# Optimization of input parameters for real-time room acoustics simulations

D. Schröder

*Institute of Technical Acoustics, RWTH Aachen University, 52066 Aachen, Email: dsc@akustik.rwth-aachen.de*

## Introduction

Virtual Reality (VR) systems are today's state-of-the-art tools in many development and analysis processes, e.g., in the fields of architecture, engineering, psychology and medical science. The quality of a VR-system is assessable by the user's degree of immersion, which improves with the number of simulated coherent stimuli and level of interactivity. In this context, the visual perception is significantly augmented by matching sound stimuli, as auditory information helps to assign meaning to visual information, and vice versa. The user evaluates these events on attributes such as spaciousness and source localization, which enforces the feeling of actual presence especially in the case of indoor-scenarios.

The CAVE-like environment at RWTH Aachen University is equipped with a real-time 3D-sound rendering system that is currently being developed at the Institute of Technical Acoustics (ITA), RWTH Aachen University. The system enables a physical-based and fully interactive real-time auralization of enclosed spaces [1], usually rooms. Among other software modules, the system features the software RAVEN[1] that provides high quality room- and building acoustics auralization based on a hybrid simulation approach of geometrical acoustics. RAVEN uses a fast image source method to provide a precise temporal resolution of early specular reflections and a stochastic ray-tracing algorithm that accounts also for diffuse reflections [2]. More details on room- and building acoustics simulation methods can be found in [3].

This contribution presents a preliminary study, which aims to find the just noticeable difference (JND) of the simulation parameters "Image Source Order" and "Number of Ray-Tracing Particles" for the immersive simulation of a concert hall. Two listening tests were carried out where the first test featured only auditory stimuli, while the second test was performed with additional visual stimuli of a concert hall. Subjects had to evaluate situations with changing simulation parameterization, i.e., image source order and number of launched energy particles, to find a good balance between the accuracy of the room acoustics simulation and the simulation's computational costs.

[1] www.akustik.rwth-aachen.de/raven

## Listening Tests

As input data for the room geometry, a polygonal model of the concert hall at Eurogress, Aachen, was chosen. The model is built from 266 polygons and has a volume of about 15,000 m³. At first, an appropriate range for the investigated parameters had to be determined. Kuttruff pointed out in [4] that from reflections of order two or three, scattering becomes a dominant effect in the temporal development of the room impulse response. Thus, only low-order image sources (specular reflections) had to be taken into account and their range was set from order 0 to 3. For finding a reasonable range for the number of ray-tracing particles, two things had to be considered. On the one hand, it can be shown that for the given room 30,000 particles relate to a deviation of less than 1 dB in the envelope of the impulse response [3]. On the other hand, RAVEN features a method called "Diffuse Rain" for an improved detection of scattered energy [5]. Here, each scattered reflection triggers a secondary radiation of scattered energy from the hit wall to the receiver. Using this technique reduces the number of required ray-tracing particles, and thus computation time while keeping the statistical error of the decay constant. Unfortunately, no detailed information on its efficiency has been determined so far, but it is assumed to work very efficiently in rooms with many scattering surfaces such as the chosen concert hall model. The lowest number of particles per simulated frequency band was set to 100 to create distorted sounding stimuli that are easily detectable. 40,000 particles were chosen as reference.

| | Test Type | Task | Stimuli Comparison | Additional Information |
|---|---|---|---|---|
| **ITA** | 3 Alternative Forced Choice Test (3 cycles) | Find differing stimulus | Fixed image source order (up to order 3)<br>Variable number of particles (100, 1000, 5000, 10000, 20000, 40000) | 9 subjects (experts) (age range 25-31)<br>Dark room, eyes closed<br>Stimuli replay allowed |
| | Auditory stimuli | | | |
| **CAVE** | 3 Alternative Forced Choice Test (2 cycles/each) | Find differing stimulus | Variable/Fixed image source order (up to order 3)<br>Variable/Fixed number of particles (100, 1000, 5000, 10000, 20000, 40000) | 14 subjects (mostly experts) (age range 23-35)<br>Visual stimulus by Cave-System<br>Stimuli replay not allowed |
| | Auditory stimuli<br>Visual stimuli | | | |

