Proceedings of the International Symposium on Room Acoustics, ISRA 2010

29-31 August 2010, Melbourne, Australia

Virtual Reality System at RWTH Aachen University

Dirk Schröder (1), Frank Wefers (1), Sönke Pelzer (1), Dominik Rausch (2), Michael Vorländer (1), Torsten Kuhlen (2)

(1) Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany
(2) Virtual Reality Group, RWTH Aachen University, Aachen, Germany

PACS: 43.55.KA

ABSTRACT

During the last decade, Virtual Reality (VR) systems have progressed from primary laboratory experiments into serious and valuable tools. Thereby, the amount of useful applications has grown in a large scale, covering conventional use, e.g., in science, design, medicine and engineering, as well as more visionary applications such as creating virtual spaces that aim to act real. However, the high capabilities of today's virtual reality systems are mostly limited to first-class visual rendering, which directly disqualifies them for immersive applications. For general application, though, VR-systems should feature more than one modality in order to boost its range of applications. The CAVE-like immersive environment that is run at RWTH Aachen University comprises state-of-the-art visualization and auralization with almost no constraints on user interaction. In this article a summary of the concept, the features and the performance of our VR-system is given. The system features a 3D sketching interface that allows controlling the application in a very natural way by simple gestures. The sound rendering engine relies on present-day knowledge of Virtual Acoustics and enables a physically accurate simulation 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, but also modifications and manipulations of the environment itself. The auralization concept is founded on pure FIR filtering which is realized by highly parallelized non-uniformly partitioned convolutions. A dynamic crosstalk cancellation system performs the sound reproduction that delivers binaural signals to the user without the need of headphones. The significant computational complexity is handled by distributed computation on PC-clusters that drive the simulation in real-time even for huge audio-visual scenarios.

INTRODUCTION

Room acoustic auralization was developed from simulation algorithms and binaural technology in a historic process of more than 20 years. Full-immersive Virtual Reality (VR) systems, such as CAVE-like environments, have been in use for more than 15 years. While computer graphics and video rendering are far developed with applications in film industry and computer games, high-quality audio rendering is still not on a comparable level. Watching a recent 3D movie production and comparing the quality of the visual and auditory representation can best illustrate this mismatch. While advanced projection systems deliver a good 3D vision, the auditory 3D impression lacks a realistic spatial sound impression, although rather complex surround sound systems are usually installed.

Real-time auralization systems have been investigated by many groups. Here to mention is the 'EVE'-project at the TKK Helsinki University [1], Finland, and the 'REVES' research project [2] at INRIA, France. Lately, Open Source projects were launched such as "UNIVERSE" [3] or the software framework by Noisternig et al. [4]. The aim of our VR activities is to create a reference platform for development and application of multimodal environments including high-quality acoustics. Such a system can be used for scientific research and testing as well as for development of complexity-reduced surround sound systems for professional audio or home entertainment. The group working on the acoustic VR-system is supported by the German Research Foundation, DFG, in a series of funded projects, where the Institute of Technical Acoustics, ITA, jointly worked with the Virtual Reality Group of RWTH Aachen University. The latter is the core group of a consortium of several institutions of our university and external partners covering the disciplines of computer science, archi-tecture, civil engineering, mechanical engineering, electrical engineering and information technology, psychology and medicine [5].

HISTORY OF ACOUSTIC VIRTUAL REALITY

VR Technology

In the early days of VR, head-mounted displays (HMDs) usually formed the heart of any VR-system in order to provide stereoscopic vision to the user. A HMD is a helmet-like display that features two small monitors positioned directly in front of the user's eyes to achieve stereopsis. Typically, a HMD is also equipped with earphones, where binaural synthesis has most often been used for the presentation of acoustic stimuli in the virtual 3-D space. However, due to fundamental problems such as wear comfort and user isolation from the real environment, today's HMDs are mainly used in low cost and mobile/portable VR-systems. Instead, especially in scientific and industrial applications, HMDs have been more and more replaced by CAVE-like displays [6]. These displays are room-mounted installations based on a combination of large projection screens that surround the user. Here,

(a) Exploration of a concert hall. (b) Online insertion and modification of a reflector panel.

Figure 1: The immersive environment at RWTH Aachen University.

the stereoscopic vision is realized by means of light-weight polarized glasses that separate visual information from the stereoscopic projection. Since room-mounted VR-systems aim at an ergonomic, non-intrusive interaction as well as co-located communication and collaboration between users, headphones should be avoided. As such, a sound reproduction based on loudspeakers is preferable. Head-mounted as well as room-mounted VR displays typically come along with a tracking system that captures the user's head position and orientation in real-time. This data is required for adapting the perspective of the visual scene to the user's current viewpoint and view direction. In addition, if binaural synthesis is applied for auralization, the user's position and orientation must be precisely known at any time in order to apply the correct pair of Head Related Transfer Functions (HRTFs). A variety of tracking principles for VR-systems are in use, ranging from mechanical, acoustic (ultrasonic), and electromagnetic techniques up to inertia and opto-electronical systems. Recently, a combination of two or more infrared (IR) cameras together with IR light reflecting markers that are attached to the stereo glasses, have become the most popular tracking systems. This type of tracking system is nearly non-intrusive, affordable and works with higher precision and lower latency in comparison to other technologies.

