Forum Acusticum 2011 - Aalborg - Denmark

RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments

Dirk Schröder and Michael Vorländer
Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany.

Summary
In analogy to visualization, the auralization of virtual environments describes the simulation of sound propagation inside enclosures, where methods of Geometrical Acoustics are mostly applied for a high quality synthesis of aural stimuli that go along with a certain realistic behavior. In this contribution, the design and implementation of the real-time room acoustics simulation framework RAVEN will be briefly described, which is today a vital part of the implemented 3D sound rendering system of RWTH Aachen University's immersive Virtual Reality system. RAVEN relies on present-day knowledge of room acoustics simulation techniques and enables a physically quite accurate auralization of sound propagation in complex environments, including important wave effects such as sound scattering, airborne sound insulation between rooms and sound diffraction. In spite of this realistic sound field rendering, not only spatially distributed and freely movable sound sources and receivers are supported at runtime, but also modifications and manipulations of the environment itself. All major features will be evaluated by investigating both, the overall accuracy of the room acoustics simulation and the performance of implemented algorithms.

PACS no. 43.20.Dk, 43.55.Ka

1. Introduction

In recent years, the development of room acoustics prediction tools and auralization techniques has made a major leap forward enabling a physically based simulation of virtual environments in real-time. The speed-up is usually achieved by adapting acceleration techniques from Computer Graphics (CG) since most simulation methods are based on the basic principles of Geometrical Acoustics (GA). In GA, the sound field is decomposed into specular reflections and scattering components, which perfectly matches simulation methods known from CG. However, the frequency range in acoustics involves three orders of magnitude (20 Hz to 20 kHz and wavelengths from about 20 m to 2 cm), where neither approximations of small wavelengths nor large wavelengths can be assumed with general validity. Therefore, it is of major importance that wave phenomena, such as diffraction at low frequencies, scattering at high frequencies, and specular reflections are considered by the applied simulation method to gain a physically based sound field modeling. While today low-reflective outdoor scenarios can be simulated rather fast and accurate, the determination of the spatial sound field in enclosures, i.e. rooms, is still a difficult task. Especially complete dynamic environments become quite challenging, where the user is not only allowed to move freely in all three dimensions and interact with sound sources in any order, but also to modify the scene geometry itself and directly perceive the impact on the sound field in real-time. Such changes significantly affect the whole auralization chain and, therefore, approximations must be made to keep the computational workload also in such demanding situations within the given real-time constraints. Fortunately, the resulting sound does not have to be physically absolutely correct for auralization, but perceptively plausible.

In this contribution, the current state of the real-time room acoustics simulation software RAVEN (Room Acoustics for Virtual ENvironments) will be briefly described. RAVEN is being developed at the Institute of Technical Acoustics, RWTH Aachen University, and is today a vital part of the implemented 3D sound rendering system of RWTH Aachen University's immersive Virtual Reality system [1, 2]. Detailed information on RAVEN's simulation concept and implementation are given in the first author's doctoral thesis [3].

2. The RAVEN Framework

The hybrid room acoustics simulation RAVEN combines a deterministic Image Source (IS)-method [4]

