are trained to extract the room dimensions and absorption configurations from room transfer function (TF) measurements. This study investigates the performance of DNNs in room acoustic analyses, which are trained with wave-based (WB) and geometrical acoustics (GA) simulation data. WB simulation data provide a physically accurate representation of room acoustics including diffraction and interference, albeit with substantial computation demands. In contrast, GA data can be obtained more rapidly, but with reduced accuracy. We found that the DNN trained with WB training data exhibits enhanced estimation performance and generalization capabilities when applied to real-world measurements. This study underscores the trade-offs between training dataset generation speed and their performance of machine learning algorithms in the inverse problem.

Keywords: Machine learning, inverse problem, room acoustic simulation, absorption coefficient, geometrical acoustics, wave-based simulation.

1. INTRODUCTION

Room acoustics is a vital indoor climate quality, such as concert halls, classrooms, and conference rooms. The precise estimation of a room's dimensions and absorption configuration is the initial step toward optimizing or enhancing the acoustic quality of a space. Numerous studies have employed supervised machine learning to predict reverberation time [1, 2]. However, researchers often utilize GA solvers to generate synthetic datasets for training, which may compromise accuracy and result in suboptimal performance with real-world data [3, 4]. Consequently, the application of machine learning methods to real-life scenarios remains a challenge and is anticipated to be effective only when the dataset is both accurate and extensive. To tackle this issue, one viable solution is to incorporate measurement data. However, acquiring measurement data for training can be labor-intensive and vary in quality due to factors such as instrument quality and measurement conditions.

Recent advancements in computational algorithms and architecture have made numerical methods in room acoustics increasingly feasible. Solving the wave equation, in particular, offers enhanced physical accuracy and more accurate real-world replications [5, 6]. In our recent research, we proposed a novel approach that leverages the WB dataset to estimate the absorptivity of surfaces within a room [7]. However, the extent to which the WB dataset enhances the performance is still not fully understood. As a complementary study, this paper aims to compare WB and GA datasets regarding the performance of DNN models with real-life data. We utilize open-source acoustic simulation software RAVEN as a GA method, which combines the deterministic image source method with the stochastic ray-tracing algorithm [8]. For WB data generation, commercial software Treble is employed due to its

*Corresponding author: chje@dtu.dk.

computational efficiency [9].

2. WAVE-BASED AND GEOMETRICAL ACOUSTICS DATA

2.1 Dataset overview

This study targets shoebox-shaped rooms without furniture, where each of the 6 surfaces is covered with a specific material. The dataset comprises 5 room dimensions, 5 frequency-dependent materials, and 5 frequency-independent materials, resulting in 30 material configurations derived from these combinations (see the tables in Appendix). These configurations include uniform frequency-independent and frequency-dependent (UI and UD) scenarios, as well as non-uniform distribution frequency-independent (NI) and frequency-dependent (ND) scenarios, detailed in the Appendix. The Latin hypercube sampling strategy was employed to effectively sample the sound field by positioning the source and receiver. For each room configuration and dimension, 6 sources and 125 receivers were used, yielding a total of 112,500 impulse responses (1 s) in the dataset. The highest study frequency is 707 Hz in WB simulation, corresponding to the 500 Hz octave band. For material properties, the random incident absorption coefficient is employed for both methods, with zero scattering. Detailed information about the room surface labeling, as well as the distribution of sources and receivers illustration, can be found in Ref. [7]. To enable a comprehensive comparison, datasets with identical settings were generated using both WB and GA methods.

2.2 Convergence test of GA method

In the GA method, a convergence test is often conducted to ensure the reliability and accuracy of simulation results [10]. The convergence test involves systematically refining simulation parameters, such as the number of rays and image source orders, to identify the point at which results reach a stable and consistent solution.

For single-room RAVEN simulations, we determined that the image source order does not significantly impact the results; hence, a default value of 2 is typically employed. However, adjusting the number of emitted rays (referred to as Energy Particles in RAVEN) is critical for accurate simulations. We performed convergence tests on all rooms in the dataset using non-uniform frequency-independent (NI) setting.

The convergence test was carried out for all five rooms, with sources and receivers randomly placed in the space no less than 0.25 m from each wall. The number of rays (P) was increased from an initial value of 100 to 50,000 in increments of 100. The deviation in sound pressure level (SPL) was calculated concerning the ray count (P) using the following equation:

$$\delta(P) = \frac{1}{N}\sum_{i=1}^{N} |SPL_i^{P+100} - SPL_i^{P}|(dB) \tag{1}$$

In this equation, P denotes the ray count, and N represents the number of sampling frequency points. Our analysis used a total of 707 frequency points, covering a range from 1 to 707 Hz. The results of the convergence test are depicted in Figure 1. As illustrated, when the number of rays exceeds 30,000, the δ value tends to stabilize, indicating that the statistical properties of the detection sphere source become stable. Consequently, 30,000 rays will be used in the dataset synthesis.

