



Differences between measured and simulated room impulse responses

Lukas Aspöck¹ and Michael Vorländer²

Institute for Hearing Technology and Acoustics, RWTH Aachen University
Kopernikusstraße 5, D-52074 Aachen, Germany

ABSTRACT

Simulation models based on geometrical acoustics mostly do not immediately deliver a simulated room impulse response, but intermediate results such as an energy histogram or an energy decay curve. At this point, further models are required to generate a room impulse response of the simulated environment, which are essential for the process of auralization. While for various simulated scenarios, the application of a reflection model based on a basic theory is sufficient, detailed comparisons of simulated and measured room impulse responses reveal differences which are attributed to the lack of diffuse reflections in the applied reflection model. These deviations can also substantially affect derived room acoustic parameters. This work presents and explains the shortcomings and discusses potential improvements to be considered in the simulation process.

1. INTRODUCTION

Computer simulations for the prediction of a room's acoustics have been developed many decades ago and now are an established tool for acoustic consultants, integrated in professional software environments. While the most typical application is the calculation of room acoustic parameters such as the reverberation time, simulation tools also offer the possibility to obtain a room impulse response (RIR). In recent validation studies involving simulation models based on Geometrical Acoustics (GA), simulated RIRs have been compared to the corresponding measured RIRs [1]. For such comparisons it needs to be considered that the measured RIR should not directly be compared to the simulated RIRs, unless at least the most relevant data which makes a measurement not correspond to the ideal situation, i.e., the sound source directivity, is also considered in the simulation. However, even a qualitative comparison of measured and simulated RIRs can help to identify shortcomings in the simulation process. With respect to the Round Robin results, it was discussed in how far the RIR synthesis method led to relevant deviations of room acoustic parameters if the simulated RIRs were used for the parameter evaluation [2]. If used for auralizations, an inadequate RIR synthesis might also introduce audible colorations. This paper revisits these investigations and focuses on the observed differences in temporal structure between measured and simulated RIRs, mostly discussing the synthesis of the early diffuse part and the late part of the RIR.

¹ las@akustik.rwth-aachen.de

² mvo@akustik.rwth-aachen.de

2. ROOM IMPULSE RESPONSE SYNTHESIS

To obtain broadband RIRs based on wave-based simulations models, which typically provide results in the frequency domain, an inverse Fourier-Transformation of the frequency response needs to be conducted. This is rarely done as the computational effort to calculate results for the full audible frequency range is very high, despite powerful CPUs and computer clusters being available [3]. For, in some cases, more efficient wave-based simulations in the time-domain, such as the finite difference time domain method (FDTD), the RIR can be obtained by setting a Dirac impulse as the pressure signal of the sound source point of the FDTD grid [4]. In addition to the high computational effort, for wave-based simulation it is not impossible but challenging to also account for directional characteristics of the sound source and receiver and integrate them into the RIR. The synthesis of RIRs based on GA models is less challenging, but also depends on the chosen method.

2.1. Image sources: Early (specular) reflections

Many of the GA-based simulation tools use the image source model [5] to obtain the direct sound and the early reflections. The calculation of each audible image source contribution to the RIR is straightforward: According to the distance of the image source to the receiver of the scene and the related surface absorption data, the delay of the impulse as well as the amplitude can be obtained. This process becomes only slightly more complicated when also directional data of sound source and receiver (Head-Related Transfer Functions) and air absorption is additionally considered. The full synthesis, processed in the frequency domain, for a set of audible image sources is illustrated in Figure 1. This method leads to a high temporal accuracy of each inserted reflections, only limited by the chosen sampling rate of the audio processing environment. The time delay, which is not included in the figure, is easily obtained by division of the distance r by the speed of sound.

