DOI: 10.61782/fa.2023.1277

forum acusticum 2023

Full bandwidth wave-based virtual acoustics with directive sources and receivers

Finnur Pind[1*] **Matthias Cosnefroy**[1] **Steinar Guðjónsson**[1]
[1] Treble Technologies, Reykjavík, Iceland

ABSTRACT

A wave-based virtual acoustics framework that enables full-bandwidth room acoustic simulations in small-medium sized rooms is presented. The framework utilizes a time-domain formulation of the discontinuous Galerkin finite element method. We present methods for modeling directional sound sources within the framework, to account for realistic sound sources such as loudspeakers. Additionally, we present methods for modeling directional receivers using large virtual microphone arrays, which enables very high order ambisonics sound field encoding in the simulations. Taken together, this yields a system that can simulate the interaction of room acoustics and audio devices with high precision, and outputs binaural room impulse responses with high spatial accuracy. The system is entirely cloud-based, which yields flexibility and scalability, e.g., for running simulations at scale for generating synthetic data for training audio machine learning algorithms. We present benchmarks that demonstrate the performance of the system.

Keywords: Wave-based, Room acoustic simulations, High-order ambisonics.

1. INTRODUCTION

Room acoustic simulations are useful for various design and virtual prototyping purposes, e.g. in building design, automotive design and product design. It is desirable to have both fast and accurate simulations, but generally accuracy is obtained at the expense of computational efficiency. This fact motivates further research on room acoustic simulation methods, where the objective is to alleviate such a trade-off between accuracy and speed.

Acoustic simulation methods are typically divided into two main categories: wave-based methods and geometrical methods. Geometrical acoustics approaches, such as the ray-tracing and the image-source methods, are essentially high-frequency approximations [1]. They tend to yield short compute runtimes but the accuracy can suffer, particularly in the lower frequency range, due to the inability to model wave phenomena such as diffraction and interference. There are also use-cases where these methods can yield insufficiently accurate results even at high frequencies, e.g., when it is desirable to have precise phase modeling.

Wave-based methods inherently account for wave phenomena such as diffraction and phase effects, and are therefore generally considered more accurate than geometrical acoustics approaches [2]. However, wave-based methods tend to be considerably more computationally intensive, which has up until now hindered their widespread adoption in practice.

This paper presents a wave-based simulation framework that enables accurate room acoustic simulations across a wide range of frequencies within what might be considered practical runtimes. The framework can include directional sound sources and directional receivers. Some benchmarks are provided which demonstrate the accuracy of the framework.

2. SIMULATION FRAMEWORK

2.1 Governing equations and numerical discretization

Acoustic wave propagation in a 3D space can be described by the homogeneous linearized Euler equations,

[*] ***Corresponding author***: *fp@treble.tech*

$$\frac{\partial \boldsymbol{v}}{\partial t} = -\frac{1}{\rho}\nabla p,$$
$$\frac{\partial p}{\partial t} = -\rho c^2 \nabla \cdot \boldsymbol{v},$$

with p and $\mathbf{v}$ the acoustic pressure and particle velocity vector, respectively, together with appropriate initial and boundary conditions. The numerical framework solves these equations in time and space by estimating the spatial derivatives with the (nodal) discontinuous Galerkin method (DGM), and by integrating the temporal derivatives using a fourth-order explicit Runge-Kutta time stepping method. For the sake of brevity, the details of the numerical discretization are avoided here, but can be found in [3]. The DGM method is attractive due to its parallelizability, geometric flexibility, and ability to extend to arbitrary orders of accuracy in a straightforward way. Frequency-dependent boundary conditions can also be accounted for with the method of auxiliary differential equations, see further details in [3].

2.2 Directional source modeling

When simulating the acoustics of a room, the main quantity of interest is often the room impulse response for a given source-receiver combination (a.k.a. the Green's function), which describes the pressure signal recorded by a receiver at a specific location in response to an impulsive excitation at the source location. One way to account for such a source in the simulations is to use a broadband point-source excitation (e.g., via a forcing term in the governing equations or via an initial pressure disturbance). However, in many cases it is desired to account for a complex directional sound source, e.g., to model how the radiation pattern of a particular loudspeaker influences the sound field.