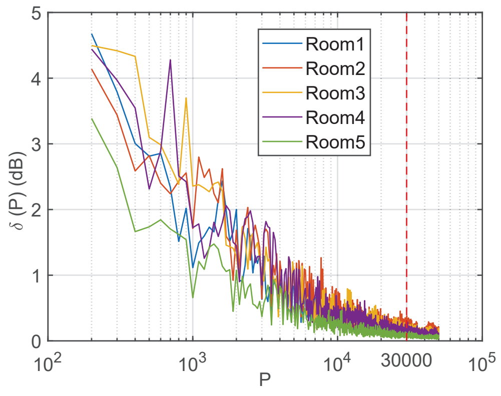

Figure 1. SPL convergence test with respect to the number of rays using RAVEN.

2.3 WB method settings

To guarantee adequate sound field sampling, the maximum mesh size for the room is set to 5 points-per-wavelength (PPW) based on the highest study frequency of 707 Hz, which is 0.097 m. The simulation featured a Gaussian pulse source with a width of 0.25 m. Therefore all sources are at least 0.25 m away from the closet wall.

3. NEURAL NETWORKS STRUCTURE AND TRAINING STRATEGY

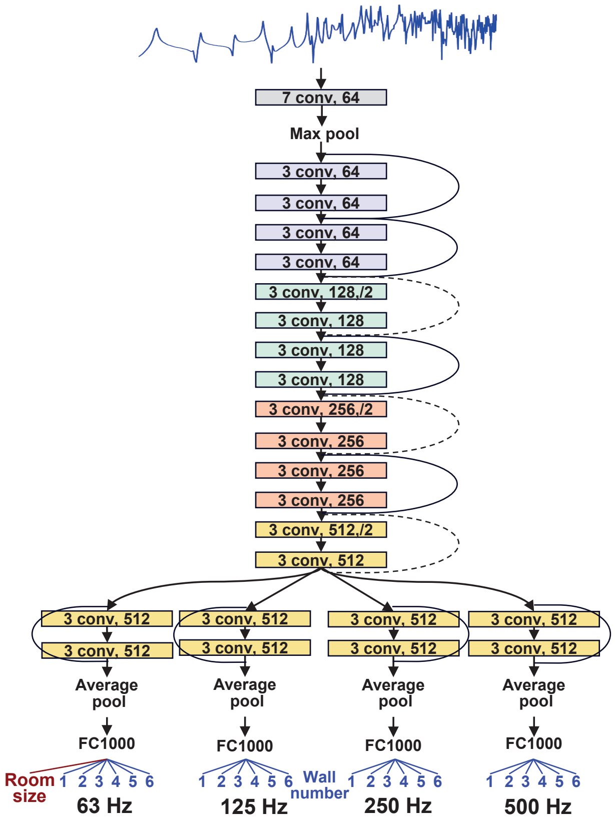

Figure 2. Multi-task ResNet-18 as a backbone structure.

In this study, we employ ResNet-18 for a multi-task learning network, consisting of four branches targeting 63 to 500 Hz octave bands to regress absorption coefficients and a room dimension branch connected to the lowest frequency band [7]. The loss weighting is uniform, which means each task is equally weighted to prevent any potential distraction. The networks initialize with Kaiming initialization without pre-trained weights. The loss function composes two parts of room dimension and absorption coefficient and is detailed in Ref. [7].

During training, we employ an Nvidia Tesla A100 GPU, the Adam optimizer [11], and a batch size of 32 for efficiency and stability [12]. With an initial learning rate of 1×10^{-4} for optimal convergence, networks are trained for 200 epochs.

4. COMPARATIVE ANALYSIS

4.1 Trade-off between data generation speed and accuracy

We employ an AMD R9 5900HX processor (8 core, 16 thread, and 16GB RAM) for the GA simulations and Nvidia A40 (SXM4, 40GB) GPUs for the WB simulations. The Treble platform leverages multiple GPUs for parallel computation, facilitating the simultaneous solving of multiple rooms. Figure 3 demonstrates that the GA simulation time remains constant regardless of room volume, while the WB simulation time increases with the room volume.