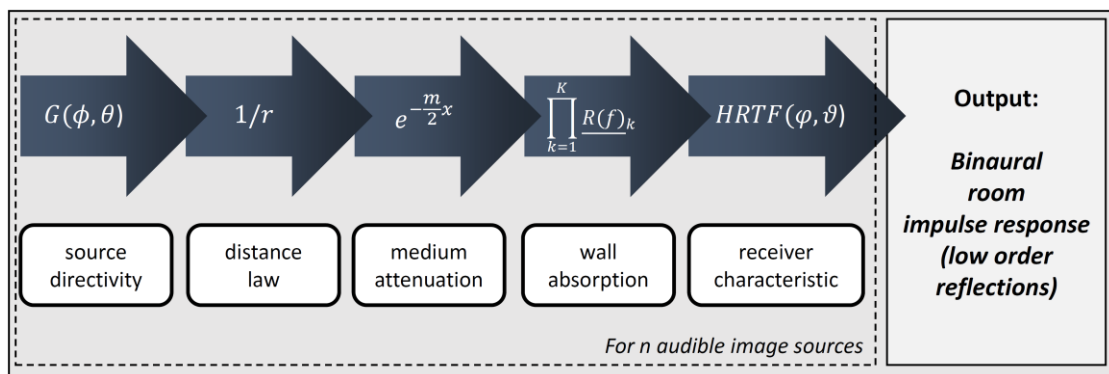


Figure 1: Filter synthesis process for image sources in the frequency domain.

2.2. Late (diffuse) reflections

Even without any application of a ray-based model the later part of the RIR can be synthesized based on an estimated reverberation time. Moorer developed a RIR synthesis already in 1979 using stationary white noise multiplied by an exponential decay based on frequency-dependent reverberation times [6]. While a RIR generated by this process would correctly model the decay time, it does not account for any room-dependent reflection patterns.

This can be achieved by different methods such as ray or beam tracing or the radiosity approach, which all generate an energy decay for the investigated room, typically with a lower temporal resolution resulting in an energy histogram, for which the energy of multiple paths is summed up in a time bin. This energy histogram is often mistaken for the energy impulse response of the room. Due to its low temporal resolution (length of a time bin is typically in the range of 1-10 ms), it does not include information about the actual reflections but only models the decay process of the energy in the room. Reducing the length of a time bin and therefore increasing the temporal resolution, e.g., up to the chosen audio sampling rate, is not a suitable to obtain a valid RIR (suitable for auralization) as

in these cases, the amount of reflection in the RIR would depend on the chosen number of rays, beams or patches of the simulation process and not on the geometrical properties of the room model.

A reflection density model describing the temporal structure of a RIR can be derived from an image source model of a simple rectangular room, leading to Eq. 1 [7].

$$N_r(t) = 4\pi \frac{c^3 t^2}{V} \left[\frac{1}{S} \right] \quad (1)$$

With c being the speed of sound, t the time and V the volume of the room. Interestingly this equation only includes the volume as the only room parameter, but not the surface area or the mean free path ($4V/S$) of the room.

Without providing a detailed derivation, Kuttruff claimed that Eq.1 also applies to rooms with arbitrary shape [8]. This, however, only under the assumption of isotropic sound incidence or perfectly exponential energy decays [9]. Especially for the early part of the RIR, isotropic sound incidence is not be expected in most rooms, indicating a potential limitation of this reflection density model.

In the implementation of the filter synthesis, this model can be applied as follows: For each sample within a time bin, a random process will decide if a reflection occurs or not, depending current reflection density [10]. This process is currently implemented in the GA-based simulation software RAVEN [11].

3. COMPARISON OF SIMULATED RIRs WITH MEASURED RIRs

The measurement of a RIR is typically conducted according to the ISO 3382 standard using an omnidirectional sound source and a measurement microphone. These measurements, in the standard, are intended for processing room acoustic parameters, but can also be directly analysed. When comparing these RIRs with simulated RIRs, several aspects should be considered:

- The measurement speaker is not a perfectly omnidirectional sound source and deviates from the ideal point source in the simulation with respect to directionality and frequency response.
- The microphone used in measurements is not an ideal point receiver. This effect, however, is negligible, especially compared to the sound sources.
- The simulation's input data has a relevant uncertainty and might not accurately represent the real situation.
- The chosen simulation model might be based on ideal assumptions (e.g., diffuse field conditions) and is often limited by frequency range.

Thus, it is not expected that a simulation's results will perfectly match the measured results. Nevertheless, the comparison between measured and simulation results might reveal shortcomings of the simulation process. In the following sections, GA-based simulations and the corresponding measurements are presented and discussed with respect to deviations.

3.1. Results of the Round Robin on Auralization

In the round robin on auralization [1], in total six participants contributed simulated RIRs for three room scenes, a small seminar room ($V=146 \text{ m}^3$), a medium-sized chamber music hall ($V=3300 \text{ m}^3$) and a large auditorium ($V=8657 \text{ m}^3$). For each scene, the round robin participants had to generate 10 RIRs (two sound source positions and five receiver positions) without any knowledge of the measured results. The room geometry and input data related to the surfaces (absorption and scatterings coefficients for one-third octave bands) was provided to the participants, who all used GA-based software. The applied tools, with one exception, all supported the calculation of specular and diffuse

reflections, diffraction was only considered by two participants and only if the direct sound was blocked by a room element, which was not the case in the three investigated room scenes. The simulated RIRs were primarily intended for the evaluation of room acoustic parameters according to ISO 3382, but could also be compared to the measured RIRs. The evaluation was done by the round robin team. All relevant input data and the measured reference data of the scenes, along with a general documentation is publicly available in the *Benchmark for Room Acoustical Simulation (BRAS)* [12].