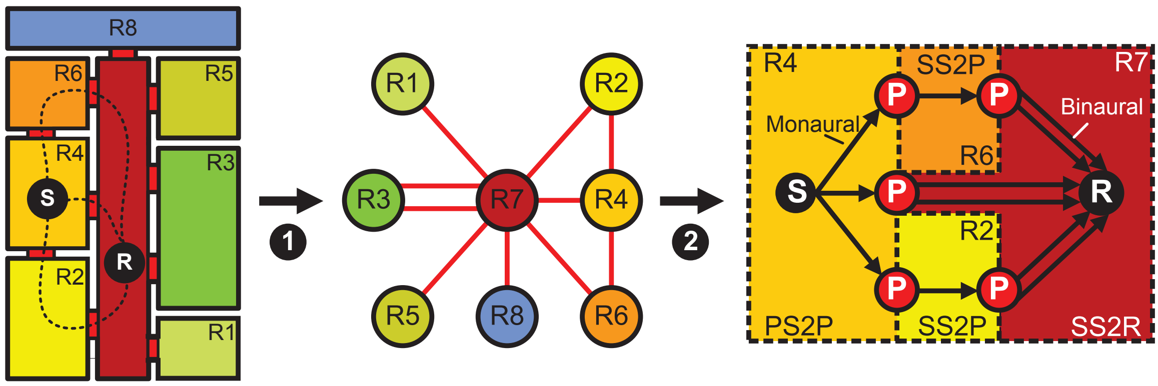


Figure 1. Example of tracking sound propagation paths throughout an office floor and constructing a corresponding filter network. The floor consists of eight rooms, where all doors, i.e. portals, are closed resulting in eight room groups (the office floor is shown on the right-hand side). A primary source S is located in room $R4$, while the receiver R resides in room $R7$. At first, all relevant propagation paths are extracted from the ASG (shown in the middle) and encoded in a directed acyclic graph (shown on the left-hand side of the figure).

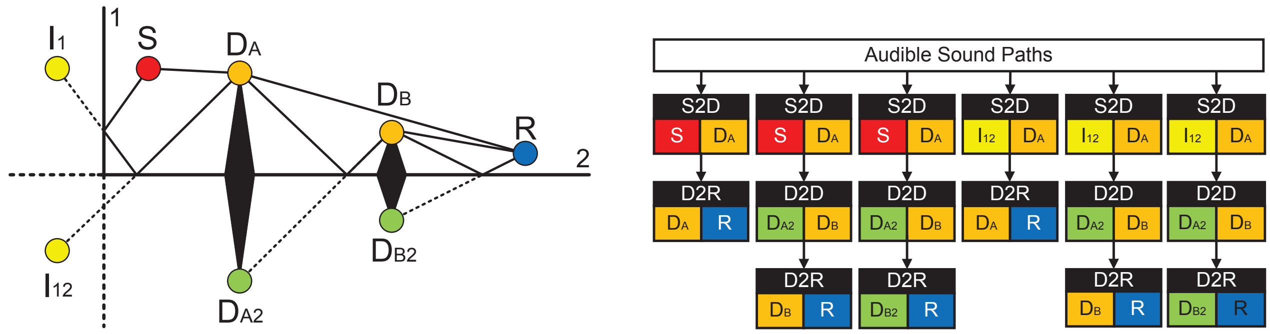
with a stochastic Ray-Tracing (RT) algorithm [5]. The framework enables a physically based computation of high quality room impulse responses in real-time, where – apart from the specularly reflected sound field components – the sound phenomena of sound scattering, sound transmission and sound diffraction are taken into account, too. In addition, RAVEN imposes no constraints on scene interaction meaning that not only sound sources and receivers are allowed to move freely in the virtual environment, but also the scene geometry can be manipulated by the user at runtime. The framework is completely written in C++, supports all major operating systems such as Windows, Linux and Mac OS X, and enables efficient parallel computing on local shared memory machines, distributed memory machines via network, or combinations of both.

2.1. Sound Transmission

In contrast to simple one-room situations, the real-time auralization of complex environments requires a very fast scene handling and interaction management. To overcome the complexity of such huge geometric models, RAVEN uses a logical and spatial subdivision of the virtual environment that enables a dynamic decomposition into acoustically separated subspaces. Here, the topological structure of the scene is represented by a graph, which is a general data structure for the logical concatenation of entities, called nodes. Each node stores the spatial representation of a single room, including the polyhedral model (encoded in special spatial data structures for fast intersection tests), material data, temperature, humidity and air pressure. The connectivity between two nodes is steered by the state of the respective portal (opened/closed), which can be seen as a logical and physical room separator, e.g., a wall, a door or a window. In the following, this graph will be referred to as Acoustical Scene Graph (ASG), which is further illustrated in Figure 1 by the example of an office floor. Here, the floor plan is given on the figure's left-hand side depicting a scene, which contains eight rooms that are interconnected by ten doors, i.e. portals. All room information is then encoded into the respective ASG (Figure 1, step 1). In this exemplary scene configuration, all portals are closed (denoted through red edges in the ASG) leading to a maximum subdivision of the virtual environment into eight single rooms representing the currently coupled acoustic spaces.