In recent years, VR has proven its potential to provide an innovative human computer interface for applications areas such as architecture, product development, simulation science, or medicine. VR is characterized as a computer-generated scenario of objects. A user can interact with these objects in all three dimensions in real-time. Furthermore, multiple senses should be included to the interaction, i.e., besides the visual sense, the integration of other senses such as the auditory, the haptic/tactile, and the olfactory stimuli should be considered in order to achieve a more natural, intuitive interaction with the virtual world.

Room acoustics simulation

At RWTH Aachen University, room acoustics simulation is in the focus since the mid 1980's, initially based on ray tracing and image source algorithms and later on combinations of both approaches. Vorländer [7] and Vian et al. [8] presented the basis for the cone, beam and pyramid tracing dialects, e.g. [9, 10, 11, 12], by showing that forward tracing is a very efficient method for finding audible image sources. Since then, the specular components of the room impulse response (RIR) were computable with high efficiency. The concepts of spatial subdivision were added for a quick processing of intersection tests which is the crucial subroutine in methods of Geometric Acoustics (GA). Then, during the 1990's it was shown that GA cannot solely be based on specular reflections [13]. The era of the implementation of scattering began with activities on the prediction, measurement and standardization of scattering and diffusion coefficients of corrugated surfaces [14].

Progress in binaural technology [15] enabled the incorporation of spatial attributes to room impulse responses. The key equation of the contribution of one room reflection, H_j, is given in spectral domain [16] with

$$H_j = H_{Source}(\vartheta, \varphi) \cdot \frac{\left(e^{-j\omega t_j}\right)}{t_j} \cdot H_{air} \cdot H_{HRTF}(\Theta, \Phi) \cdot \prod_{i=1}^{n} R_i$$

where t_j is the reflection's delay, $j\omega t_j$ the phase, $1/t_j$ the distance law of spherical waves, H_{source} the source directivity in source coordinates, $H_{\text{air att}}$ the low pass of air attenuation, R_i the reflection factors of the walls involved, and H_{HRTF} the head-related transfer function of the sound incidence at a specified head orientation. The complete binaural room impulse response is composed of the direct sound and the sum of all reflections. This filter is appropriate for the convolution with anechoic signals to obtain audible results. However, basic methods of GA do not cover two important wave phenomena, that is sound transmission and sound diffraction. Thus, these methods fail to correctly simulate sound propagation from a hidden source to a receiver where the direct line of sight is blocked by other objects, e.g. an obstacle or a door to an adjacent room. Therefore, more sophisticated simulation techniques are required that reflect the real world experience.

Sound transmission prediction tools are well established. They are based on statistical energy analysis and enable the calculation of energy transmission via sound and vibration transmission paths. The implementation of sound transmission in auralization software can be done using filters that are interpolated from spectra of transmission coefficients [17], with secondary sources radiating transmitted sound in adjacent volumes [18, 19]. In contrast, diffraction is - especially in real-time systems - often neglected or poorly modeled due to its analytical complexity. However, the lack of diffraction