The visual comparison of the RIRs for all three rooms in general reveals that the simulated and the measured RIRs, in most cases, have a similar decay process and, after removing systematic time shifts for some participants, correctly model the delay of the direct sound arrival. With respect to the temporal structure in the early part of the results (below 100 ms), the simulated RIRs are often sparser than the measured RIRs. While strong (specular) reflections are mostly accurately modelled by the simulation tools, the energy in between is often missing. This can be partially attributed to the non-ideal sound source conditions of the measurements and to complex-valued surface impedances. Both of these effects cause the reflections to spread out over a longer period of time and to correspond less and less to an ideal Dirac impulse. However, especially for RIRs with a rather low distance between sound source and receiver in relation to the room dimensions, the measured RIRs often include early scattered (and potentially also diffracted) energy in between the specular reflections which is not present in the simulated RIRs. This is not an issue of concern for the small room, but this difference becomes apparent for some of the RIRs of the medium-sized room and the large auditorium.

To further analyze this difference, in this work, an example selection of the large room's results is presented. The top view including two sound source positions LS1 and LS2 and two receiver positions MP1 and MP2 is shown in Figure 2.

**Complex room 4 (CR4, Scene 11)
of the BRAS database**

$V = 8657 \text{ m}^3$
 $S = 5,851 \text{ m}^2$
 $\sim 1200 \text{ Seats}$

Distances:
 $d(\text{LS1-MP1}) : 9.62 \text{ m}$
 $d(\text{LS2-MP2}) : 7.01 \text{ m}$

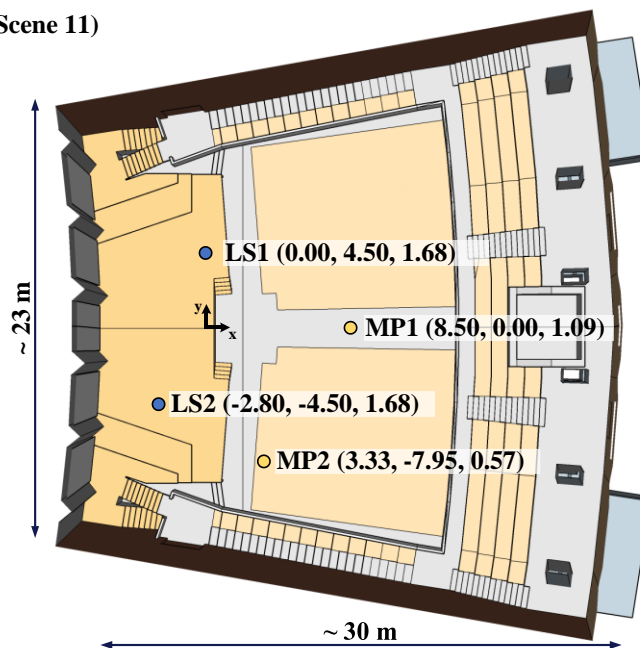


Figure 2: Top view for the large room scene of the Round Robin: Lecture hall at TU Berlin

The analysis of RIRs plotted with a linear y-axis demonstrates the mentioned effect of less pronounced reflection impulses in case of the measured result. In Figure 3 the early part of the measured RIR and the simulated RIRs of four Round Robin participants are shown for positions LS2-MP2. To improve the visibility of the reflection, all RIRs have been multiplied with a factor of 3 just after the direct sound, which arrives at around 21 ms. The results suggest that the applied simulation model of *Sim1* either is different from the other three participants or the input data for this source receiver combination was not correctly entered as three prominent specular reflections between 41

and 43 ms were not detected and instead, two distinct reflections are present shortly after the direct sound. Although the presentation of the result with a linear y-axis might not appropriately demonstrate the magnitude of the difference, it is apparent that especially in the early part, between 21 and 40 ms of the RIR, the measured result contains substantially more energy. Only the result of *Sim4*, to some extent, also accounts for this early energy.