For the simulation of sound transmission, the building acoustics auralization concept by Thaden et al. [6] was adapted and further extended to a portal-related secondary source model. In Thaden's model, sound transmission is described by sound paths from a primary source (PS)/secondary source (SS) to a receiver via air (indoor) and structural elements. The performance of each structural element is described by a transfer function, which is built from interpolated transmission coefficients of the corresponding coupling element (these coefficients are computed according to standardized simulation models, e.g., the European Standard EN 12354). However, since the complete simulation of *all* transmitting parts between coupled rooms is infeasible under real-time constraints, sound transmission paths are usually reduced to a small number of directly coupling joints with a low sound level difference, such as doors and windows, as they dominate the overall level of transmitted sound. To incorporate this secondary source model into the overall simulation management, the portal model is extended from a simple logical and physical separator of the ASG to a coupling element with integrated sender, receiver and corresponding transfer function. Here, the portal's sender is always modeled as a point source, i.e., a SS that is located at the center of the portal's surface, while the hybrid implementation requires two different types of receivers: A

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767



(a) Sound paths between the primary source S and the receiver R. ISs and DSs are denoted by I and D, respectively. Some sound paths are omitted for the sake of a clear arrangement.

(b) Construction plan for impulse responses extracted from edge visibility information. Each container denotes one sound path and the respective sound subpaths are described by their type (S2D: Source (PS/IS) to edge (DS/DIS), D2R: Edge (DS/DIS) to receiver, and D2D: Edge (DS/DIS) to edge (DS/DIS)).

Figure 2. Extension of the traditional IS-model for edge diffraction.

point receiver[1] in the case of the IS-method and a surface receiver in the case of stochastic RT, where the portal's surface is used as detector area. Each sound propagation path from a PS/SS through air to a receiver/portal requires a room acoustical simulation for computing the respective impulse responses, while the performance of the portal itself is described by the corresponding transfer function as described above.

Considering this simulation model, four different types of sound propagation paths can occur: (1) PS2R: PS to receiver, (2) PS2P: PS to portal, (3) SS2P: SS to portal, and (4) SS2R: SS to receiver. Cases 1 and 4 state the sound propagation to a binaural receiver, where the sound is emitted from either a PS or a SS in the corresponding coupled room group. Since sources can be spatially localized here, binaural simulations of the respective sound paths are required for a good spatial impression[2]. In contrast, case 2 and 3 resemble the sound propagation from a PS and SS, respectively, to a portal. These sources reside outside the receiver room group, i.e., they are located behind closed portals, which makes a spatial localization of the source rather impossible for the receiver. Hence, monaural simulations are sufficient for both path types. With the given scene configuration shown in Figure 1, all relevant sound propagation paths from a PS to a receiver are determined by unrolling the ASG in a depth-first manner and encoding information on audible sound propagation paths for each PS/receiver-pair in a directed acyclic graph (see Figure 1, compare left-hand and right-hand side). The decision on sound path audibility is drawn by summing up the sound reduction indices of all involved portals and comparing the result with a given energy threshold, usually set to the sound level of the investigated primary source. If the threshold is exceeded, the path is skipped since it is not audible anymore for the receiver. The final graph then comprises both, information on required simulation steps and a construction plan for the corresponding filter network that represents the scene's total transmission characteristics from a PS to a receiver via several structural elements. For the latter, all edges of the directed graph are substituted by their respective room impulse responses that come from room acoustical simulations, while portal nodes are replaced by their corresponding portal filter functions. More details are given in [3].