**Tabular 1:** Details on the two performed listening tests.

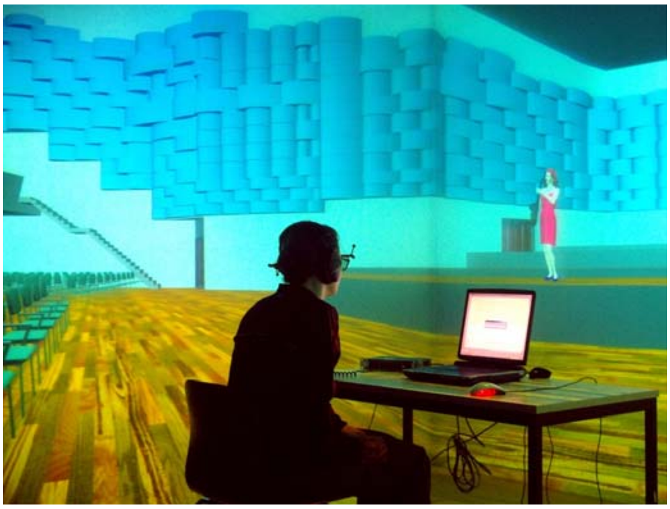


**Figure 1:** Listening test in a Cave-like environment.

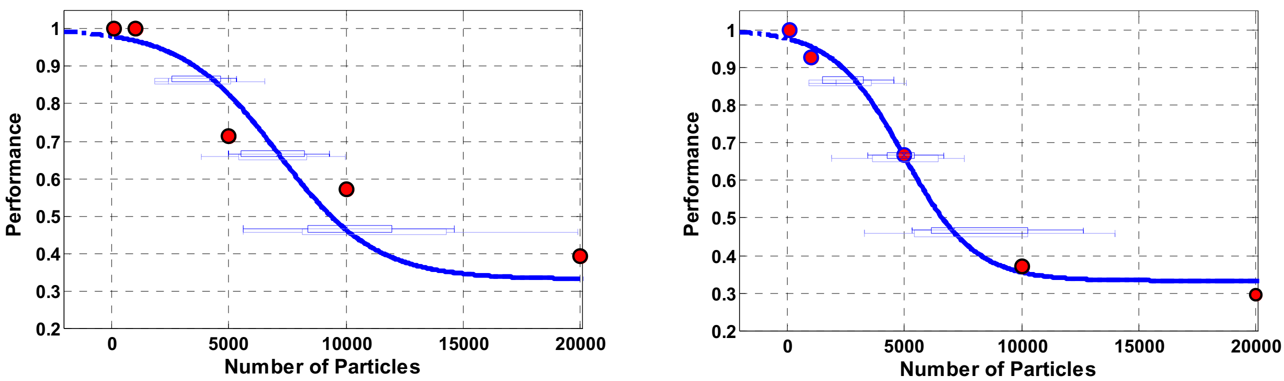
To not overstrain the subjects later on, the overall number of stimuli had to be limited to a moderate amount. Therefore, additional stimuli were created with 1,000, 5,000, 10,000 and 20,000 particles as input parameters for the ray-tracing. This resulted in the total number of 24 stimuli where the respective simulated impulse responses were convolved with a dry recording of a saxophone player. For the listening tests, the method of 3-Alternative-Forced-Choice-Test was applied as it avoids a response bias [6]. The subjects had to evaluate series of stimuli triples where each triple consists of two identical stimuli and the differing stimulus had to be identified. A GUI, that was displayed on a laptop, guided the subjects through the tests. The stimuli were played back using a Sennheiser headphone and a RME Hammerfall DSP. The first test was carried out in a darkened room and subjects had to additionally close their eyes to secure that the hearing is stimulated solely. The second test was performed in the Cave-like environment at RWTH Aachen University. Here, additional stereoscopic visual stimuli that represented the simulated scene were projected from outside onto the five screens. The subjects were asked to look at the saxophone player during the test (see Figure 1). More details on the two listening tests are given in Table 1.

## Evaluation

The test results were analyzed using the Matlab toolbox `psignifit`[2] by Wichmann and Hill [7, 8] which is a free multi-platform software package that performs maximum-likelihood fitting and significance testing for psychometric functions. Figure 2 shows exemplarily the calculated functions for the test series with/without additional visual stimulus where the image source order was set to 2 and the number of particles was varied. The red dots show the averaged performance of the subjects in respect to the stimulus. Additionally the psychometric functions are plotted (computed by a maximum likelihood estimation after Wichmann and Hill) as partially solid and dashed line. The 68.2% and 95.4% confidence intervals are given in the horizontal bars at the 0.2, 0.5 and 0.8 level of the respective psychometric functions.

Most noticeable, all functions showed similar behavior when comparing the related results of the two listening tests with and without visual stimuli. All JNDs were significantly shifted to a lower number of particles (an average decrease of about 2,000 particle was observed) if the additional visual stimuli were presented (compare Figure 2a,b). The JNDs lay always in the range between 5,000 and 8,000 particles, which states that the image source order had only minor influence on these test results and that the "Diffuse Rain" method seems to work quite efficiently in this type of room. Therefore, 10,000 particles were regarded as appropriate for the immersive simulation of the chosen concert hall. As for the image source order, an additional test series was carried out during the second listening test in the CAVE-like environment. Here, the number of particles was fixed and the image source order ranged between 0 and 3.