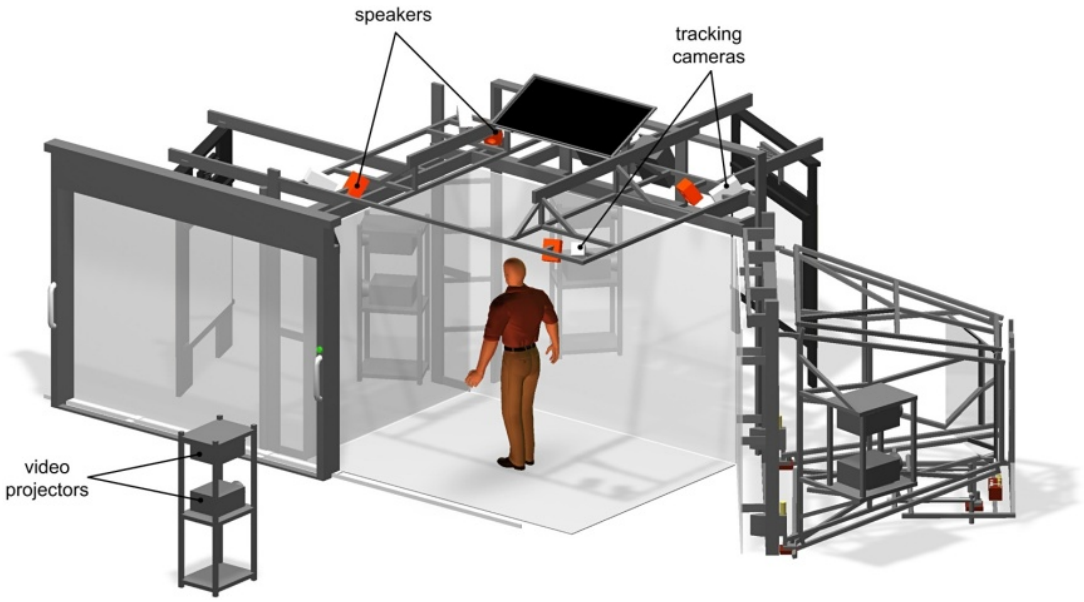


Figure 2: Cave-like environment at RWTH Aachen University. Four loudspeakers for dynamic CTC and infrared tracking cameras are mounted on top of the projection screens.

causes a significant error in most simulations. This becomes more evident by the example of a simple noise barrier that separates a sound source from a receiver. Here, a shadow zone grows clearer and sharper with increasing sound frequency. This zone results from a total cancellation of the incident wave by the diffraction wave which is radiated from the object's edges or perimeter to the receiver. Due to a matter of principle of GA, that is the linear propagation of sound rays, basic methods fail to detect any sound energy inside the shadow zone of such a barrier. Fortunately, analytical and stochastic diffraction models based on GA have been developed which maintain a smooth transition from the view zone to the shadow zone, meanwhile even in more complex scenarios (e.g. [20, 21, 22, 23]).

Audio Rendering

The final step in the auralization chain is the convolution of the simulated impulse response with a dry excitation signal, usually speech, music, or ambient sounds [24]. Since room impulse responses are usually quite long, the convolution can become a computationally very intensive task. By now, powerful hardware and fast convolution algorithms exist, that enable the realization of the entire audio rendering by means of high-quality FIR filtering. The mathematical background of convolution is well known and it can be easily implemented in time- or frequency-domain using MATLAB or similar tools. In contrast, convolution-based real-time audio rendering is an advanced problem in itself. Various requirements must be fulfilled, such as low latencies, rapid exchangeability of filters without audible artifacts and high signal-to-noise ratios. These can only be met by specialized algorithms. The state-of-the-art method is non-uniformly-partitioned convolution in the frequency-domain [25, 26]. It unites a high computational efficiency with low input-to-output latencies, by partitioning the filter impulse responses into a series of subfilters with increasing size. For a smooth exchange of filters cross-fading in the time-domain is commonly applied.

THE VIRTUAL REALITY SYSTEM

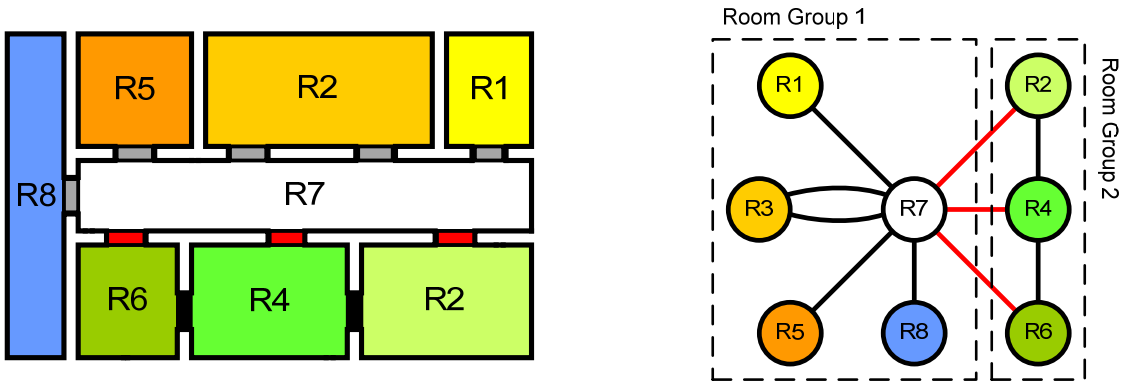
Background and base technology

Shortly after the establishment of the first VR developments at RWTH Aachen University, the activities in computer science were joined with those in acoustics. The advantage was that both groups had deep knowledge in their specific field so that the competences could be combined with high synergy. The initial step was the integration of interactive VR technology (visual and haptic) with headphone-free audio reproduction. At that time the decision was made in favor of a stereo loudspeaker setup for an adaptive crosstalk cancellation. The first task was integrating head tracking and adaptive filters into a Cross Talk Cancellation (CTC) system [27] that turned out to be a flexible solution for various display environments.

