DOI: 10.61782/fa.2025.0028

FORUM ACUSTICUM EURONOISE 2025

A PRELIMINARY EVALUATION OF RENDERING TECHNIQUES FOR ROOM IMPULSE RESPONSES

Rahul Roy Chowdhury[1,2*] **Isaac Engel**[2] **Lorenzo Picinali**[1]

[1] Audio Experience Design, Imperial College London, United Kingdom
[2] Huawei Munich Research Center, Germany

ABSTRACT

This paper provides a review of various Room Impulse Response (RIR) synthesis methods for auralisation applications, followed by a comparative analysis of selected ones using a common framework. This is driven by the rapid expansion of the extended reality (XR) industry and its diverse applications. In XR implementations, realistic immersion is typically achieved when acoustic and visual cues are congruent. The challenge in achieving this is related to computational complexity, acoustical parameter estimation, and, more generally, perceptual relevance of processing choices. Focusing on reverberation, the review comprises geometric-based, data-driven, and algorithmic methods. These differ, among other things, in input requirements: some rely on physical measurements, some on a high-level description of the room's geometry, while others necessitate a 3D mesh. The latter is of particular significance within the XR community, as the impact of incorporating visual information into the synthesis process on the overall spatial quality of a virtualised scene remains uncertain. Objective comparisons of these methods are made using real-world recordings following the work of previous round-robin studies. This paper aims to provide a comprehensive resource for researchers working on immersive audio and acoustic simulation, highlighting the advances made over the past decade.

Keywords: *Room Impulse Response Synthesis*

**Corresponding author*: *rahul.roy.chowdhury@huawei.com.*

1. INTRODUCTION

Room Impulse Responses (RIRs) are critical in various audio applications, such as extended reality (XR), spatial audio rendering, speech enhancement, and architectural acoustics. RIRs capture the way sound propagates from a source and interacts with its environment, including reflections, diffractions, and scattering. RIR synthesis for acoustic simulations must be done with enough accuracy to achieve plausible results. Traditionally, RIRs are obtained through real-world measurements, but this approach is time-consuming and impractical for real-time or large-scale applications. As a result, synthetic methods for generating RIRs have gained significant attention in recent years given the rapid advancements in computational capabilities.

Attempts to synthesize room acoustics date back to the 1960s with Schroeder's Reverb, an artificial reverberation technique that uses digital filters to simulate reverberation [1]. The first physics-driven model for RIR synthesis utilised Ray Tracing (RT) inspired by research in optics [2]. This was followed by the Image Source Method (ISM), which Allen & Berkley had developed in 1979 [3]. Approximating sound waves as rays works for high frequencies, but the deviation from wave-like behavior is prominent in lower frequencies. This paved the way for wave-based techniques such as Finite Difference Time Domain (FDTD). For a comprehensive review on the history of room acoustic synthesis methods, the reader is referred to [4–7]. At present, machine learning-based RIR synthesis, and hybrid geometrical methods have emerged as powerful alternatives that leverage real-world data to improve realism, albeit at the expense of increased computational complexity and data requirements.

This paper reviews a selection of relevant RIR synthesis techniques, highlighting their principles, advantages, and limitations. Section 2 discusses various RIR synthesis methods, and Section 3 presents a framework for a comparative analysis of four such methods. Section 4 presents the results of the analysis, and Section 5 concludes with a discussion of current challenges and future research directions in the synthesis of RIR. This exploration is part of a larger study which will also evaluate BRIRs and conduct subjective tests.

2. RIR SYNTHESIS METHODS

While several techniques are available for RIR synthesis, including those that do not necessarily rely on physical room and acoustic properties, this report will focus on methods that integrate these physical aspects. While other approaches, such as machine learning-based methods that use images as input, may offer distinct advantages, they fall outside the scope of this discussion. Furthermore, the methods chosen for the evaluation were selected from commercially available and open source solutions.

2.1 Perceptually Driven Methods

Historically, artificial reverberation gained popularity with Schroeder's Reverb Algorithm, a digital interconnection of comb filters and delay lines. Their popularity led to the development of Feedback Delay Networks (FDNs) [8], which consist of interconnected feedback, delay and filter stages tuned to create a comb filter, mimicking the temporal effects of reverberation. The emphasis was on creating perceptually convincing results with low computational complexity. Developments have allowed tuning the filters to account for frequency-dependent decay, and further modifications have reduced instabilities in the feedback filter that give rise to resonances and metallic ringing sounds. These techniques are popular in music technology due to their efficiency in generating plausible reverberation.