One approach for modeling directional sources in the proposed simulation framework is to include the geometry of the speaker box in question directly into the 3D room model, and then apply an impulsive surface velocity excitation to the loudspeaker membrane. This yields a directional source due to the radiation pattern of the membrane and subsequent edge diffraction that the baffle geometry introduces. Figure 1 shows an example of a simple loudspeaker geometry with a single driver; note that the approach can be applied to loudspeakers with more than one driver.

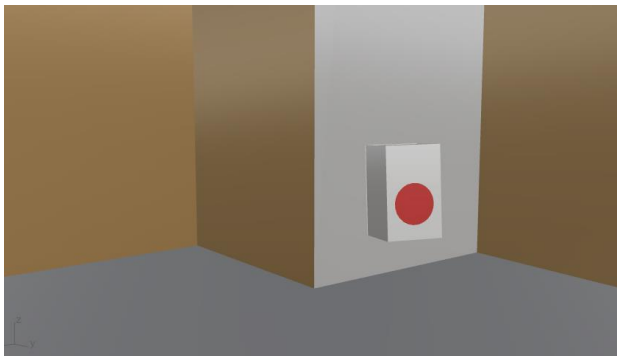

Figure 1. Example of a loudspeaker box in a 3D room model. The red membrane is excited with the surface velocity condition described in section 2.2.

More specifically, the loudspeaker is excited by prescribing the normal acoustic velocity along the surface of the (fixed) membrane geometry (i.e., a Neumann boundary condition). Due to numerical constraints, the excitation signal is a Gaussian-like impulse. Such an excitation does not have a flat energy spectrum at all frequencies; this is remedied in post-processing by normalizing the simulated room impulse responses to the simulated free-field on-axis response at a distance of one meter. The rest of the speaker box is modeled as an acoustically hard boundary.

This approach has the advantage of being straightforward to use, as it only requires a sufficiently accurate model of the speaker geometry. The effects of the diffraction from the baffle are then inherently included, making it particularly well suited for low-frequency simulations. It should be kept in mind that using a flat piston to represent the membrane is an approximation that may neglect potential near-field and nonlinear effects. The approach is also best suited to modeling loudspeaker sources and may be less relevant to modeling, e.g., speech sources, or more generally to accounting for arbitrary directivity patterns.

2.3 Directional receiver modeling

It can be desirable to capture certain spatial characteristics around the receiver locations, e.g., to enable binaural or multi-channel loudspeaker auralizations. Just like in measurements with large microphone arrays [4], an analogous approach can be taken in simulations, where an array of receivers is placed around the sampling location to encode the resulting sound field into an ambisonics spatial impulse response, which can then be decoded to binaural or

multi-channel loudspeaker playback systems [5]. Figure 2 shows an overview of the process.

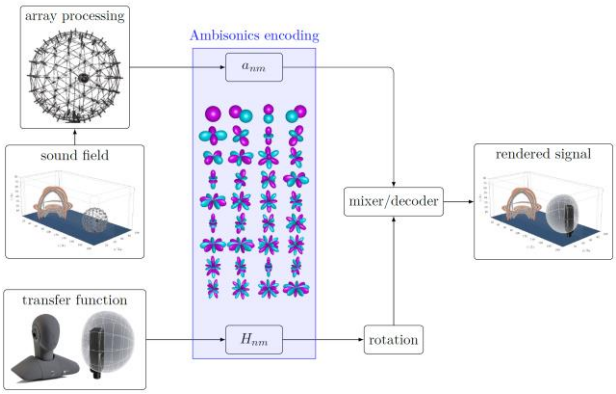

Figure 2. Overview of the receiver modeling approach.

In physical measurements with microphone arrays, some physical constraints are introduced: the microphones take up space which limits how many microphones can be placed in a given area, the microphones may disturb the sound field, and the measurement setup can be expensive and time consuming. Obtaining a sufficiently large signal-to-noise ratio can also be challenging. In simulations, these limitations are less severe: signal-to-noise ratio is generally better and there are no limitations on how many receivers can be used and where they can be placed. Adding many receivers comes at little additional compute cost.