In 2004 a CAVE-like five-sided surround-screen projection system was installed at RWTH Aachen University. It has a size of 3.6m·2.7m·2.7m and can be reconfigured using a slide door and a movable wall (see Fig. 2). Stereoscopic images are produced by two images per screen with a resolution of 1600x1200 pixels each and are separated by polarized glasses. It uses several IR cameras for tracking several input devices and the user's head position/orientation. For the reproduction of acoustic signals, four loudspeakers are installed at the top of the CAVE. This setup was chosen over a simple stereo system in order to achieve a good binaural reproduction that is independent from the current user's orientation (see below).

The VISTA Software platform

At RWTH Aachen University, the VR Toolkit ViSTA has been under development for more than 10 years now in order to provide an open, flexible and efficient software platform for the realization of complex scientific and industrial applications [28]. One of the key features of ViSTA comprises functionality for the creation of multimodal interaction metaphors, including visual, haptic, and acoustic stimuli. For such elaborate, multimodal interfaces, flexible sharing of different types of data with low latency access is needed while main-



(a) Floorplan.

(b) Corresponding Acoustical Scene Graph.

Figure 3: Example of the applied scene decomposition: (a) floorplan of a common office building and (b) corresponding acoustical scene graph describing the topological structure of a room acoustic scene and acoustic coupling of interconnected rooms via portals.

taining a common temporal context. Therefore, ViSTA comes along with a high-performance device driver architecture. It provides a novel approach for history recording of input by means of a ring buffer concept that guarantees both a low latency and a consistent temporal access to device samples at the same time [29].

In acoustical reproduction based on binaural synthesis and crosstalk cancellation, latency of the (optical) tracking system is especially critical. For this reason, a compensation scheme has been developed for ViSTA that, based on current tracking samples, can predict the state of the human head position and orientation for the time of application.

Sketch-based modification

Apart from the multimodal reproduction of a scenery, it is also important that one can interact with the virtual environment in an intuitive way. An example for this is a virtual room acoustics laboratory, where the user can perform modifications of the scenery – e.g. by changing the material properties of a wall, creating a piano, or adjusting a reflector panel (see Fig. 1(b)), and then directly perceive the impact on the acoustics. For this purpose, special interaction techniques are required that match the demands of immersive virtual environment, i.e. they are easy-to-use, use small and light-weight input devices and avoid disturbing graphical interface elements. Consequently, a sketch-based interface was developed for the interaction with architectural sceneries where the user can draw three-dimensional command symbols that are then recognized by a real-time symbol matching algorithm. Recognized symbols execute commands such as the creation of a window, and can also contain additional information such as the size and position of the window. The sketch-based interaction provides a considerable number of possible commands that are quickly executable by using an intuitive pen-like input device. More details are given in [30].

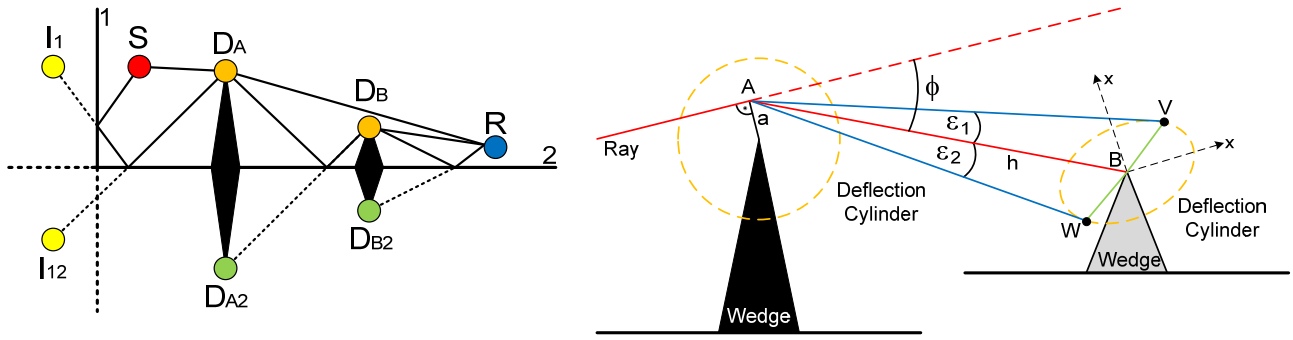
SOUND FIELD RENDERING

For real-time sound field rendering, the hybrid room acoustics simulation software RAVEN (Room Acoustics for Virtual ENvironments) was integrated into the VISTA framework as a network service. RAVEN combines a deterministic image source method [31] with a stochastic ray-tracing algorithm [32] in order to compute high quality room impulse responses on a physical basis. This also includes the simulation of the sound phenomena of sound transmission and sound diffraction (see below). RAVEN imposes no constraints on scene interaction, meaning that not only sound sources and receivers can move freely in the virtual environment, but also the scene geometry can be manipulated by the user at runtime. This is achieved by using advanced modularized and flexible data structures that separate the simulation into single parallel processes that are then distributed and processed on a computing cluster (see below).