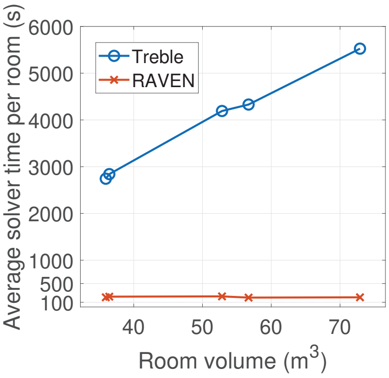

Figure 3. Comparison of synthesis data time by Treble and RAVEN.

4.2 Accuracy of room dimension and absorption estimation

In this section, we train a total of 10 models based on datasets generated by two simulation methods, GA and WB. Each method produces five distinct training sets corresponding to various scenarios, which include uniform frequency-independent (UI), frequency-dependent (UD), and non-uniform distribution scenarios (NI and ND), as well as a mixture of the four datasets (UI, UD, NI, and ND), which is referred to as ALL. The networks are trained on the full dataset without a training and testing split. After training, the models are saved and reloaded for use with measurement data. The measured TF dataset from Ref. [6] comprises three material configurations (All concrete, Ecophon Akusto Wall-A on Surface 1, and Ecophon Industry Modus on Surface 1) and includes two sources (S1, S2) and two receivers (R1, R2), ending up with four source-receiver pairs. Readers are referred to Ref. [6] for detailed information.

For single TF input, the network generates a corresponding prediction. To assess the accuracy of this prediction, we define specific error measures related to room dimensions (ΔL) and absorption coefficients ($\Delta\alpha$):

$$\Delta\alpha = \frac{1}{N_f \cdot N_s} \sum_{i=1}^{N_s} \sum_{j=1}^{N_f} |\alpha_{i,j}^{est} - \alpha_{i,j}^{true}| \quad (2)$$

$$\Delta L = \frac{1}{N_d} \sum_{i=1}^{N_d} |L_i^{est} - L_i^{true}| \quad (3)$$

where N_f and N_s are the number of frequency bands and number of surfaces, N_d is the number of dimensionality of the room. For our case which is a rectangular room, N_d is 3 and N_s is 6. These measures serve as quantitative indicators of the prediction's correctness.

As illustrated in Figure 4, for measurements with uniform concrete surfaces, the neural networks trained with the WB dataset outperform those trained with the GA dataset, particularly in estimating absorption coefficients. The mean errors for absorption coefficient estimation, $\Delta\alpha$, are 0.25 for GA and 0.07 for WB, with the latter exhibiting a significantly lower standard deviation of 0.02 compared to that of GA as 0.10. The networks trained using GA data exhibit higher sensitivity to the source-receiver pair, whereas those trained with WB data display a remarkable robustness to the measurement position. The UD-model yields the best performance for the absorption coefficient estimation, which is logical as this measured TF belongs to the UD category. Surprisingly, the NI-model demonstrates the best performance in the room dimension estimation task. The mean errors for room dimension estimation, ΔL, are about 1.06 m for GA and 0.91 m for WB, both accompanied by a standard deviation of 0.04 m.

In the scenarios involving absorption material installation, as illustrated in Figures 5 and 6, the difference in performance between WB and GA data is more pronounced than in the uniform concrete wall condition, specifically when determining absorption coefficients. The GA data demonstrate a worse performance with $\Delta\alpha$ of 0.25 and 0.23 in Figures 5 and 6, respectively, whereas $\Delta\alpha$ for WB data are reduced to 0.16 and 0.15, respectively. Moreover, the increased complexity leads the ALL dataset to outperform the UD in absorption coefficient estimation, underscoring the significance of varied training data for achieving more precise estimations in intricate scenarios.

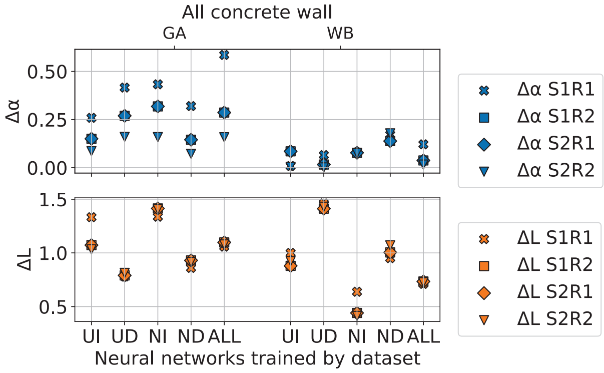

Figure 4. Absorption and dimension estimation error for the uniform concrete case.