2.2 Wave-Based Modeling

Wave-based methods attempt to solve the wave equation numerically and are by far the most accurate method for RIR synthesis since all the wave-like properties of sound are inherent in such a model. One downside of wave-based methods is that they are computationally expensive for RIR synthesis (making them unsuitable for real-time applications), especially at higher frequencies where the smaller wavelengths require finer spatial resolution. The most popular techniques are Finite Difference Time Domain, Finite Element Methods and Boundary Element Methods. These methods are not considered in the evaluation due to their high computational cost. The study focuses on techniques that are capable of real-time or near-real-time rendering.

2.3 Geometrical Acoustics

Geometrical methods treat sound waves as rays and approximate well for high frequencies, where acoustic reflections and occlusion are more relevant, and wave-based phenomena such as diffraction are less frequent. Image Source Method (ISM) is a geometrical method with a deterministic approach as it models reflections by representing reflective surfaces as virtual sources. ISM does not support scattering and only creates specular reflections, which work well for early reflections. Ray tracing (RT) is a geometrical method with a stochastic approach that operates by tracing the paths of individual rays as they interact with surfaces. This technique models scattering by introducing randomness into the reflection angles of sound rays when they interact with surfaces.

2.4 Hybrid Methods

Currently, several implementations such as WayVerb, GWA [9,10] and Treble [1] use hybrid approaches that combine the speed and accuracy of wave-based synthesis at low frequencies, capturing diffraction and other wave effects, with geometric methods at high frequencies, where sound waves behave more like rays, to synthesize RIRs quickly. Some even combine geometrical acoustics for the early part and then model the late part using perceptually driven artificial reverberators. Scattering Delay Network (SDN) [11, 12] is an example of this, since it models the early reflections with image sources and the late part is modeled using a network of feedback and delay blocks.

2.5 Data Driven Methods

The synthesis of RIR using a machine learning model has attracted a great deal of attention over the past decade. This has been driven mainly by the desire to improve the algorithms for speech enhancement and automatic speech recognition with data augmentation [13–15], and to investigate whether the addition of visual cues can improve the

[1] Treble Technologies - Acoustic Simulation Platform for Machine Learning & AI

synthesis of RIR [13, 16]. One limitation of most of these techniques is that they have been trained with a dataset created using geometric methods and therefore exhibit deviations at lower frequencies. Additionally, these implementations prioritize data augmentation over the physical accuracy of the synthesized RIR. For example, MESH2IR [13] is an implementation which takes a 3D mesh of the room as input, along with source and receiver positions, and it outputs an impulse response with a length of 0.25 s at $f_s = 16\,\text{kHz}$. Aside from the short length and low sampling rate, the model does not rely on the acoustic properties of the materials. As a result, the generated RIRs may appear plausible but may not accurately represent the scene, either physically or perceptually.

3. OBJECTIVE EVALUATION

3.1 Framework

The model of the room is selected from the BRAS dataset [17], which was used in a round robin comparison of geometric-based RIR synthesis by Brinkmann et al. [18]. Among others, the dataset features a shoebox-like empty room with an approximate volume of $145\ \text{m}^3$. This room was selected for the comparative study since it also provided absorption and scattering coefficients at $1/3$ octave bands for all of its materials from 20 to 20000 Hz. The dataset for the room contains 32 monaural measurements comprising 2 source positions with 4 orientations (Genelec 8020c speakers) and 4 receiver positions (1/2" B&K 4134 diffuse-field measurement microphone). The measured impulse responses were upsampled to 48 kHz (originally, they were 44.1 kHz) and truncated to 3.5 s. A 3D model of the room is provided in the .skp format and is shown in Figure 1. Since two of the methods mentioned below do not support source directivity, the speaker orientation closest to the microphone's direction was chosen, resulting in 8 measurements in total. The ambient temperature was measured to be $19.5°$ C, along with 41.7% humidity.