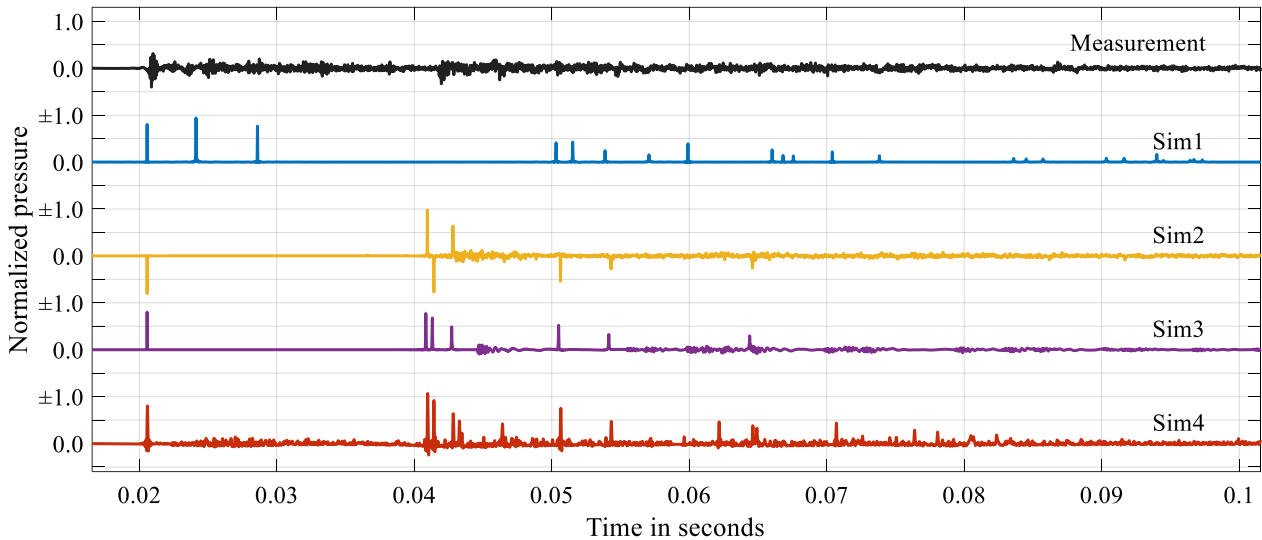


Figure 3: Measured and four simulated RIRs for the large room (CR4), for positions LS2-MP2 from Round Robin investigation. Reflection amplitudes were multiplied by 3 to improve the visibility.

For another RIR, for positions LS1-MP1, similar deviations between simulation and measurement are observed, the results of *Sim2*, *Sim3* and *Sim4*, however, to some extent account for the early scattered energy in between the direct sound at ~28 ms and the first specular reflection at ~57 ms.

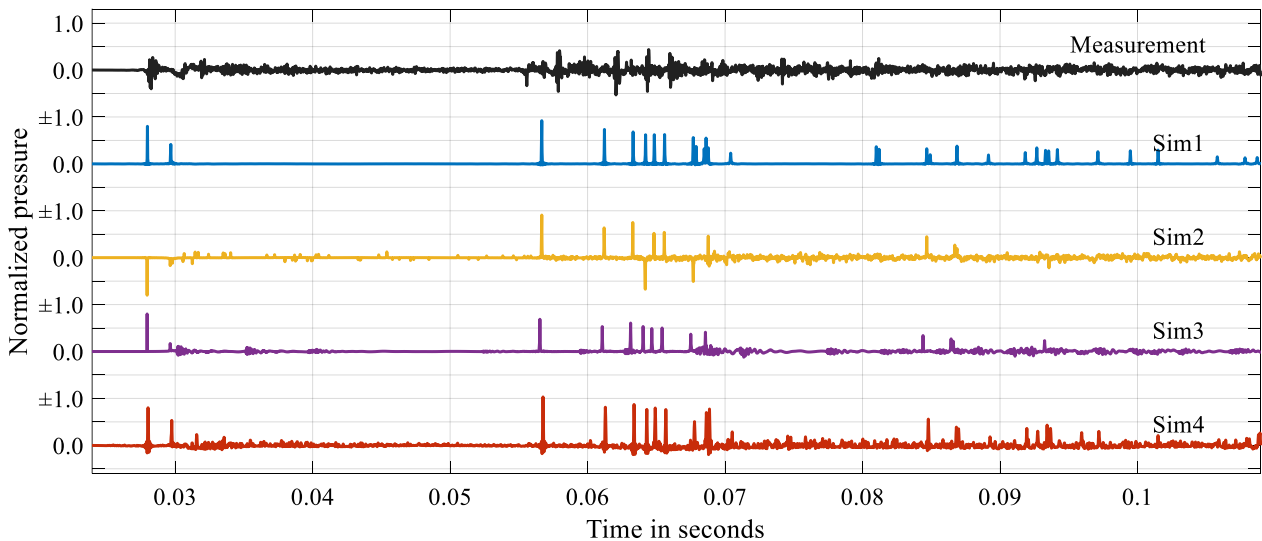


Figure 4: Measured and four simulated RIRs for the large room (CR4), for positions LS1-MP1 from Round Robin investigation. Reflections amplitudes were multiplied by 3 to improve the visibility.