The science of designing spherical microphone arrays is well established [6]. Multiple design choices such as microphone placement, microphone type, microphone count, as well as surface type (open vs rigid) must be taken into consideration. In our case, we opted for using an open spherical array of cardioid receivers. This array design is very robust and allows a wide operating frequency range. The number of receivers and the array radius can be modified freely to control the valid frequency range of the array and the ambisonics order which can be achieved. We have for example used receiver arrays of 32 receivers to enable 2nd order ambisonics impulse responses which are valid up to roughly 2 kHz and 434 receivers to enable 16th order ambisonics impulse responses which are valid up to 8 kHz for an averaged sized head.

3. EXPERIMENTAL VALIDATION

3.1 Source modeling validation

We conduct a directivity measurement in an anechoic chamber of a single driver loudspeaker placed in a closed box. The measured directivity is then compared against simulated directivity using the source excitation method described in section 2.2. In the simulation, the loudspeaker is placed at the center of a rather large room, receivers placed in a sphere around the box at 1m distance, and the simulation time length is truncated such that no reflections from the walls arrive at the receiver positions. Figures 3-4 show the results at 125 Hz and 1000 Hz, respectively. We observe a fair match between the measurements and the simulations.

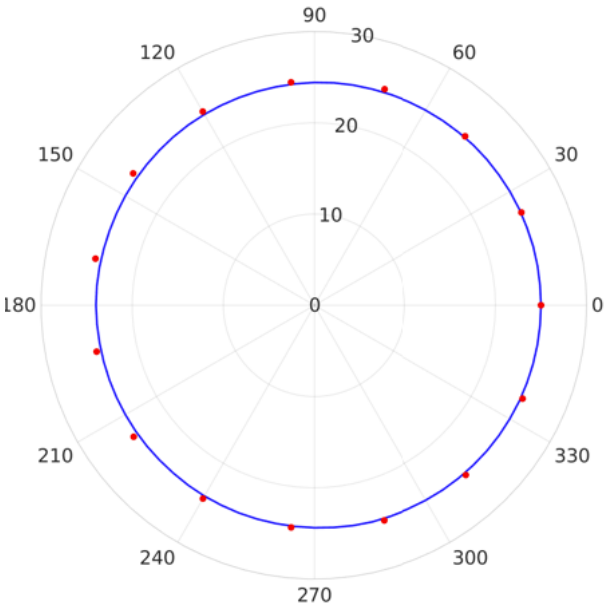

Figure 3. Measured (red dots) and simulated (blue line) directivity of the loudspeaker at 125 Hz.

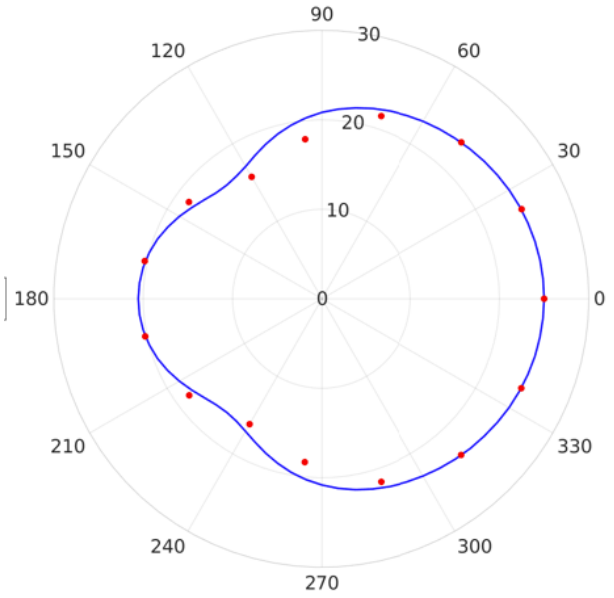

Figure 4. Measured (red dots) and simulated (blue line) directivity of the loudspeaker at 1000 Hz.

3.2 Receiver modeling validation

To validate the receiver modeling, we carry out a test on simulating an HRTF, where the receiver is modeled with the approach described in section 2.3, i.e. with a virtual microphone array and not by having the physical geometry of the head and torso present in the simulation.