3.2 Methods

Four methods were selected for the comparative study, differing in terms of their approach for RIR synthesis and their computational times. Data-driven methods such as MESH2IR were considered but not evaluated due to several restrictions, such as a low sampling rate (16 kHz) and the length of the RIR (0.25 s). **Real-time-SDN** is an implementation based on [11, 12] and is capable of real-time rendering using the SDN approach. **RAZR** [19] uses ISM for the early part and FDN for the late part, **RAVEN** [20] also computes image sources for the early part but synthesizes the late part with ray tracing. Lastly, **WayVerb** [9] employs a similar approach to RAVEN for synthesizing high frequencies, while low frequencies are generated using a wave-based synthesis technique called Digital Waveguide Mesh (DWM) [21]. The transition between geometric and wave-based approach was set to 500 Hz. For the remainder of the study, they will be referred to as methods A, B, C and D, respectively.

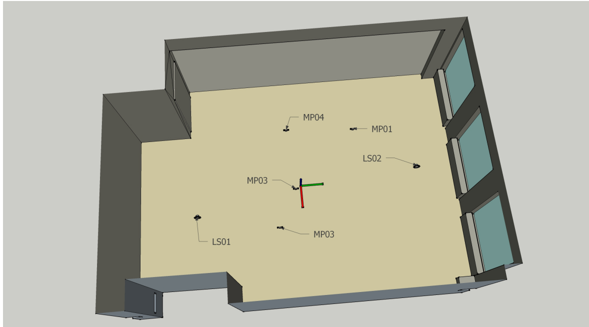

Figure 1. 3D rendering of Scene 9 from the BRAS dataset. The scene consists of a seminar room with 3 window panels on one side. The positions of the speakers and the microphones are also shown (LS = Loudspeakers, MP = Microphone).

The methods were also selected on the basis that they were all provided with a similar amount of information about the room and its acoustic properties. All the methods were set to create up to 3rd order image sources, except method A, which was fixed to 1st order. Furthermore, all methods considered frequency-dependent absorption and scattering, except A and B, which did not take scattering effects into account. Methods C and D support complex geometries with 3D mesh as input, while A and B require rectangular dimensions for the room. Since methods C and D take mesh as input, they are the only methods that correctly assign the labelled acoustic properties to the materials. For the others, the properties were manually assigned to each of the six surfaces. Since this evaluation is concerned with monaural RIRs, the receiver's directivity was also omitted. A summary of this is presented in Table 1.

Table 1. Summary of four methods chosen for evaluation.

Method	A	B	C	D
Name	Real Time SDN	RAZR	RAVEN	WayVerb
Year	2024	2017	2021	2017
Model Early Part	ISM	ISM	ISM	ISM + DWM
Model Late Part	SDN	FDN	RT	RT + DWM
Supporting Geometry	Shoebox	Shoebox	Mesh	Mesh
Absorption Coefficients	Yes	Yes	Yes	Yes
Scattering Coefficients	No	No	Yes	Yes
Source Directivity	No	Yes	Yes	No
Receiver Directivity	Yes (N/A)	Yes (N/A)	Yes (N/A)	Yes (N/A)

4. RESULTS & DISCUSSION

Each method was employed to synthesize 8 RIRs at a sampling rate of 48 kHz. The performance of the methods was evaluated based on:

- Reproducibility of Early Reflections
- Accuracy of Energy Decay Curve (EDC)
- Room Acoustic Parameters

The early reflections and EDC are evaluated for a single measurement (LS02 and MP03 from Figure 1), while the acoustic parameters are presented as an average across the 8 measurements. The room parameters include T20, T30, Early Decay Time (EDT), Clarity, Definition and Center Time. The first three are indirect methods for calculating the reverb time. Clarity is a measure of the proportion of the early sound in comparison to the late, expressed in dB, and it is typically used to evaluate the speech intelligibility in a room ($h(t)$ is the impulse response):

$$C_{80} = 10 \log_{10} \left(\frac{\int_0^{0.08} h^2(t)\, dt}{\int_{0.08}^{\infty} h^2(t)\, dt} \right) \tag{1}$$

Definition refers to the percentage of energy the early sound contributes to the impulse response; it is closely related to the clarity, but expressed in percentage and also relates to speech intelligibility. A higher percentage indicates better speech intelligibility due to a stronger presence of early reflections:

$$D_{50} = \frac{\int_0^{0.05} h^2(t)\, dt}{\int_0^{\infty} h^2(t)\, dt} \cdot 100\% \tag{2}$$

Finally, the center time refers to an instant in time where the total energy is equal both before and after, also known as the center of gravity. A low value typically indicates a strong presence of early reflections:

$$T_S = \frac{\int_0^{\infty} t \cdot h^2(t)\, dt}{\int_0^{\infty} h^2(t)\, dt} \tag{3}$$

In terms of rendering time, method A took less than 5 ms, followed by method B, which took around 1.4 s on average. Method C took around 11 s while method D was the longest, averaging around 5 minutes, probably due to the inclusion of a wave-based synthesis at low frequencies. All simulations were run on a computer equipped with a modern processor with no GPU acceleration. All RIRs were post-processed for time alignment and normalization, ensuring the direct sound arrived at the same time with a peak amplitude of 1.