2.2. Sound Diffraction

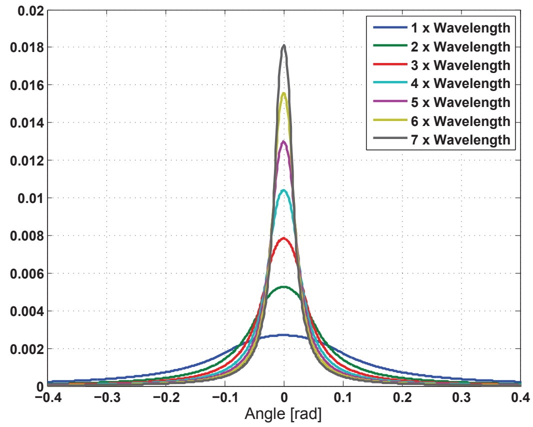
As mentioned above, RAVEN's hybrid simulation approach also accounts for the wave phenomenon of diffraction by adapting and partly expanding state-of-the-art GA simulation methods of edge diffraction for both, the IS-method and RT. Since the traditional IS approach neglects energy that is actually bent into shadow zones caused by obstacles, the implemented IS-method is extended by another type of SS, called Diffraction Source (DS) in the following. These spatially extended DSs are located on top of each edge (see Figure 2(a), D_A and D_B) and set active if the direct path between a source and a receiver is blocked[3]. Similar to the generation of ISs for a PS, each edge along with its respective DS is mirrored on the room's surfaces up to a predefined reflection order. These mirrored edges are called diffraction image

[1] The point receiver is also located at the center of the portal's surface.

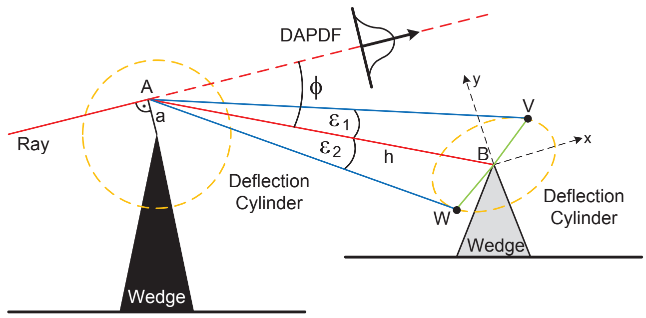
[2] Receivers are always modeled binaurally since they represent the real-world user.

[3] This is certainly not exact since diffraction also occurs in the view zone, but it can be assumed that the contribution of the diffracted waves in the view zone is quite small compared to the primary wave and, thus, can be neglected in order to reduce the overall simulation complexity at runtime.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767



(a) DAPDF as a function of deflection angle and fly-by distance (in wavelengths).



(b) *Diffracted rain* method for an edge-to-edge situation using the 2D-DAPDF. The outgoing energy is dispersed on a plane which intersects with a second DC.

Figure 3. Extension of the RT algorithm for edge diffraction.

source (DIS) and allow the additional computation of specular sound reflections in between the wedges (see Figure 2(a), D_{A2}, and D_{B2}). For the fast computation of audible sound subpaths across and in between (w)edges, visibility information for both, sources and edges are computed and stored in tree-like data structures that allow a fast access and update during runtime. Similar to the extraction of audible sound paths in the case of sound transmission (see above), a special search algorithm is applied on these visibility trees that returns logical and geometrical information of each audible sound path from a primary sound source to a receiver across the involved edges. These visibility trees are used in a subsequent step to construct the corresponding filters (see Figure 2(b)). Using this decomposition scheme, the transfer function of each sound path can efficiently be computed in frequency domain by multiplying the transfer function of the PS's directional pattern H_{PS} with the transfer functions of each sound subpath H_n and the receiver H_R. The total impulse response h_{sum} is then obtained by summing up the respective impulse responses of the respective subpaths in time domain. While the computation of impulse responses goes straight forward for subpaths where only PSs/ISs and the receiver are involved, more complex computations are required for sound paths that include also edges. These edge diffraction impulse responses are determined by using the concept by Svensson et al. [7], which is based on the exact and well-known Biot-Tolstoy solution. In this extended method analytical directivity functions are derived for secondary edge sources (not to be mixed up with the DS/DIS that were introduced above) that give the exact solution not only for infinite wedges, but also for finite ones - at least for first-order diffraction. Very good results are also achieved for higher-order diffraction though.