A single plane wave is simulated passing by the receiver array. The open virtual spherical array contains 434 cardioid receiver and has a 10 cm radius. The resulting receiver signals are encoded to 16th order ambisonics and then decoded to binaural using the method described in [5]. An example result is shown in figure 5, where an excellent agreement between measured HRTF data and simulated HRTF response via the microphone array processing technique is observed.

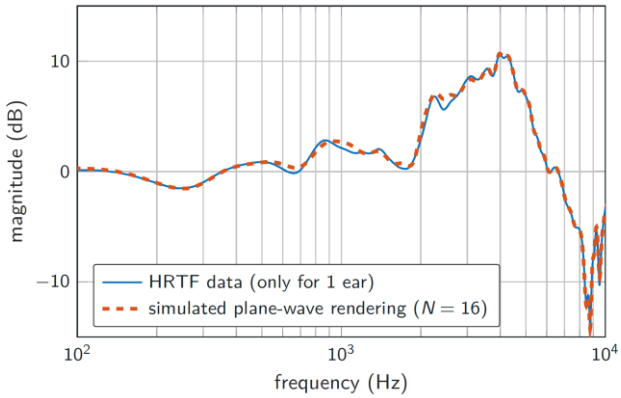

Figure 5. Simulated HRTF using the receiver modeling approach from section 2.2. compared against a measured HRTF.

3.3 Full room simulation case study

We now consider a benchmark study, where we compare measured monaural and binaural room impulse responses against simulated responses using the proposed framework. A wide frequency range of 40-8000 Hz is considered. The room in question is an 80 m3 shoebox shaped room. Most surfaces are concrete, but a 100 mm rockwool absorber has been placed on the walls and floor in various locations. Figure 6 shows a photograph from the room.

Figure 6. The room used for benchmarking.

The loudspeaker used for the measurements is a two-way Yamaha HS-8 loudspeaker. The monaural room impulse responses are measured using a GRAS ½ inch free field microphone, whereas the binaural room impulse responses are measured using a KEMAR mannequin with GRAS KB1090 shore 35 ears and GRAS 40A0 pressure mics

mounted flush with the ear canal entrance, i.e. no ear coupler is used. Figure 7 shows the source and receiver positions used. To enable binaural decoding, the HRTF dataset for this mannequin is also measured in an anechoic chamber.

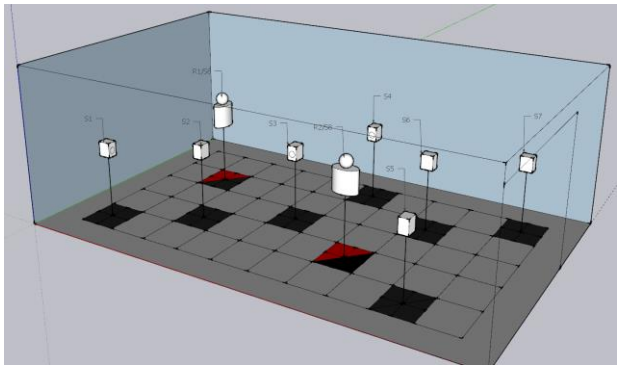

Figure 7. Source / receiver positions used.

The surface impedance of the rockwool absorber is measured in an impedance tube according to ISO 10534-2. This produces valid results in the frequency range of 230 Hz to 2000 Hz. The impedance at higher and lower frequencies was estimated using Komatsu's model, by using the known thickness of the absorber and fitting the flow resistivity value to give a good match to the measured impedance values for the mid frequency range from the impedance tube measurement. The absorption properties of the remaining surfaces are taken from established databases. Two simulations are performed, one using the larger membrane in the loudspeaker box (up to 3 kHz) and another using the tweeter membrane (3-8 kHz), and the results of these two are merged in post-processing.