4.1 Early Reflections

The early reflections for all the methods along with the reference for a single source/receiver location are presented in Figure 2. The first 20 ms of the RIR are presented to evaluate the presence of the first 6 specular reflections. W_{Right} & W_{Left} correspond to the walls on the positive and negative X axis in Figure 1 (X axis is the line parallel to the surface with the windows). Visual inspection suggests that all methods reconstruct most of the early reflections successfully; the slight differences in their time of arrival (ToA) can be attributed towards estimation error of the room's geometry and measurement error for the location of the source and receiver. The reflections arising from method B do not appear to be as salient as the other methods; perhaps this is due to the filtering that is being

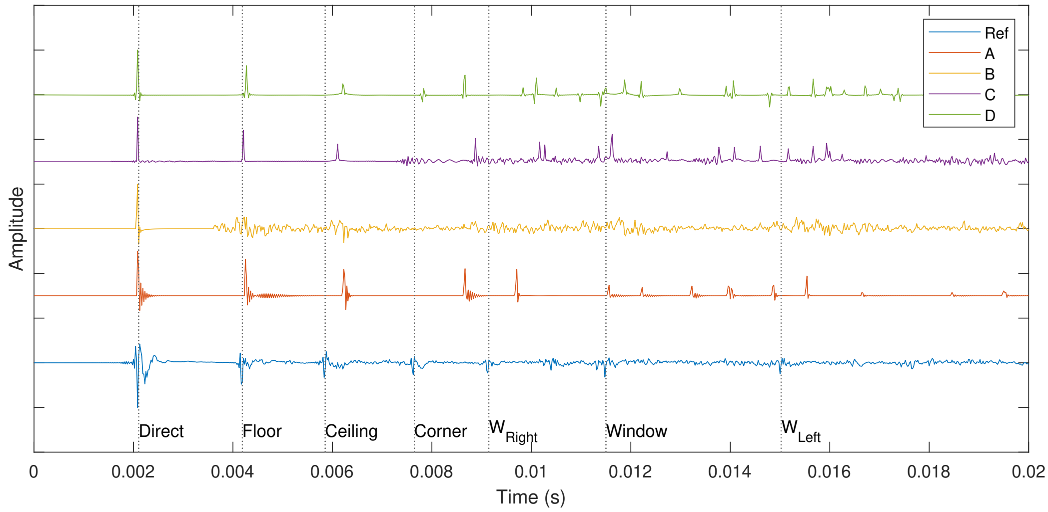

Figure 2. First 20 ms of the monaural RIRs for the measurement (LS02 and MP03) and the RIR synthesis methods. The first 6 reflections are labelled starting from the floor and until the left wall.

done to the image sources. The "Corner" refers to the reflection occurring after the direct sound has passed the microphone and reflected from the corner wall directly behind (see Figure 1). Since this level of detail in the geometry is only considered in methods C & D, the reflection is missing in A & B. All the methods use a variation of ISM for the early reflections; the main exception being method D, which uses a wave-based technique to model the low frequencies.

4.2 Energy Decay Curve

The EDC for 6 octave bands spanning from 125 Hz to 4 kHz for the RIRs explored above is presented in Figure 3. At $f_C = 125$ Hz, all the methods underestimate the decay. The methods begin to converge from $f_C = 250$ Hz, where methods C and D are close to the reference, while method B underestimates the decay and method A largely overestimates. Method B begins to converge towards the measured slope, along with the other methods, at higher frequencies. FDN in method B is a lower-complexity technique that achieves similar EDC to more complex methods like RT for monaural RIRs. If we were to evaluate RIRs containing spatial information (e.g. BRIRs), then the FDN would likely miss directional components of the late reverberation due to its simplicity. Method D estimates decay most accurately at lower frequencies, likely due to the inclusion of a wave-based model. Aside from model A underestimating the slope for the first 4 bands, the other methods begin to converge at the mid-frequency bands, and then the slope is overestimated towards the higher frequencies. Both deviations (underestimation at low frequency and overestimation at high frequency) probably result from inaccurate estimation of the absorption and scattering coefficients. The BRAS dataset documentation revealed that the properties of the acoustic materials were acquired by measurement, extrapolation, and matching from an external database.

