Next-Generation Synthetic Data Techniques for Training, Evaluation, and Prototyping in Audio Signal Processing

Finnur Pind [1], *Georg Götz* [1], *Daniel Gert Nielsen* [1]

[1] Treble Technologies, Reykjavík, Iceland

Abstract—**The development, benchmarking, and optimization of modern audio signal processing algorithms—such as speech enhancement, source separation, echo cancellation, and blind room acoustics estimation—critically depend on access to high-quality acoustic data. While such data can be measured in real environments, simulation and synthetic data generation offer superior scalability, traceability, controllability, and coverage of diverse acoustic conditions. However, conventional simulation methods often fail to produce data accurate enough for algorithms to generalize reliably to real-world scenarios. Recent advances in acoustic simulation have enabled the generation of high-fidelity synthetic data that closely matches measured responses. This demonstration introduces a novel wave-based simulation engine, delivered as an accessible Python SDK, which allows researchers and practitioners to generate high-fidelity, multi-channel, optionally device-specific Room Impulse Responses (RIRs) and Spatial Room Impulse Responses (SRIRs) at scale. Unlike conventional geometric acoustics methods, our approach achieves a substantially closer match to real-world measurements. This fidelity, in turn, enables algorithms trained on simulated data to perform and generalize more effectively, as validated by both perceptual and algorithmic evaluation metrics. It also provides evaluation results comparable to those obtained with measured data, but with far greater scalability to test a broader range of scenarios and uncover performance limitations. In addition to algorithm development, these capabilities support fully virtual workflows for acoustic hardware prototyping. To support the community, we will release an open benchmark dataset of measured and simulated RIRs as part of the demonstration, facilitating reproducible experiments and validation.**

1. INTRODUCTION

Machine learning-based audio processing systems depend on diverse, realistic datasets to achieve robust performance. RIRs paired with anechoic signals underpin training pipelines, algorithm tuning, and reproducible evaluation. While measurement campaigns in real environments remain the gold standard for realism, they impose high logistical and financial costs. This often leads to incomplete coverage of challenging scenarios such as occluded sources, near-field effects, and complex geometries, and prevents data collection specific to a particular audio device [1,2].

Simulation-based RIR generation offers a scalable alternative. However, traditional geometric acoustics frequently fall short of capturing wave phenomena essential for realism—diffraction, interference, and scattering. As a result, models trained or evaluated on such synthetic data often underperform when deployed in real-world environments [3-5].

Our approach addresses these limitations through an efficient, massively parallelized wave-based simulation engine [6] that delivers physically and perceptually accurate results. Table 1 below considers the case when this simulation engine is applied to DSP/AML algorithm evaluation across three different algorithm tasks and benchmarked against ground truth measured data across six different rooms, that our simulations (DG-FEM) achieve the highest correlation coefficients with measured ground truth data (up to 0.92) and substantially reduced RMSE across key metrics such as PESQ, ESTOI, and SI-SDR compared to conventional simulation methods.

Table 1: The table reports the Pearson correlation coefficient ρ and the root mean squared error (RMSE) between measured evaluation results and the corresponding results obtained from DG-FEM, GA-RR, and GA-RT simulations. Results are compared for three different ASP/AML algorithms with their respective performance metrics. Across all metrics, DG-FEM simulations yield similar evaluation results as measurements, while geometrical acoustic simulations could not replicate the measured evaluation results as precisely.

	SDE (ML)		**Dereverberation (ML)**						**Dereverberation (DSP)**					
	Distance est. error		PESQ		ESTOI		SI-SDR		PESQ		ESTOI		SI-SDR	
	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓	$\rho\uparrow$	RMSE ↓
DG-FEM	**0.76**	**0.16**	**0.92**	**0.22**	**0.91**	**0.05**	**0.75**	**3.09**	**0.91**	**0.21**	**0.90**	**0.03**	**0.73**	**2.35**
GA-RR	0.59	0.24	0.68	0.51	0.70	0.11	0.61	3.68	0.77	0.35	0.76	0.06	0.57	2.86
GA-RT	0.51	0.25	0.28	0.81	0.45	0.17	0.23	5.71	0.14	0.70	0.56	0.07	0.10	4.19

Equally important, this simulation engine is delivered within a comprehensive Python SDK that streamlines every stage of the workflow—from defining complex virtual scenes and advanced source models to scalable cloud execution and spatial audio post-processing. The SDK's intuitive interface, automation tools, and seamless integration with popular data pipelines make it practical to generate large, diverse, and reproducible datasets at scale. This combination of high-fidelity simulation and robust auxiliary functionality empowers researchers and practitioners to design, evaluate, and deploy learning-based signal processing systems more effectively.