[2] www.bootstrap-software.org




**(a) auditory stimuli** **(b) auditory and visual stimuli**

**Figure 2:** Psychometric functions fitting the test results of the test series with fixed image source order 2 and variable number of ray-tracing particles. A performance of 1 means that the subjects have always found the differing stimulus, while a performance of 0.33 shows that the subjects have always guessed. The JND relates to a performance of 0.66.

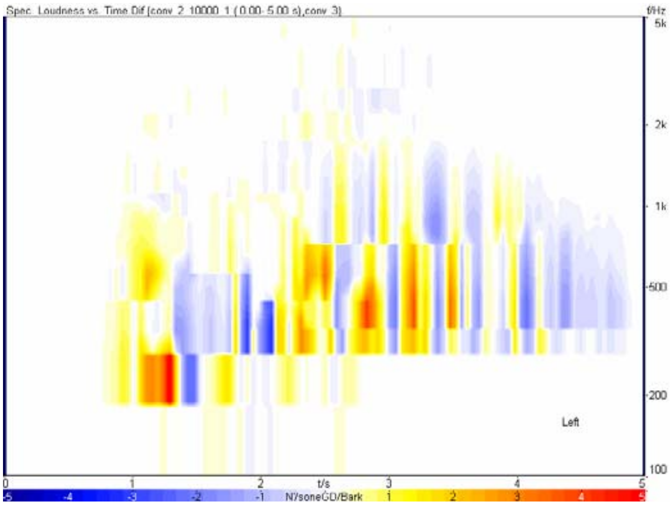



**Figure 3:** Difference plot of stimuli loudness over time (image source order 2 vs. 3, equal loudness of the stimuli were about 40 sone). The number of particles was set to 10,000.

However, the test results did not show any useful information as all subjects mostly guessed the answers. It is assumable that the subjects lacked concentration during the test as it was performed right after the first test series with fixed image source order and variable number of particles due to certain time constraints for using the Cave-like environment. However, this indicates that the stimuli were very similar, which was also confirmed by the subjects that had to fill out a questionnaire right after the listening tests. To further analyze this, difference plots of the stimuli loudness over time were determined (see Figure 3). The equal loudness of the two compared stimuli was about 40 sone and the difference between them (second order vs. third order image sources) was only +/- 4 sone and in a direct comparison no subject was able to differ between the two sounds. Summing it all up, the subjects were unable to distinguish between stimuli that were simulated with an image source order greater than 2 and a number of particles greater than 10,000. This relates to a workload of 7.45 ms for testing the image sources on validity and 1.7s for the ray-tracing (10 frequency bands and 10,000 particles per band) on a standard personal computer (Intel Core 2 Quad Q6600/2.4 GHz). Thus, real-time capability is still maintained (more details on applied update strategies are given in [1]).

## Outlook

This contribution presented only a preliminary study and investigations that are more detailed will be published in the future. Tests will be carried out with more subjects, different room types (reverberant/dry sound field, different shapes) and refined listening tests. Furthermore, the performance of the "Diffuse Rain"- method will be further examined to figure out their performance within different room types. In a subsequent step, all these tests will be carried out under stress, e.g., search operations, where a further drop of the found JNDs is expected.

## Acknowledgments

The author would like to thank the German Research Foundation (DFG) for funding this project. Furthermore, thanks to Michael Vorländer, Sebastian Fingerhuth and Ingo Witew for valuable discussions.

## References

[1] Lentz, T., Schröder, D., Vorländer, M., Assenmacher, I., *Virtual Reality System with Integrated Sound Field Simulation and Reproduction*, EURASIP Journal on Advances in Signal Processing , 2007

[2] Schröder, D., Vorländer, M., *Hybrid method for room acoustic simulation in real-time*, 19th International Congress on Acoustics (ICA), Madrid, Spain, 2007

[3] Vorländer, M., *Auralization*, Springer Berlin, 2007

[4] Kuttruff, H., *A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries*, The Journal of the Acoustical Society of America, 98(1), 288–293, 1995

[5] Schröder, D., Dross, P., Vorländer, M., *A Fast Reverberation Estimator for Virtual Environments*, AES 30TH International Conference, Saariselkä, Finland, 2007

[6] Gescheider, G. A., *Psychophysics- The Fundamentals*, Lawrence Erlbaum Associates Inc, Mahwah, New Jersey,1997

[7] Wichmann, F. A. & Hill, N. J., *The psychometric function: I. Fitting, sampling and goodness-of-fit.* Perception and Psychophysics 63(8), 1293-1313, 2001

[8] Wichmann, F. A. & Hill, N. J., *The psychometric function: II. Bootstrap-based confidence intervals and sampling*. Perception and Psychophysics 63(8), 1314-1329, 2001