Nonetheless, the estimation of room dimensions does not display a substantial disparity between the WB and GA datasets. The NI-model reveals the lowest dimension estimation error. In both datasets, the mean errors for room dimension estimation are 1.06 m with GA and 0.85 m with WB in Figure 5. ΔL with GA and WB are 1.09 m and 0.95 m, respectively, in Figure 6. A possible interpretation is that these errors may be more influenced by the limited sampling (only five examples) of the room dimension parameter in the training set than the choice of GA or WB methods.

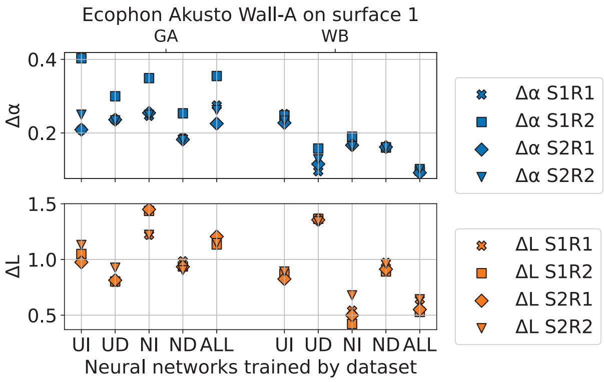

Figure 5. Absorption and dimension estimation error for the non-uniform case with Ecophon Akusto Wall-A (40 mm, flow resistivity of 47 $\mathrm{kNs/m^4}$) on surface 1 and concrete on the other surfaces.

Table 1. Summary of means and standard deviations of the inference errors

Indicator		Absorption coefficient		Dimension	
Data generation method		GA	WB	GA	WB
All concrete wall	Mean error	0.25	0.07	1.06	0.91
	Mean standard deviation	0.10	0.02	0.04	0.04
Ecophon Akusto Wall-A on surface 1	Mean error	0.25	0.16	1.06	0.85
	Mean standard deviation	0.05	0.01	0.06	0.04
Ecophon Industry Modus on surface 1	Mean error	0.23	0.15	1.09	0.95
	Mean standard deviation	0.03	0.01	0.10	0.08
Averaged	Error	0.25	0.13	1.07	0.91
	Standard deviation	0.06	0.02	0.06	0.05

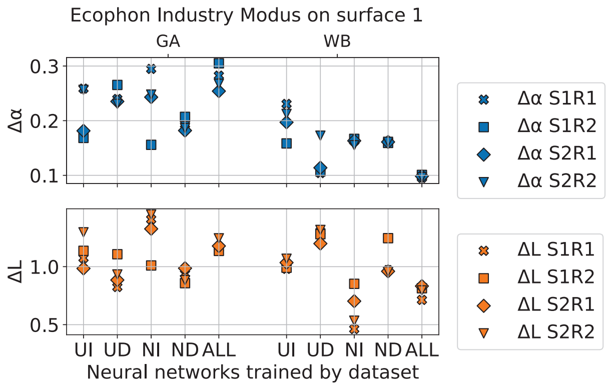

Figure 6. Absorption and dimension estimation error for the non-uniform case with Ecophon Industry Modus (100 mm, flow resistivity of 10.9 $\mathrm{kNs/m^4}$) on surface 1 and concrete on the other surfaces.

5. DISCUSSION AND CONCLUSION

In this study, we assess the performance of DNN models trained with WB and GA synthetic TF data for the inverse estimation of absorption configuration and room dimensions. Our measured data reveals that despite the increased computational time required for WB simulations, the absorption estimation error from the DNNs trained with GA data is nearly twice (92%) as large as the error with WB data. These absorption errors are relatively large as 0.25 and 0.13 for GA and WB, respectively, because the model has not seen these materials, particularly the porous absorbers, during the training step. ΔL is increased by 18% with GA than WB. Furthermore, the WB data also result in a significantly lower variance, with a reduction of about 67% in the absorption coefficient standard deviation and approximately 14% in the estimated dimension standard deviation. This demonstrates better generalization performance and robustness by using WB data than GA data. Additionally, the importance of training data variations for achieving more accurate estimations is emphasized, suggesting that DNNs trained with a diverse range of room configurations and acoustic materials produce more precise and consistent results.

6. ACKNOWLEDGEMENT

The authors extend their gratitude to IHTA, RWTH Aachen University for the access to the RAVEN software, to Treble for providing their Synthetic Data Generation API, and to the DTU Computing Center for providing access to its GPU cluster. Special acknowledgments go to Lukas Aspöck for his indispensable guidance in RAVEN, and to Haukur Hafsteinsson for his invaluable assistance throughout the modeling process with Treble.

7. REFERENCES

[1] J. Nannariello and F. Fricke, "The prediction of reverberation time using neural network analysis," *Applied Acoustics*, vol. 58, no. 3, pp. 305–325, 1999.