4.3 Room Acoustic Parameters

The deviations from the reference are even more evident in Figure 4, which shows the Room Acoustic Parameters: T20, T30, Early Decay Time (EDT), Clarity (C80), Definition (D50) and Center Time (T_S). By visual inspection, it is clear that the slope of the decay does not change much between -20 dB (T20) and -30 dB (T30). The low frequency RT60s are overestimated for all the methods, confirming the observations made from the EDC curve. Similarly, the RT60 at 4 kHz and 8 kHz appear to be underestimated for methods A and C, in agreement with the EDC curve. The mid-frequency values appear close to the reference, except for method A. Brinkmann et al. [18] found similar trends in the round-robin com-

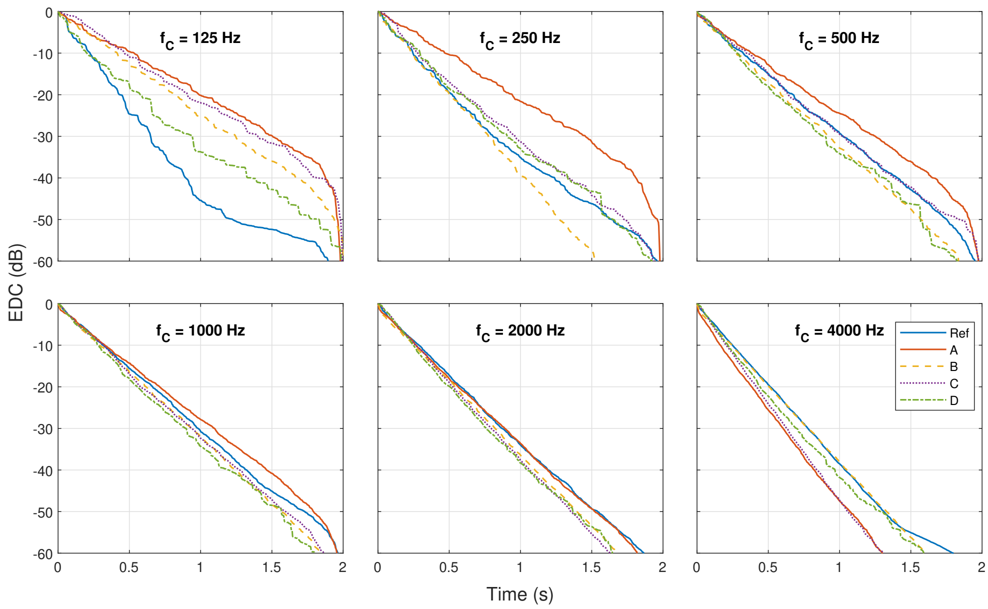

Figure 3. Energy Decay Curve for the RIRs of the measurement (LS02 and MP03) and the synthesis methods.

parison for this room. They reported that the parameters matched (within the critical perceptual threshold 5 %) in the medium frequency range (500 Hz - 2 kHz), with overestimation at lower frequencies and underestimation at higher frequencies. According to them, this occurred because of inaccurate absorption coefficients, which were only optimal for the mid-frequency range. Method A estimates the RT60 to be around 3 seconds in the 125 Hz octave band, twice as high as the reference signal. One possible reason method A overestimates RT60 at low frequencies is that the model does not account for surface scattering. Fontana et al [11] used method A for RIR synthesis and computed the T60 for the same room from the BRAS dataset with a single reference (not averaged) and found no significant deviation across all frequency bands. However, it is unclear how they estimated T60, since the reference was supposed to be around 1.5 s at the 250 Hz octave band (consistent with Brinkmann et al. [18]), whereas they reported it to be closer to 2.0 s. The EDT is similar across all methods, with method D closely matching the reference at low frequencies, indicating accurate modeling of the initial decay by the hybrid method.