To extend the stochastic RT method for edge diffraction, a new type of particle detector is used, so-called deflection cylinders (DCs). These detectors are of cylindrical shape, where the cylinder's axis is identical with the respective edge and the cylinder's radius depends on the regarded frequency with $r = 7\lambda$. If such a DC is hit by a energy particle, the edge's influence on the particle can be described by a special function that was proposed by Stephenson [8]. The basic idea of Stephenson's uncertainty-based diffraction method is to deflect energy particles around an edge as a function of the respective shortest fly-by distance, i.e., the closer the particle passes the edge, the more it gets diffracted into the wedge's shadow zone. This is expressed by a 2D-deflection-angle-probability-density-function (DAPDF) (see Figure 3(a)), which can be seen as the simplified spatial Fourier transform of a slit's transfer function. In Stephenson's traditional interpretation, each impacting particle energy is logged and further distributed among a large number of outgoing secondary energy particles in order to keep stochastic fluctuations at a minimum level. To overcome this complexity, a selective transfer of sound energy to detectors is used, called *diffracted rain* (see Figure 3(b)). Here, the energy portion of a certain angle field $\Delta\epsilon$ is calculated by integrating the 2D-DAPDF over the respective angle range. This energy is dispersed on a 2D-plane in a recursive way meaning that any DC can forward the received energy to other visible detectors (detection spheres, portals and DCs) that intersect with the plane in direction of sound propagation. In a subsequent step, the detected energy particle is deflected around the initially hit edge and further traced, whereby direct hits of detectors are skipped that were already covered by the diffracted rain. A complete description of RAVEN's edge diffraction concept is published in [3].

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

2.3. Geometry Manipulation

In order to maximize the user's interaction potential, another important design aspect is the creation of highly flexible algorithmic interfaces that uphold a real-time auralization even in quite demanding interaction events. While code adjustments for operations such as the exchange of material parameters and the manipulation of portal states are relatively easy to implement, the requirements of a modifiable geometry is quite an algorithmic challenge since it affects the whole auralization chain. First test implementations showed that RAVEN's acceleration algorithms based on Binary Space Partitioning (BSP) [4] do not meet the criteria of dynamically manipulable geometry, because any modification of this hierarchical approach calls for a recalculation of at least large parts of the respective BSP trees leading to significant lag. In order to solve this issue and maintain a fast system responsiveness, the auralization framework was further expanded by introducing two types of scene geometry: Static and dynamic objects. Static objects, such as walls, are not modifiable during the simulation and are therefore processable in a quick and efficient way. In contrast, dynamic objects are adjustable by the user at runtime (an example is given in Figure 4). However, it should be emphasized that there is no general limitation on the object's shape meaning that the whole room geometry can be defined as dynamic as long as the available computing power is sufficient.

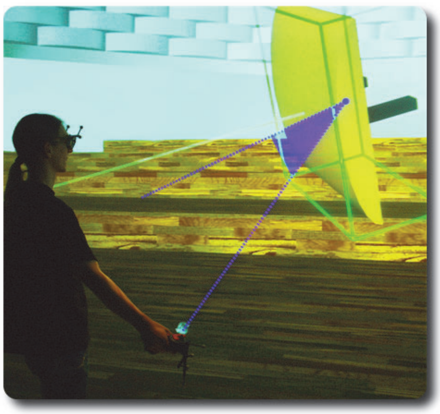


Figure 4. User working in an immersive virtual environment by the example of the real-time adjustment of a reflector panel.