Participants will learn how to integrate this technology into workflows spanning model training, large-scale evaluation, and device prototyping.

2. DEMONSTRATION DETAILS

This demonstration will be structured into multiple thematic sections, each illustrated with live examples and supporting material:

We will introduce the numerical foundations of the wave-based simulation engine and contrast its performance with conventional geometric methods. Comparisons against measured RIRs will illustrate strengths and trade-offs in accuracy, computational efficiency, and perceptual realism [7,8].

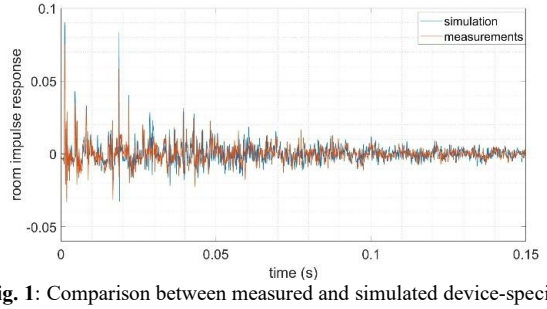

Fig. 1: Comparison between measured and simulated device-specific room impulse response in complex acoustic conditions.

Participants will learn how to programmatically create rich, complex 3D acoustic scenes, including multi-room apartments, restaurants, offices, vehicles, and outdoor spaces. We will demonstrate applying realistic material absorption profiles, directional sound sources (e.g., human talkers, loudspeakers), and environmental noise layers. Additionally, we will show how to configure simulation workflows and

dataset generation pipelines optimized for specific signal processing tasks. This includes balancing compute cost, target fidelity, and scene complexity, as well as scripting parameter sweeps to efficiently cover diverse scenarios.

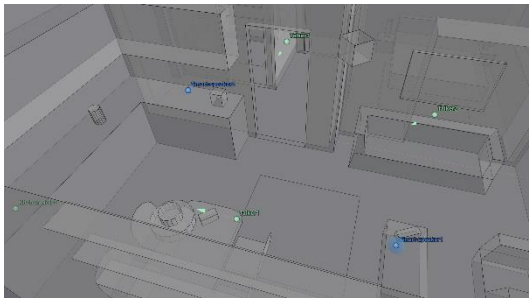

Fig. 2: Complex acoustic scene with three talkers, noise sources, and two smart speakers with microphone arrays.

The demonstration will cover how to simulate spatial RIRs up to 32nd-order ambisonics, enabling physically accurate spatial sound across a broad frequency range. We will show how to incorporate device-specific microphone array geometries (e.g., AR headsets, smart speakers) to produce multi-channel outputs. Anechoic device transfer functions can be measured or simulated as part of the workflow.

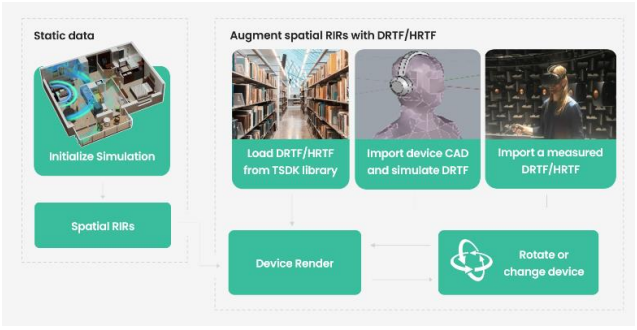

Fig. 3: Workflow for augmenting data by mixing spatial RIRs and device transfer functions in post-processing.

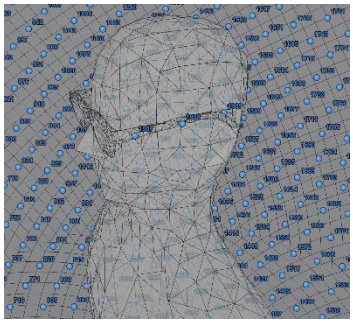

Fig. 4: Device transfer function simulation with multi-mic AR headset.

We will illustrate how to model near-field phenomena critical to many applications, such as mouth-to-microphone transmission paths, close-proximity loudspeaker interactions, and scenarios with partial occlusion or strong boundary reflections.