Sound transmission

In contrast to simple one-room situations, the real-time auralization of complex environments requires a very fast data handling and convenient interaction management. Imagine an office floor where sound-emitting sources are located in each room. Sound propagation and transmission paths have to be computed from any source to the receiver. To overcome the complexity of such huge geometric models, RAVEN uses a logical and spatial subdivision that enables a dynamic scene decomposition into acoustically separated volumes. Rooms inside a building are such volumes, but also the building's outer surroundings resembling a more free-field-like situation with fewer reflections. Rooms are interconnected by portals, which are logical and physical room separators, e.g., walls, doors and windows. Portals can have individual states, for instance, a door can have the state 'closed' or 'opened', while a solid wall has to be understood as a permanently closed portal. Rooms that are interconnected by open portals must be handled as one integral acoustic space, called room groups.

The topological structure of the scene can be represented by a graph, which is a general data structure for the logical concatenation of entities, called nodes. In the following, this graph will be referred to as Acoustical Scene Graph (ASG) where each node stores the spatial representation of a single room, including the polyhedral model (encoded in special spatial data structures), material data, temperature, humidity and air pressure. The ASG's edges represent polygonal portals, which connect adjacent rooms, i.e. nodes. As sound waves will pass the portals from both sides, these edges are undirected. The connectivity between two nodes is steered by the state of the respective portal. The state ‘closed’ dissociates two rooms from each other, while the state ‘opened’ pools two adjacent rooms in the same room group. Fig. 3 further illustrates the concept of ASG by the example of an office floor. On the left-hand side the floor plan is given. The scene contains eight rooms that are interconnected by ten doors, i.e. portals. All this information is encoded in the ASG, which is shown in Fig. 3(b). With the given portal configuration, the



(a) Sound paths between a source S and a receiver R. Image sources and diffraction sources are denoted by I and D, respectively. Note that some sound paths are omitted for the sake of a clear arrangement.

(b) Selective energy transfer method for an edge-to-edge situation using the 2D-DAPDF. The outgoing energy is dispersed on a plane which intersects with a second deflection cylinder.

Figure 4: Implemented edge diffraction models in RAVEN for both prediction methods: (a) Image sources and (b) Ray tracing.

scene can be divided into two room groups representing the currently coupled acoustic spaces (compare Fig. 3 (a,b)). For the simulation of sound transmission, the building acoustics auralization concept by Thaden et al. [18] was integrated and further extended to a portal-related secondary source model. Here, in addition to the physical and logical scene representation through the ASG, a portal constitutes both, a receiver and a secondary source. While the portal's sender is always modeled as a point source, two different types of receivers are applied using a point receiver and a surface receiver for image sources and ray-tracing, respectively. Each sound propagation path from a primary sound source through air requires a room acoustics simulation to compute the respective impulse responses, while the performance of the portal itself can be described by a transfer function, which is built from interpolated transmission coefficients of the corresponding structural element. More details on this concept of sound transmission auralization are given in [19, 33].

Diffraction

As mentioned above, RAVEN also accounts for the wave phenomenon of diffraction using a hybrid approach for the simulation of sound diffraction, which allows the simulation of higher-order edge diffraction. For this purpose, existing GA methods of edge diffraction have been adapted and optimized. The concept of secondary sound sources by Svensson et al. [20] was chosen for the image source method, as the method allows an exact analytic description of higher order diffraction of finite edges. Here, the demand for a very fast and accurate prediction conflicted with the problem of multiple diffracted reflections and the mirroring of the secondary sources, as the complexity of diffraction path searches and number of secondary sources rises exponentially with diffraction order. Therefore, two types of diffraction edges were introduced, static and dynamic edges. Static diffraction edges cannot be manipulated during the simulation, i.e., they cannot be moved or changed in size. This allows the precomputation of visible edges and secondary sources, which are organized in efficient tree-graphs. By using these data structures diffraction paths up to a range from three to five can be taken into consideration for the online simulation (see Fig. 4(a)). For dynamic diffraction edges, i.e., edges that are fully scalable and moveable, this order must be reduced to at most order two due to the complexity of regenerating the graphs for higher order diffraction. However, it should be kept in mind that this affects only the actual process of manipulation. Once it has been modified, the object's state switches back into static mode.