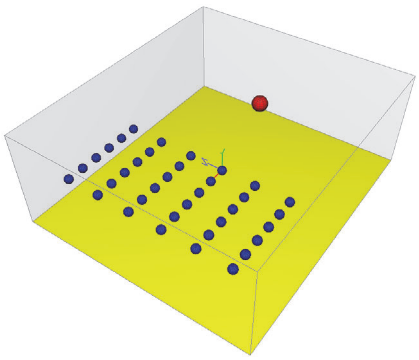
For the unconditioned modification of dynamic objects, RAVEN switches to another approach of geometry processing, called Spatial Hashing (SH). SH is a method from CG, which is usually applied for collision tests with deformable objects. The concept of SH is based on a spatial subdivision by primitive volumes (voxels), where the infinite voxelized space is mapped to a finite set of one-dimensional hash indices using simple hash functions. The advantage of SH over other hierarchical spatial data structures such as BSP-trees is that the insertion/deletion of m vertices takes only O(m) time, i.e., these operations comes almost for free. On the other hand, a comprehensive performance analysis showed that BSP-accelerated intersection tests were much more efficient than methods based on SH. However, they also evidenced the non-negligible lag that comes from hierarchical data structures due to their geometrical dependency causing discontinuities of the rendered sound field [3]. Thus, SH is meaningful only during the actual modification events to enable fast updates of the sound field, but as soon as these events are finished, the framework regenerates all hierarchical data structures and switches back to fast BSP-tree acceleration.

While the concept of SH is easily embeddable to the applied stochastic RT algorithm - where it is sufficient to update just both the geometry and the corresponding spatial data structures - a dynamic handling of ISs is much more complex as ISs have to be generated, destroyed and updated (audibility and position) at runtime. For this purpose, a hierarchical tree data structure was introduced that efficiently organizes ISs for a convenient processing. A detailed description of the whole concept for real-time geometry modification is given in [9, 3].

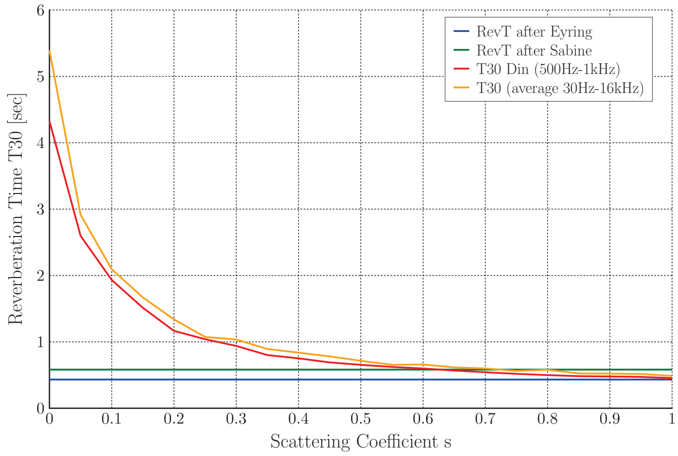
3. Validation

Due to limited space, a detailed validation of RAVEN cannot be given in the course of this contribution. Therefore, only one test scenario concerning the hybrid simulation approach itself will be exemplarily shown, whereas numerous others can be found in, e.g., [5, 3]. In this comparison a simple shoebox-shaped room (volume of $1344\,m^3$) was chosen with a calculated reverberation time of $0.43\,s$ after Eyring and $0.58\,s$ after Sabine, respectively[4]. The room's side wall materials were parameterized with an absorption factor of 0.05, while a factor of 0.8 was assigned to both, the floor and the ceiling materials. In contrast, the same scattering coefficient was used for both materials and varied in steps of 0.05 during the simulations. A sound source was positioned in one half of the room and 36 in-plane receivers were evenly distributed outside the source's critical distance in the opposite half of the room (see Figure 5(a)). Room acoustical simulations were carried out by taking into account ISs up to order 4, performing RT simulations with 25000 particles per frequency bands (octave band resolution), and averaging the results of all 36 receivers (more details are given in Figure 5(b)).