To get a better impression of the magnitude of the deviations, the early RIR part for two participants, *Sim2* and *Sim3*, are shown in comparison with the corresponding measured result in Figure 5. Here, the simulated results show a clear gap of reflections between 21 and 40 ms of the RIR in case of LS2-MP2 (see Figure 5b and d), while for LS1-MP1 the simulated reflections are sparser and of lower amplitude than in the measured RIR during this period (see Figure 5a and c).

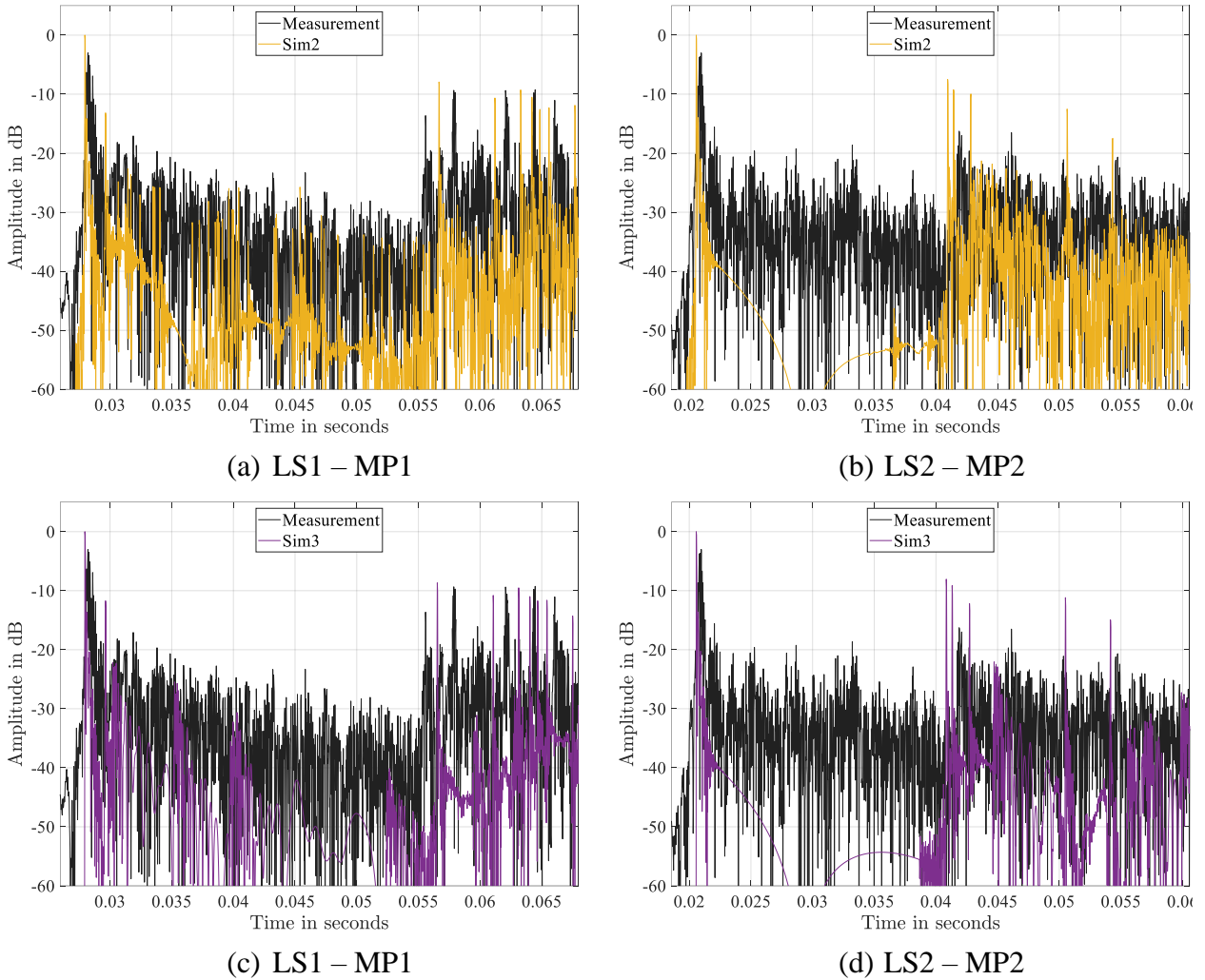


Figure 5: Comparison of simulated and measured room impulse responses for two round robin participants, large auditorium scene (CR4).