The simulations were carried out on a cluster with six Nvidia A100 GPU cards running in parallel. The finite element mesh resolution is set at roughly 1.6 elements per wavelength at the highest frequency of interest (8000 Hz) and fourth order basis functions are used. These settings ensure minimal dispersion errors across the frequency range of interest. The compute time for 0.5 sec impulse response length was around 14 hours. Figure 8 shows a comparison between measured and simulated monaural IRs for an arbitrarily chosen source/receiver combination. A good match is observed, all prominent early reflections are seen in the simulated response with approximately correct amplitudes. Figure 9 shows a frequency response for another source/receiver combination, again a good match is observed. The measured low frequency modes are seen in the simulation results and the high frequency trend is captured too.

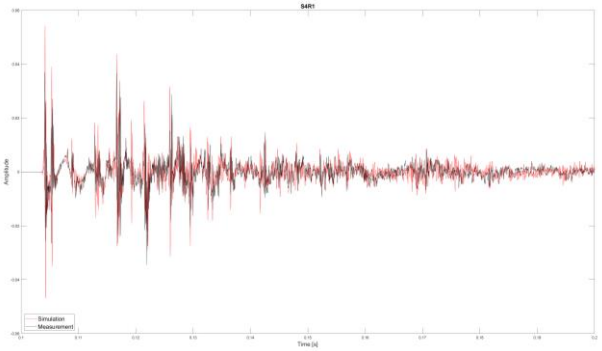

Figure 8. Simulated versus measured monaural room impulse responses compared for S4R1.

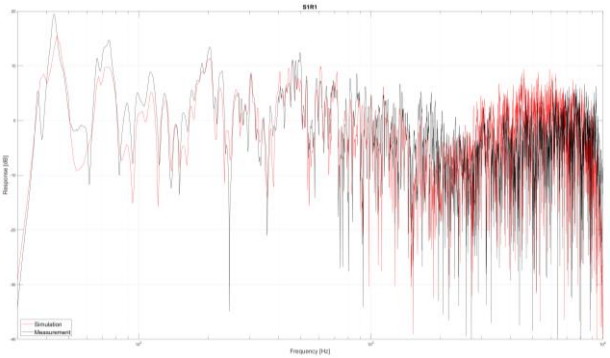

Figure 9. Simulated versus measured monaural room impulse responses compared for S1R1.

Figure 10 shows simulated and measured acoustic parameters, T20 and EDT, compared for an arbitrarily chosen source / receiver location, where again a good match is observed.

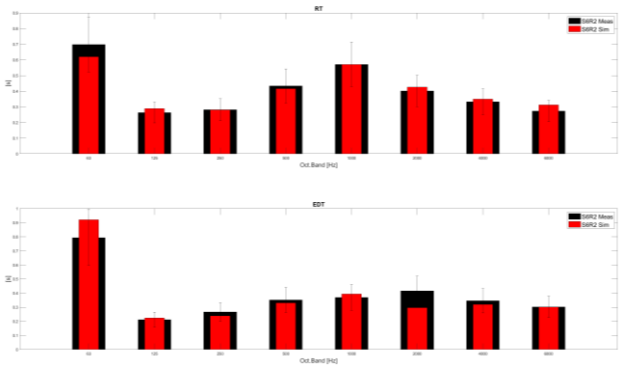

Figure 10. Simulated versus measured room acoustic parameters (T20 above, EDT below) for S6R2.

Figure 11 shows an example of a simulated versus measured BRIR. Again, a good match is observed, validating the receiver modeling methodology in a full room setting.

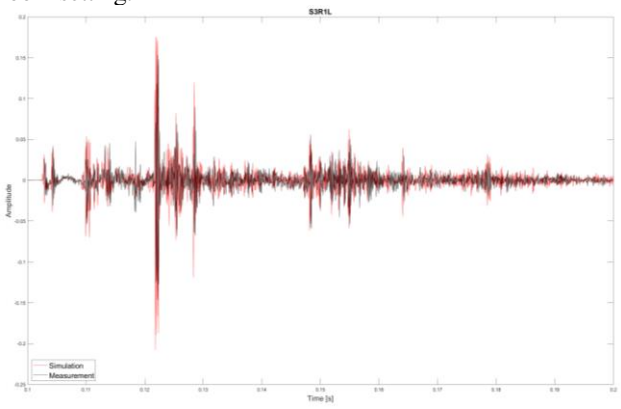

Figure 11. Simulated versus measured monaural room impulse responses compared for S3R1, left ear.