The diffraction method by Stephenson [21], which is based on Heisenberg's uncertainty principle, was integrated in RAVEN's stochastic ray tracer. The core of this approach is the computation of a 2D-deflection-angle-probability-density -function (DAPDF) of energy particles when they pass an edge. This diffraction model fits perfectly to algorithms that model sound propagation as the dispersion of energy particles, such as stochastic ray tracing. Unfortunately, this approach is also computational demanding due to the underlying principle of energy dispersion for higher order diffraction that leads to a very large number of required energy particles. This problem was tackled by the introduction of cylindrical edge detectors, called deflection cylinders. A deflection cylinder counts impacting energy, which is then distributed to other detectors (see Figure 4 (b)). This approach is not exact, but prevents the explosion of computing and delivers very good results in simple test scenarios. A detailed validation is still pending, though. For this purpose, edge diffraction measurements from a noise barrier were carried out in cooperation with Peter Svensson at the NTNU Trondheim [34]. A complete description of RAVEN's edge diffraction concept is published in [35].

Geometry Manipulation

Another important design aspect for interactive room acoustics simulation is the creation of highly flexible algorithmic interaction interfaces that support a maximum degree of freedom in terms of user interactivity. While code adjustments for operations such as the exchange of material parameters and the manipulation of portal states were relatively easy to implement, the requirements of a modifiable geometry turned out to be a quite an algorithmic challenge. After first test implementations it became apparent that RAVEN's acceleration algorithms based on Binary Space Partitioning (BSP) [31] do not meet the criteria of dynamically manipulable geometry since any modification calls for a recalculation of at least large parts of the BSP trees. It was therefore decided to introduce two different modi operandi for scene objects: static and dynamic (similar to the states of diffraction edges). Static objects, such as walls, are not modifiable during the simulation and are therefore processable in a quick and efficient way. A dynamic object, for instance a reflector panel, is adjustable by a user at runtime (see Figure 3 (a)), though it should be emphasized that there is no limitation on the object's shape in general, i.e., the whole room geometry can be defined as dynamic. For the unconditioned modification of dynamic objects, RAVEN switches to a new approach in geometry processing – that of Spatial Hashing (SH) [36, 37].

SH is a method in Computer Graphics, which is usually applied for collision tests with deformable objects [38]. The concept of SH is based on the idea of subdividing the space by primitive volumes called voxels and map the infinite voxelized space to a finite set of one-dimensional hash indices, i.e., a Hash Table (HT), which are en/decoded by a hash function. The advantage of SH over other spatial data structures such as BSP-trees is that the insertion/deletion of m vertices into/from the HT takes only O(m) time. Thus, this method is perfectly qualified to efficiently handle modifications of a polygonal scenery in order to enable a real-time auralization of a dynamically-changing environment. However, a comprehensive performance analysis has shown that the principle of SH can never compete with the performance of the fast BSP tree on a single core computing unit [39]. On the other hand, the approach significantly gains performance from any additional CPU core as the HT data structure is efficiently schedulable in parallel. On a state-of-the-art multicore CPU (four or more cores), the SH approach will therefore outclass the BSP-based method in all situations where the geometry is modified.

While the concept of SH was easily embeddable to the applied stochastic ray tracing algorithm – where it was sufficient to update just both the geometry and the corresponding spatial data structures – a dynamic handling of ISs turned out to be more complicated 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 this approach has been submitted to the journal Acta Acustica united with Acustica, where a short version was presented at the International Congress on Acoustics (ICA) in Sydney, Australia [37].

AUDIO REPRODUCTION SYSTEM

Sound generation

The system supports several sound generation methods: In the simplest case, a virtual sound source plays back a single mono audio file, which can be looped, if necessary. This simple modeling is sufficient for transient sounds, such as background noise. However, interaction with virtual objects often results in a multitude of individual sounds and sound transitions. In order to adequately model these sounds, more advanced sound generation concepts are required. For this purpose, the system implements a sequencer, which allows to playback arbitrary sound samples on a virtual sound source. Medium synchronization between audio and video is ensured by using time codes. This technique applies for a wide range of objects, such as an electric sliding door or a virtual drum set, though it is not appropriate for objects whose sounds are driven by continuous parameters. An example is a virtual electric motor with freely adjustable revolution speed. For such applications the system offers real-time modal synthesis (similar to [40]) and post-processing filters can be added for high-quality sound authoring.

Dynamic crosstalk cancellation

Binaural playback strictly demands the ability to reproduce individual audio signals at each of the user's ears. Headphones guarantee this property by their construction and are therefore ideally suited for binaural playback. In contrast, sound emitted by a loudspeaker will – at least to a certain degree – always reach both ears of the listener. If left uncompensated, this crosstalk destroys the three-dimensional binaural cues. With knowledge of the sound propagation paths (speakers to each ear), a filter network can be designed that eliminates the crosstalk (see Fig. 5). This technology is known as Crosstalk Cancellation (CTC) and has been investigated for some decades now. Several filter design methods are known and different setups of loudspeakers are possible.