As a consequence for the evaluated room acoustic parameters, in case of the LS2-MP2-RIR, the lack of early energy leads to a C80 deviation of up to 5 dB for *Sim2* (for 125 Hz and ~ 2 dB for 250 Hz) and up to 7 dB for *Sim3* (for 125 Hz and ~ 3 dB for 250 Hz).

These deviations cannot exclusively be attributed to the RIR synthesis procedure of these simulation models, the absence of any simulated reflections in case of LS2-MP2 between 21 and 40 ms, however, is an indication that this is not just an issue of inaccurate scattering data for the given scenario.

3.2. Results of informed simulations using the RAVEN software

In a second step, additional simulations of the large room scene were conducted by the authors. As for these simulations, the measured results, as well as the simulated results of the participants, these simulations are considered to be *informed* simulations. The in-house simulation software RAVEN was chosen to conduct the simulations. RAVEN uses a hybrid simulation model combining image sources (early reflections) with a ray tracing algorithm including diffuse rain for early diffuse reflections and the late reflections. The input and output data as well as the corresponding MATLAB scripts to configure and run the simulations can be found in a related data publication [13].

To reduce the impact of potentially incorrect absorption data, the reverberation time T30 of the simulated result was matched to the corresponding measured result. This was achieved by iteratively adjusting the absorption coefficients of all surfaces of the room, leading to an average T30 deviation (for 10 energy histograms) of 2.7%.

The RIR simulation using RAVEN led to the same energy gap in the early part of the RIR as it was observed in the analysis of the Round Robin (see Figure 6a). The following evaluation of the clarity parameter was done in two ways: 1) Based on the simulated energy histogram 2) Based on the RIR. The results of the C80 evaluation (see Figure 6b) show that, after adjusting the absorption coefficients in order to match the measured T30, the C80 values are all very close (within the JND of ± 1 dB) to the measured values for the evaluated octave frequency bands of 1 kHz and higher. For the lower frequency bands, deviations for both evaluation methods of the simulation are observed – the fact that also the C80 values based on the evaluated energy histogram deviate for the frequency bands 125, 250 and 500 Hz by around 1.5 to 2.0 dB from the measured values indicates that the boundary conditions (scattering coefficients) might not be accurately defined for these three octave bands. Another potential source for this deviation is the lack of diffraction modelling in the applied simulation software. The role of the seat dip effect is of less relevance as the receiver position corresponds to the second row.

The deviation, however, is on average around 1 dB lower when C80 is evaluated based on the histogram instead of the RIR. Thus, if the reflection density model is not adjusted, the problem in the temporal structure remains and is relevant, whenever the receiver is rather close to a sound source in a room with a rather large volume (which only has a slowly increasing reflection density).