[4] As commonly known, the equation after Eyring is more accurate since it is not based on the assumption of low-absorbing surfaces as in Sabine's approach.



(a) Test model with one sound source (red sphere) and 36 receivers(blue spheres).



(b) Reverberation Time T30 in comparison with the scattering coefficient s.

Figure 5. Validation of RAVEN's hybrid room acoustics simulation method.

A comparison of the simulation results is shown in Figure 5(b). Here, reverberation times (T30) were either calculated after DIN EN ISO 3382, i.e. for the center frequencies 500 Hz and 1 kHz (red line), or by evaluating all center frequencies between 30 Hz and 16 kHz (orange line). In addition, the values of reverberation times are given that were calculated after Sabine (green line) and Eyring (blue line). When taking a closer look, one can see that the simulated reverberation times show the typical exponentially decreasing behavior with increasing diffuseness of the sound field, where simulated reverberation times start to match the prediction after Sabine for scattering coefficients greater than 0.6. For a scattering coefficient of exactly one (ideal diffuse sound field) the evaluation after DIN EN ISO 3382 exactly matches the reverberation time after Eyring, while the value derived from all 10 octave-band just missed that point by a very narrow margin. From this it follows that the transition from pure specular reflections to a diffuse sound field is covered correctly by the applied hybrid simulation.

4. Conclusion

A unique auralization framework was briefly described that relies on present-day knowledge of room acoustical simulation techniques and enables – among many other features – a physically accurate auralization of sound propagation in complex environments, including the important wave effects of sound scattering, sound transmission and sound diffraction. In spite of this realistic sound field rendering, not only spatially distributed and freely movable sound sources and receivers are supported at runtime, but also modifications and manipulations of the geometrical environment itself. A complete description can be found in the first author's doctoral thesis [3].

Acknowledgement

The authors would like to thank the German Research Foundation (DFG) for funding this project.

References

[1] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmacher, "Virtual reality system with integrated sound field simulation and reproduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, p. 19, 2007.

[2] D. Schröder, F. Wefers, S. Pelzer, D. Rausch, M. Vorländer, and T. Kuhlen, "Virtual Reality System at RWTH Aachen University," in *Proceedings of the International Symposium on Room Acoustics (ISRA)*, (Melbourne, Australia), 2010.

[3] D. Schröder, *Physically Base Real-Time Auralization of Interactive Virtual Environments*. PhD thesis, RWTH Aachen University, 2011.

[4] D. Schröder and T. Lentz, "Real-time processing of image sources using binary space partitioning," *Journal of the Audio Engineering Society*, vol. 54, no. 7/8, pp. 604–619, 2006.

[5] D. Schröder, P. Dross, and M. Vorländer, "A fast reverberation estimator for virtual environments," in *Proceedings of the 30th AES International Conference*, (Saariselkä, Finland), 2007.

[6] R. Thaden, *Auralisation in Building Acoustics*. PhD thesis, RWTH Aachen University, Aachen, Germany, 2005.

[7] U. P. Svensson, R. I. Fred, and J. Vanderkooy, "An analytic secondary source model of edge diffraction impulse responses," *Journal of the Acoustical Society of America*, vol. 106, p. 2331, 1999.

[8] U. M. Stephenson, "Simulation of diffraction within ray tracing," *Acta Acustica United with Acustica*, vol. 96, p. 516, 2010.

[9] D. Schröder and A. Ryba, "Real-time auralization of dynamically changing virtual environments," in *Proceedings of the 20th International Congress on Acoustics (ICA)*, (Sydney, Australia), 2010.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767