Since the user may freely move, the sound propagation paths change over time. Consequently, CTC filters must be adapted in order to keep the listener within the sweet spot and maintain a proper crosstalk cancellation. The position and orientation of the listener is obtained from the motion tracking with a frequency of 60 Hz. A threshold for translation (1 cm) and rotation (1°) is used to trigger the recalculation and update of

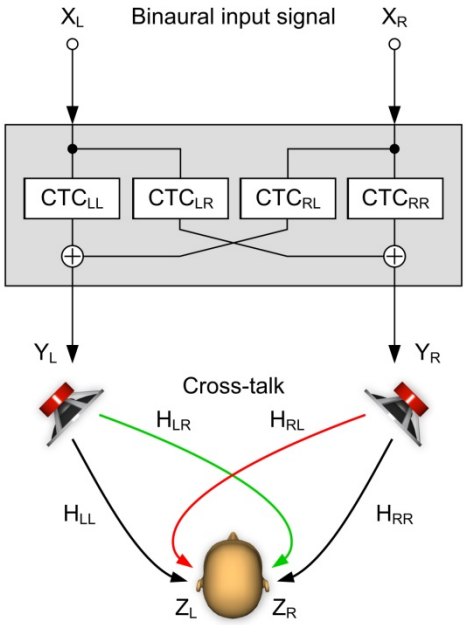


Figure 5: Crosstalk-cancellation for two speakers and one listener. All variables are in the frequency-domain.

the CTC filters. CTC filters prove to be stable only within certain angular ranges that depend on the orientation of the listener with respect to the loudspeaker setup. Only two loudspeakers are not sufficient to cover all possible user orientations within the CAVE. This problem is solved by combining multiple two-channel CTCs over a setup of four loudspeakers into a Dual-CTC algorithm. During runtime this method chooses the best speaker configuration by minimizing the compensation energy [41].

Using free-field HRTFs for compensating the sound propagation paths is valid only for anechoic conditions. The CAVE-like environment, however, is a confined space that is surrounded by acrylic glass walls. Reflections occur, which influence not only the crosstalk compensation, but also the binaural perception. In [42], Lentz investigated the impact of this issue on the localization performance for binaural playback. A distortion of the perceived directions occurred, mainly in the elevation angle, though a combined audio-visual scenario reduced this mislocalization significantly.

Real-time convolution engine

The VR-system that is presented here performs real-time auralization on the basis of high-order FIR filters. The audio rendering itself is a complex and computationally intensive task, which is handled by a dedicated convolution engine. It filters the audio signal of each sound source with an independent binaural room impulse response (BRIR). For realistically sounding scenes, the signals of a high number of virtual sound source (50-100) must be convolved with the results of the room acoustics simulation. These auralization filters (BRIRs) typically consists of 20,000 – 200,000 filter coeffi-

cients. The latencies of the filtering must be very small (<20ms) in order to reproduce the systems' reaction on user input (e.g. movement, rotation, actions) without audible delays. The overall latency depends on the processing delay (input-to-output latency, but also additional delays for the exchange of filters). Moreover, the exchange of filters must not produce audible artifacts.

Summarized under the term 'fast convolution', several efficient methods for FIR filtering are known today. All of them use the efficient Fast Fourier Transform (FFT) to perform the convolution in the frequency-domain by the simple multiplication of discrete Fourier spectra. The FFT is calculated block-wise and thus introduces a fixed input-to-output latency equal to the block length. In order to keep the delay low, partitioned convolution is used for real-time applications. Here, the filter impulse responses are split into several parts, which are convolved individually. The latency can be adjusted by choosing the subfilter sizes appropriately. The most efficient of the currently known techniques is a non-uniformly partitioned convolution [25, 43, 26]. At the beginning of filters it uses small block lengths to achieve short input-to-output latencies. Where affordable, it uses longer filter parts to reduce the computational effort.

The presented system uses a dedicated convolution engine called LLC (low-latency convolver). LLC has been developed at the Institute of Technical Acoustics for several years. It implements a parallelized non-uniformly partitioned fast convolution in the frequency-domain on multi-core machines. A distinctive feature of LLC is the ability to allow an arbitrary impulse response partitioning, which is a key parameter concerning the filter exchange, runtime stability and computational efficiency. Finding an optimal filter partitioning is not trivial. Efficient optimization algorithms exist [26] in order to maximize the convolution performance. LLC uses a filter partitioning that is specifically optimized with respect to the available hardware. Even though it is very efficient, a non-uniform partitioning also puts limitations onto the filter exchange [44]. Longer subfilters can only be exchanged with lower update rates. However, for hybrid room acoustics simulation this is not a disadvantage. The early parts of the impulse responses, containing the direct sound and low-order reflections, are most important for the localization and depend strongly on the user movement. The first 6,000 filter coefficients are partitioned into small subfilters (≤1024 taps), allowing high update rates of more than 40 Hz. The late reverberation varies less when a user moves. Therefore the later parts of the impulse responses require less frequent updates and can be realized using longer subfilters. This lowers the overall computational load. Since a hard switch of filters would result in signal discontinuities and thereby causes unpleasant audible artifacts, time-domain cross-fading is applied to ensure a smooth transition.