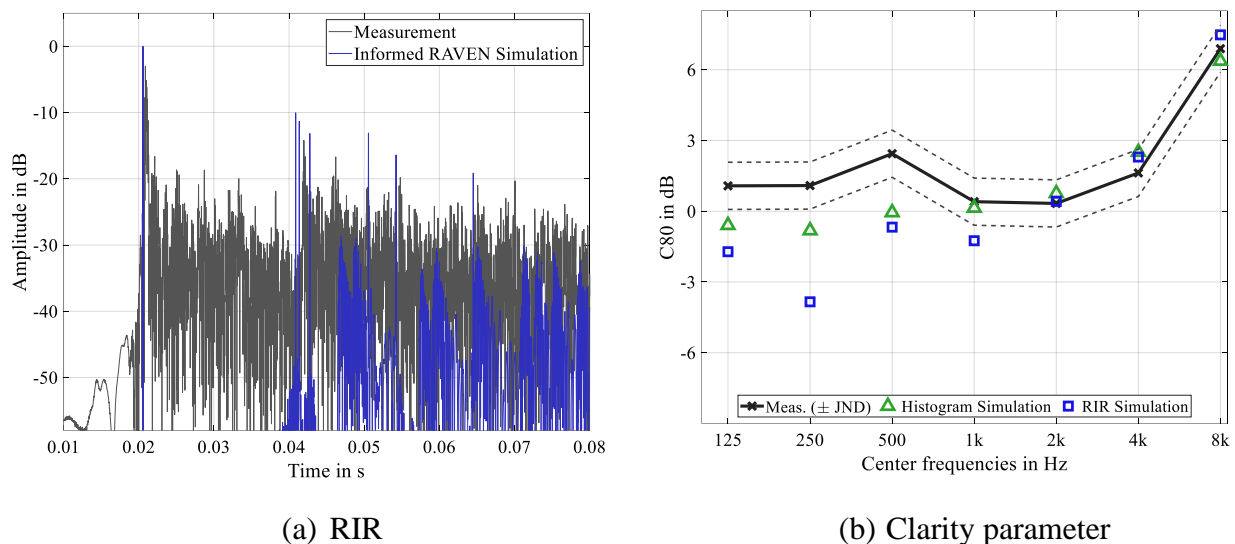


Figure 6: Comparison of measurements and informed simulation using RAVEN for the large auditorium scene (CR4), position LS2-MP2.

3.3. Potential improvements of the reflection density model

The analysis of the previous sections suggests that the reflection density model used in the filter synthesis needs to be adjusted, at least for the early part of the RIR. This could be done based on additional input parameters (e.g., geometrical complexity of surfaces or the scattering coefficients) or be based on the result of the ray tracing (temporal profile of detected particles).

The need for an adjusted reflection density model is supported by an evaluation of an image source calculation up to an order of 5 for a simple room model. In Figure 7 the number of audible image sources are visualized in a histogram (bin size: 1 ms) for three rooms with identical volume ($V=145$ m³), but different number of walls: 1) $n_w=6$; 2) $n_w=10$; 3) $n_w=32$. The number of audible image sources for the three rooms up to 30 ms are 75, 120 and 101; the total number of audible image sources (up to an order of 5) are 182, 274 and 292.

This example demonstrates that the (early) reflection density does depend on the room geometry, but cannot be described by a simple model, e.g., by the wall number, as for the chosen period of up to 30 ms, the number of audible image sources for the room model with 32 surfaces is even lower than for the

less detailed room model with only 10 surfaces. Thus, it cannot be stated that the number of audible reflections will increase in general if a model has a higher level-of-detail.

Additionally substantial deviations from the theoretical reflection density model (black curve in Figure 7) are observed in the early part of the RIR, even for the simple rectangular room shape.

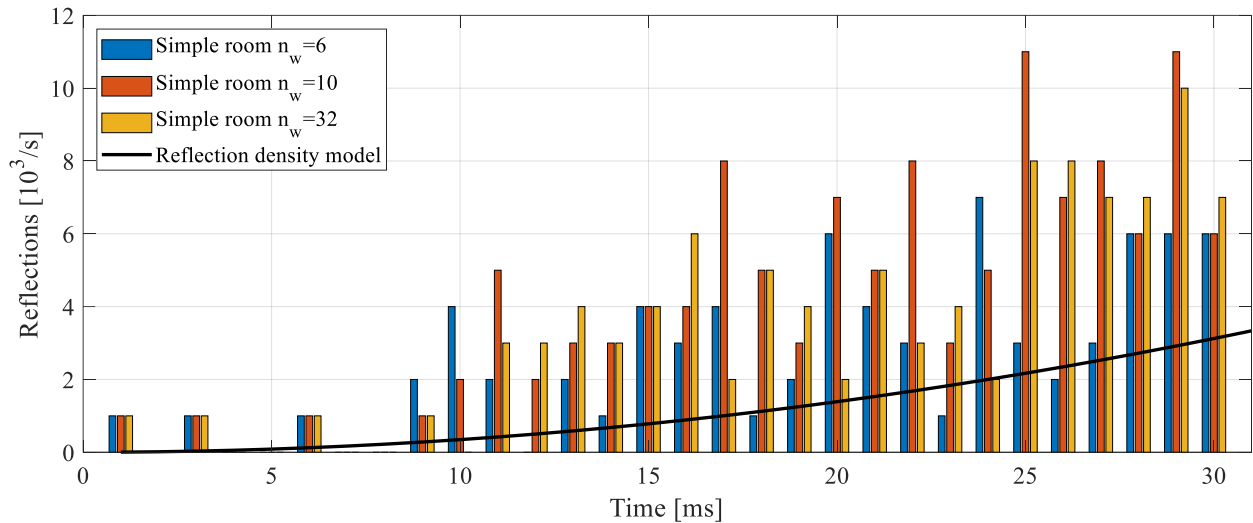


Figure 7: Reflection density evaluated based on the image source model (maximum order of 5) for three different rooms with identical volume (145 m^3). Audible image sources are added up in bins of 1ms length and compared to the reflection density model according to Eq. 1.