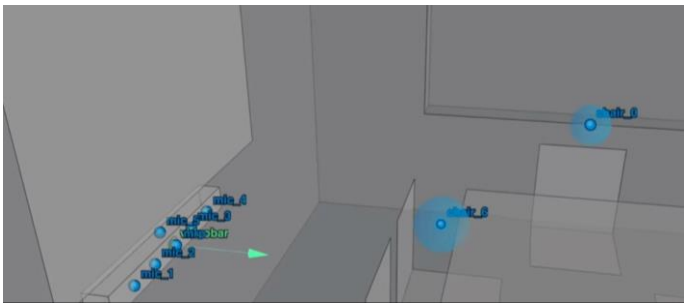

Fig. 5: Double-talk scenario, with self-noise from a videobar loudspeaker and multiple talker sources.

We will present audio-visual auralizations demonstrating the perceptual realism of simulated RIRs compared to measured references. Participants will be able to listen to examples through headphones or loudspeakers to appreciate the temporal and spatial accuracy. Our benchmark dataset (to be released open-source) will include examples of these comparisons for community experimentation.

We will show how high-fidelity synthetic RIRs can reveal algorithm weaknesses and performance boundaries more effectively than conventional synthetic data, providing insights comparable to measured data but with far greater scalability. This approach helps uncover edge cases and improve model robustness.

Through examples and case studies, we will demonstrate how large-scale synthetic datasets can be used to train ML-based signal processing models such as speech enhancement and dereverberation, resulting in improved generalization and performance.

Finally, we will show how to use simulation outputs to assess microphone and loudspeaker configurations, simulate edge cases (e.g., partial occlusion), and optimize designs before committing to physical prototypes. We will demonstrate how to build automated virtual prototyping pipelines for hardware and algorithm development.

As a part of the demonstration, we will release an open-source data set containing simulated and measured impulse responses. The simulated RIRs will be the world's first open source broadband wave-based RIRs, benefiting both the participants in the demonstration as well as the scientific community. We will invite the WASPAA community to contribute to our open benchmark dataset by submitting measured RIRs, participating in validation studies, and collaborating on extensions targeting new application domains.

3. EQUIPMENT AND FACILITIES

We ask for a standard presentation venue equipped with video projection capability. A high-quality audio playback system is desirable for group listening comparisons but not strictly needed. To supplement this, we will provide reference headphones for individual listening experiences. All attendees will get access to the software tool, as well as our open benchmark dataset, example scripts, and supporting documentation.

The demonstration can be flexibly adapted in length and depth to meet WASPAA's scheduling requirements

4. REFERENCES

[1] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "MESHRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Online conference, 2021.

[2] G. Stolz, S. Werner, F. Klein, L. Treybig, A. Bley, and C. Martin, "Autonomous robotic platform to measure spatial room impulse responses," in *Proc. DAGA*, Hamburg, Germany, pp. 59–61, 2023.

[3] T. Ko, V. Peddinti, D. Povey, M. L. Seltzer and S. Khudanpur, "A study on data augmentation of reverberant speech for robust speech recognition," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*, New Orleans, LA, USA, 2017, pp. 5220-5224, 2017

[4] R. Scheibler, E. Bezzam, and I. Dokmani´c, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*. (ICASSP), pp. 351–355, 2018.

[5] Z. Tang, R. Aralikatti, A. J. Ratnarajah, and D. Manocha, "GWA: A large high-quality acoustic dataset for audio processing," in *ACM SIGGRAPH Conf. Proc.*, pp. 1–9, 2022.

[6] A. Melander, E. Strøm, F. Pind, A. P. Engsig-Karup, C.-H. Jeong, T. Warburton, N. Chalmers, and J. S. Hesthaven, "Massively parallel nodal discontinous Galerkin finite element method simulator for room acoustics," *Int. J. High Perform. Comput. Appl.,* vol. 38, no. 3, pp. 154–174, 2024.

[7] S. Guðjónsson, H. Sampedro Llopis, M. Cosnefroy, and H. Hafsteinsson, "Validation study of broadband wave-based acoustic simulations and binaural renderings," in *Proc. DAGA*, Copenhagen, Denmark, pp. 1173–1176, 2025.

[8] H. Sampedro Llopis, S. Guðjónsson, M. Cosnefroy, H. Hafsteinsson, and F. Pind, "Validation study of broadband numerical methods for room acoustic simulations and binaural rendering in a furnished room," in *Proc. 11th Conv. European Acoust. Assoc. (Forum Acusticum)*, Málaga, Spain, 2025.