Looking at clarity, all methods converge at the $f_C = 1$ kHz, and begin to diverge towards higher and lower frequencies. On average, there is about 6 dB underestimation at $125 - 250$ Hz bands for methods A and C. Both these methods do not account for wave-based phenomena, which play a crucial role at lower frequencies. Interestingly, methods A and B are based on delay networks, yet method B follows the reference much more closely. Perhaps, method A requires some tuning to achieve a closer match. Methods B and D follow the clarity scores recorded for the reference up until the 8 kHz band, where they underestimate by around 2 dB. Methods B and D appear to estimate the clarity well with an average deviation of 1.4 dB across all bands from the reference. The overestimation of method A at higher frequencies likely stems from shortcomings of the implementation.

Visually, the trends for definition appear to be very similar to clarity, suggesting that both are comparable measures of speech intelligibility. One noticeable difference here is that method B appears to overestimate the definition for 250–4000 Hz bands by around 10%. This observation is consistent with the early reflections found in Figure 2, where method B appears to have a considerably higher amount of signal content compared to the other methods. Whether or not these differences are perceivable can be confirmed by a subjective study.

The center time also shows deviations at the low frequency bands 125–250 Hz. Methods A and C overestimate T_S by a factor of $2-2.5$. A larger value suggests that the room will be perceived as more reverberant. Method

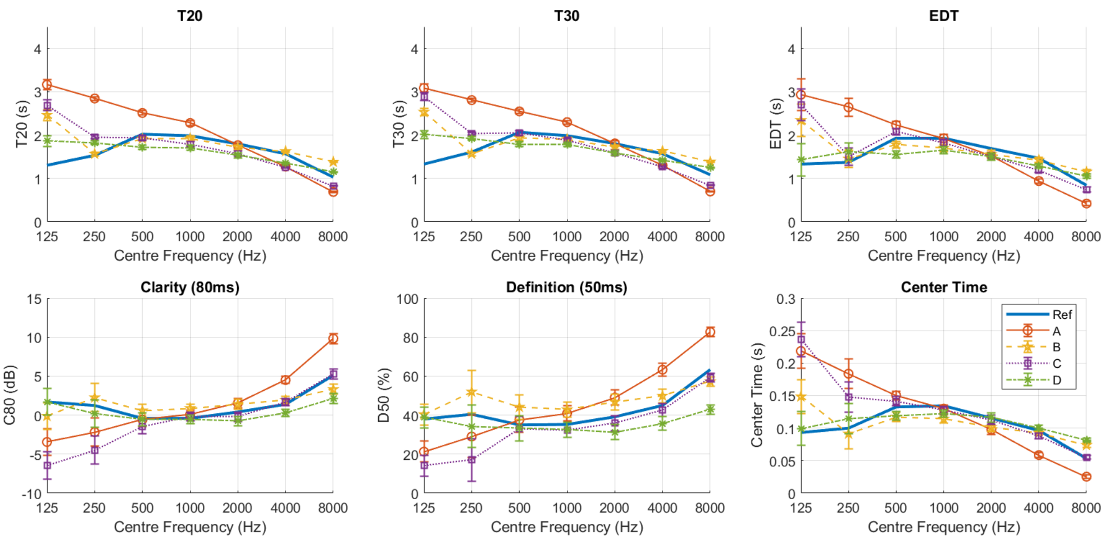

Figure 4. Room acoustic parameters, the error bars represent one standard deviation from the mean.

B overestimates by a factor of 1.5. The lack of diffraction and modal behavior at lower frequencies for these models is probably the cause of this deviation. Method D is able to estimate the center time quite well up to 4 kHz with an overestimation at 8 kHz.

5. CONCLUSION AND FUTURE WORK

In this study, four distinct methods for RIR synthesis were objectively evaluated against a reference measurement. The evaluation considered the timing of early reflections, EDC and some room acoustic parameters. For the early reflections, the comparison revealed that methods A, C and D were able to reconstruct the early reflections, despite having differences in the time of arrival. This discrepancy may be attributed towards inaccurate geometry with insufficient representation of the diffraction phenomena, something that was also found in Brinkmann et al. [18]. Method B contained significant noise early on, making it hard to visually identify the arrival time of early reflections in the RIR.