4. CONCLUSIONS AND OUTLOOK

When GA-based room acoustic simulation software is used for auralization, or is compared to corresponding measured results, a valid procedure for the synthesis of RIRs is required. This work presented possible approaches to synthesize a RIR based on image sources and an energy histogram and compared results of different GA-based simulation tools with corresponding measured results.

The comparison revealed that the temporal structure of simulated RIRs resembles the one of measured RIRs, but situations could be identified, where the simulated RIRs of all investigated software lack energy in the early part of the RIR, leading to relevant deviations of evaluated room parameters. This has been observed for rooms of larger volume, where strong reflections are sparse in the early phase after the arrival of direct sound, and in particular in situations with relatively small distances between the sound source and the receiver compared to the room dimensions. Thus, a more robust model for the reflection density in the early part of the RIR is required. Potential input parameters for such a model could be the geometry of the room, scattering coefficients or the temporal receiver hit distribution of a ray or beam tracing.

It also needs to be stated that not all GA-based simulation tools are using the described reflection density model. Especially if higher order image sources are used, assuming a time-invariant reflection density for the diffuse decay process, also in the early part, could be sufficient in many cases and would not lead to perceptual differences when applied in auralizations.

In general, the perceptual relevance of the discussed deviations needs to be assessed. While the clarity evaluation suggests that the differences are audible, a spectral comparison between simulated and measured RIRs revealed only minor, potentially inaudible, differences.

Further steps will involve data analysis of simulations (e.g., with respect to the temporal receiver hit distribution of the ray tracing) in order to develop and implement a modified reflection density model.

5. ACKNOWLEDGEMENTS

Part of this research was funded by the German Research Foundation (DFG Research Unit 1557 "SEACEN"). We also would like to thank all six participants of the Round Robin for providing their simulation results.

6. REFERENCES

1. Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., & Weinzierl, S. (2019). A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*, 145(4), 2746-2760. <https://doi.org/10.1121/1.5096178>
2. Aspöck, L. (2020). *Validation of room acoustic simulation models* (PhD Thesis, RWTH Aachen University, 2020). <https://doi.org/10.18154/RWTH-2020-12146>
3. Hamilton, B., & Bilbao, S. (2018, October). Wave-Based Room Acoustics Modelling: Recent Progress and Future Outlooks. In *Auditorium Acoustics 2018* (pp. 160-161). Institute of Acoustics.
4. Murphy, D. T., Southern, A., & Savioja, L. (2014). Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation. *Applied Acoustics*, 82, 6-14. <https://doi.org/10.1016/j.apacoust.2014.02.010>
5. Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4), 943-950. <https://doi.org/10.1121/1.382599>
6. Moorer, J. A. (1979). About this reverberation business. *Computer music journal*, 3(2), 13-28. <https://doi.org/10.2307/3680280>
7. Cremer, L. (1948). *Die wissenschaftlichen Grundlagen der Raumakustik. Band I, Geometrische Raumakustik*. Hirzel-Verlag Stuttgart, Germany.
8. Kuttruff, H. (2009). *Room acoustics*. Spon Press, New York, USA.
9. Vorländer, M. (1995). Revised relation between the sound power and the average sound pressure level in rooms and consequences for acoustic measurements. *Acta Acustica united with Acustica*, 81(4), 332-343.
10. Aspöck, L., & Vorländer, M. (2017). Synthesis of room impulse responses based on simulated energy decay curves. *Fortschritte der Akustik – DAGA 2017*, 275-278.
11. Schröder, D. (2011). *Physically based real-time auralization of interactive virtual environments* (Vol. 11). Logos Verlag Berlin GmbH.
12. Aspöck, L., Brinkmann, F., Ackermann, D., Weinzierl, S., & Vorländer, M. (2020). BRAS-Benchmark for Room Acoustical Simulation. <https://doi.org/10.14279/depositonce-6726.3> (data publication)
13. Aspöck, L. (2020). Input and output data for informed room acoustic simulations of the BRAS scene database. <https://doi.org/10.18154/RWTH-2020-09906> (data publication)