While only selected source/receiver combinations are shown here for the RIR, BRIR, frequency response and acoustic parameters, a comparable level of match between simulations and measurements is seen across all source/receiver combinations.

3.4 Computational performance

In this section, we analyze the computational performance of the numerical framework in a general way.
The runtime of the simulation engine is dictated by

- the frequency range of the simulation (grows roughly with constant*frequency^4),
- the size of the domain (grows roughly linearly with the size of the domain),
- the simulated IR length (grows linearly with the simulated time length),
- in some cases, the geometrical features of the room can have an impact. When there are fine geometrical details much smaller than the wavelength of the highest frequency of interest, the time step will have to be reduced.

The simulation runtime of an empty 50 m3 cubic domain is reported in Table 1, when using a single Nvidia A100 GPU. These numbers can then linearly be extrapolated to estimate the runtime for smaller/larger domains, and longer/shorter impulse response lengths.

Table 1. Measured runtime of the simulation engine when using a single Nvidia A100 GPU to simulate a 50 m3 cubic domain.

	Room IR length		
Upper frequency	**0.2 sec**	**0.5 sec**	**1 sec**
800 Hz	10 sec	25 sec	50 sec
1600 Hz	3 min	7 min	13 min
3200 Hz	48 min	2 hours	4 hours

The multi-GPU scalability is analyzed in terms of *weak scaling* computational efficiency. Weak scaling is characterized by how the runtime changes when fixing the problem size for each processing unit. Thus, the problem size is increased when increasing the number of processing units, to keep the overall workload constant on each processing unit. The weak scaling computational efficiency is given by $\eta = Trb / Tr$, where *Trb* denotes the runtime of a baseline case and *Tr* denotes the runtime for the comparison case. We consider a weak scaling test where 1M mesh elements are used per GPU. The resulting weak scaling efficiency is shown in Fig. 8. Excellent weak scaling efficiency is observed. This indicates that the simulation engine can be used to efficiently solve nearly-arbitrarily large problems, as long as an appropriate amount of computational resources is chosen.

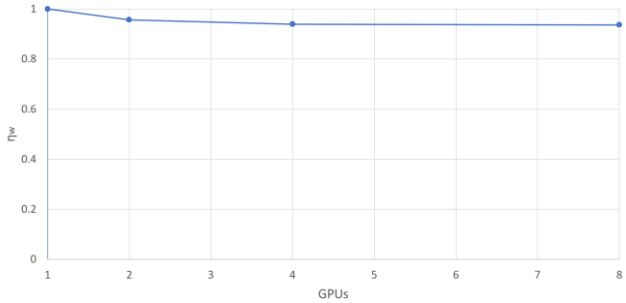

Figure 12. Weak scaling computational efficiency of the proposed DGM framework.

4. CONCLUSION

A wave-based simulation framework has been presented. The framework is based on using a time domain DGM formulation and can handle frequency dependent boundaries and directional sources and receivers. Experimental validation for the source modeling and receiver modeling approaches are shown, as well as a room simulation benchmark study which demonstrates that the framework can produce accurate results for a broadband room acoustic simulation with directional sources and receivers in an 80 m3 room.

5. REFERENCES

[1] L. Savioja, U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, 138(2), pp. 708-730, 2015.

[2] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *The Journal of the Acoustical Society of America*, 98(6), pp. 3302–3308, 1995.

[3] F. Pind, C.-H. Jeong, A. P. Engsig-Karup, J. S. Hesthaven, J. Strømann-Andersen, "Time-domain room acoustic simulations with extended-reacting porous absorbers using the discontinuous Galerkin method," *The Journal of the Acoustical Society of America*, 148(5), pp. 2851-2863, 2020.

[4] J. Ahrens and C. Andersson, "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," *The Journal of the Acoustical Society of America*, 145(4), pp. 2783–2794, 2019.

[5] J. Sheaffer, M. van Walstijn, B. Rafaely and K. Kowalczyk, "Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2125-2135, 2015.

[6] B. Rafaely, "*Fundamentals of Spherical Array Processing,"* Springer Topics in Signal Processing, vol. 8, 2015.