The room parameters and the EDC curve served as a useful tool to evaluate the late part of the RIR. Here, it was found that all methods largely overestimated RT60 at low frequencies. This is likely due to all models, except method D, not using a wave-based approach. For method A, this could be due to an error in the implementation; perhaps the parameters for modelling the late part were not tuned optimally. Another source of error could be the challenge of estimating the acoustic properties of materials at lower frequencies, as these are difficult to measure due to large-scale interactions. Over the mid frequency bands ($500 - 2000$ Hz), most of the methods were close to the parameters corresponding to the reference, in agreement with the results presented in Brinkmann et al. [18]. Method A resulted in the largest overestimation of RT60 at the lower frequencies, and in general deviated the most across all six parameters. Methods B and D produced parameters closest to the reference values, with method D performing better at low frequencies due to the inclusion of a wave-based model.

Future research will explore the perceptual impact of these shortcomings through subjective listening tests, helping to validate the objective results and connect them to real-world applications. Evaluating spatial features of RIRs as well as binaural signals will enable the assessment of metrics relevant to XR applications, such as externalisation. The current results are based on a single rectangular-like room with no furniture, so investigating more complex room setups would be valuable. Additionally, while material properties were precisely defined in this study, real-world surfaces have more variable acoustic properties, and it would be interesting to see how approximating this can affect the results.

6. REFERENCES

[1] M. Schroeder and B. Logan, ""colorless" artificial reverberation," *IRE Transactions on Audio*, vol. AU-9, no. 6, pp. 209–214, 1961.

[2] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.

[3] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 04 1979.

[4] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[5] V. Välimäki, J. Parker, L. Savioja, J. Smith, and J. Abel, "More than 50 years of artificial reverberation," in *Proc. 60th International Conference of the Audio Engineering Society* (S. Goetze and A. Spriet, eds.), (United States), Audio Engineering Society, 2016.

[6] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, pp. 708–730, 08 2015.

[7] A. G. Prinn, "A review of finite element methods for room acoustics," *Acoustics*, vol. 5, no. 2, pp. 367–395, 2023.

[8] J. O. Smith, *Physical Audio Signal Processing*. http://ccrma.stanford.edu/ jos/pasp/, Accessed Date: 26/03/2025. online book, 2010 edition.

[9] M. R. Thomas, "Wayverb: A graphical tool for hybrid room acoustics simulation," Master's thesis, University of Huddersfield, January 2017.

[10] Z. Tang, R. Aralikatti, A. J. Ratnarajah, and D. Manocha, "Gwa: A large high-quality acoustic dataset for audio processing," in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proc.*, SIGGRAPH '22, p. 1–9, ACM, Aug. 2022.

[11] M. Fontana, G. Presti, D. Fantini, F. Avanzini, and A. Reyes-Lecuona, "A highly parametrized scattering delay network implementation for interactive room auralization," in *Proc. of the 27-th Int. Conf. on Digital Audio Effects (DAFx24)*, pp. 286–293, 09 2024.

[12] E. De Sena, H. Hachabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1478–1492, 2015.

[13] A. Ratnarajah, Z. Tang, R. Aralikatti, and D. Manocha, "Mesh2ir: Neural acoustic impulse response generator for complex 3d scenes," in *Proc. of the 30th ACM International Conference on Multimedia*, MM '22, p. 924–933, ACM, Oct. 2022.

[14] A. Ratnarajah, Z. Tang, and D. Manocha, "Ts-rir: Translated synthetic room impulse responses for speech augmentation," 2021.

[15] A. Ratnarajah, S.-X. Zhang, M. Yu, Z. Tang, D. Manocha, and D. Yu, "Fast-rir: Fast neural diffuse room impulse response generator," 2022.

[16] N. Singh, J. Mentch, J. Ng, M. Beveridge, and I. Drori, "Image2reverb: Cross-modal reverb impulse response synthesis," 2021.

[17] F. Brinkmann, L. Aspöck, D. Ackermann, R. Opdam, M. Vorländer, and S. Weinzierl, "A benchmark for room acoustical simulation. concept and database," *Applied Acoustics*, vol. 176, p. 107867, 2021.

[18] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *The Journal of the Acoustical Society of America*, vol. 145, pp. 2746–2760, 04 2019.

[19] T. Wendt, S. van de Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *Journal of The Audio Engineering Society*, vol. 62, pp. 748–766, 2014.

[20] D. Schröder and M. Vorlaender, "Raven: A real-time framework for the auralization of interactive virtual environments," *Proc. of Forum Acusticum*, pp. 1541–1546, 01 2011.

[21] J. O. Smith, *Physical Audio Signal Processing*. http://ccrma.stanford.edu/ jos/pasp/, Accessed Date: 26/03/2025. online book, 2010 edition.