[2] F. L. Zainudin, S. Saon, M. N. Yahya, *et al.*, "Prediction of classroom reverberation time using neural network," in *Journal of Physics: Conference Series*, vol. 995, p. 012028, IOP Publishing, 2018.

[3] W. Yu and W. B. Kleijn, "Room acoustical parameter estimation from room impulse responses using deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 436–447, 2021.

[4] C. Foy, A. Deleforge, and D. Di Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning," *The Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. 1286–1299, 2021.

[5] F. Pind, A. P. Engsig-Karup, C.-H. Jeong, J. S. Hesthaven, M. S. Mejling, and J. Strømann-Andersen, "Time domain room acoustic simulations using the spectral element method," *The Journal of the Acoustical Society of America*, vol. 145, no. 6, pp. 3299–3310, 2019.

[6] T. Thydal, F. Pind, C.-H. Jeong, and A. P. Engsig-Karup, "Experimental validation and uncertainty quantification in wave-based computational room acoustics," *Applied Acoustics*, vol. 178, p. 107939, 2021.

[7] Y. Xia and C.-H. Jeong, "Room dimensions and absorption inference from room transfer function via machine learning," *arXiv preprint arXiv:2304.12993*, 2023.

[8] D. Schröder, *Physically based real-time auralization of interactive virtual environments*, vol. 11. Logos Verlag Berlin GmbH, 2011.

[9] F. Pind, "Cloud-based numerical room acoustic simulations using the discontinuous galerkin method: Benchmarks and industrial applications," in *24th International Congress on Acoustics*, 2022.

[10] M. Vorländer, "Die genauigkeit von berechnungen mit dem raumakustischen schallteilchenmodell und ihre abhängigkeit von der rechenzeit," *Acta Acustica united with Acustica*, vol. 66, no. 2, pp. 90–96, 1988.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *arXiv preprint arXiv:1609.04836*, 2016.

8. APPENDIX

Table 2. Absorption coefficient of material.

Material	ID	Octave band absorption coefficient								ID	Averaged
		63Hz	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz	8kHz		
Carpet with underlay	A	0.15	0.22	0.33	0.4	0.43	0.44	0.44	0.44	a	0.35
Gypsum board	B	0.08	0.07	0.06	0.05	0.04	0.03	0.03	0.03	b	0.05
Melamine based foam	C	0.08	0.18	0.56	0.95	0.95	0.95	0.95	0.95	c	0.70
Frabric wrapped panel	D	0.15	0.30	0.64	0.75	0.80	0.80	0.77	0.77	d	0.62
Window glass	E	0.40	0.34	0.25	0.17	0.11	0.07	0.05	0.04	e	0.18

Table 3. Material configuration.

Uniform Material Configuration	Frequency Independent (UI)					Frequency Dependent (UD)				
Surface 1	a	b	c	d	e	A	B	C	D	E
Surface 2	a	b	c	d	e	A	B	C	D	E
Surface 3	a	b	c	d	e	A	B	C	D	E
Surface 4	a	b	c	d	e	A	B	C	D	E
Surface 5	a	b	c	d	e	A	B	C	D	E
Surface 6	a	b	c	d	e	A	B	C	D	E
Total uniform material configuration						10				

Non-uniform Material Configuration	Frequency Independent (NI)										Frequency Dependent (ND)									
Surface 1	b	d	b	b	c	c	e	e	d	e	B	D	B	B	C	C	E	E	D	E
Surface 2	b	d	b	b	d	c	b	e	d	b	B	D	B	B	D	C	B	E	D	B
Surface 3	a	a	a	a	a	c	a	a	a	a	A	A	A	A	A	C	A	A	A	A
Surface 4	c	d	b	c	c	c	c	c	c	b	C	D	B	C	C	C	C	C	C	B
Surface 5	b	e	e	b	d	b	b	b	b	b	B	E	E	B	D	B	B	B	B	B
Surface 6	b	d	b	b	d	e	b	b	b	c	B	D	B	B	D	E	B	B	B	C
Total non-uniform material configuration											20									

Table 4. Room dimension.

Dimension ratio	Length (m)	Width (m)	Height (m)	Area(m^2)	Volume (m^3)
1:1.11:1.67	3	4.5	2.7	13.50	36.45
2:3:5	4	6.75	2.7	28.00	72.90
1:1.4:1.9	3.8	5.15	2.7	19.47	52.84
1:1.56:1.86	4.2	5	2.7	21.00	56.70
1:1:1	3.3	3.3	3.3	10.89	35.94