Filter rendering

In scenes consisting of multiple coupled rooms, the sound propagation is modeled by filter networks, which describe the sound propagation in form of directed acyclic graphs (DAGs) and is directly derived from the ASG (see above). Such structures cannot be efficiently implemented into the convolution engine without having an impact on the maximum possible number of sound sources. For the audio rendering it is more beneficial to combine filter networks into an equivalent single BRIR, which can then be used with the convolution engine. This process is called filter rendering. It is realized by evaluating all sound paths and combining successive filters using convolution. If a filter element within the DAG changes, the overall filter has to be rerendered. This process does not result in an additional latency concerning the audio streaming, but still should be computed as fast as possible in order to enable a quick adaption to changes in the sound propagation paths. For this application a uniformly partitioned offline convolution is the most efficient algorithm. Here, subfilter sizes are chosen to minimize the computation time, without respect to latencies. By using advanced evaluation strategies, which include memorization of intermediate results, the filter (re)rendering can be performed fast and efficiently. More details can be found in [33].

PERFORMANCE

Parallelization

Advanced real-time auralization concepts post high requirements on the computational performance. In order to meet these requirements, parallelization is extensively used in many components and on all architecture levels. On the most basic level, all arithmetically intensive computations are vectorized using single-instruction multiple-data (SIMD) instructions. Additionally, multi-core computers are used to increase the performance and allow faster calculations. Ray tracing in particular can be efficiently parallelized using OpenMP. Furthermore, the crucial timing dependencies of the real-time convolution demands more advanced concepts. Therefore, special concepts such as flexible data structures are utilized for an efficient parallelization.

Optimizations for multi-core machines allow realizing sceneries of medium complexity. The simulation of complex building environments exceeds the capabilities of a single PC with a limited maximum number of CPU cores. For the simulation of extensive sceneries with multiple coupled rooms, detailed geometries and many sound sources, a cluster of multiple computers can be used to achieve the required computing performance. For this purpose, a cluster-capable version of RAVEN was developed, using MPI and the Viracocha-Framework [45]. It allows distributing different subtasks of the computation, such as individual sound sources, frequency bands, or particle subsets, to different cluster nodes. Specialized scheduling strategies distribute the computation evenly among all nodes. Furthermore, simulation tasks can be prioritized, guaranteeing that IS computations will not be delayed by prior, but less important RT calculations.

Since the top-level interface of RAVEN is unified, the underlying computation hardware is transparent and can be either a single computer or a computation cluster of varying size. This makes the approach very scalable so that the hardware can be chosen to match the complexity of the scenery. All in all, the optimized parallelization strategies make the room acoustics simulation fast enough for real-time processing.

Filtering performance

Currently, a dedicated 2,4 GHz dual quad-core machine is used to realize the filtering. A RME Hammerfall series audio interface is used for sound input and output. Audio streaming is done using Steinberg's ASIO professional audio architecture, at 44.1 kHz with streaming buffersize (blocklength) of 512 samples. For BRIRs of 88,200 filter coefficients, LLC manages to filter the signals of more than 50 sound sources.

FUTURE WORK

An interesting idea for increasing the quality and speed of acoustic simulations at the same time has been introduced by [46]. This approach uses a set of models with graduated level of detail of the same scene geometry, where every single room model is optimized for a certain frequency range, which is important especially for correct reflection patterns at low frequencies. Additionally, with increasing time in the result-

ing impulse response, the level of detail can be decreased during the simulation, providing a total simulation speed-up of a factor of six when combined with the speed-up due to frequency-matched geometries. For even more complex scenes, the computation on graphic cards is very promising. This technique has been successfully applied to auralization [47, 48]. For real-time GPU-based convolution, a highly optimized and promising solution was presented in [49].

ACKNOWLEDGMENTS

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

REFERENCES

[1] T. Lokki, "Physically-based auralization - design, implementation, and evaluation," Ph.D. dissertation, Helsinki University of Technology, 2002.

[2] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, vol. 3(23), 2004.

[3] P. Lundén, "Uni-verse acoustic simulation system: interactive realtime room acoustic simulation in dynamic 3d environments," *The Journal of the Acoustical Society of America (JASA)*, vol. 123(5), p. 3937–3937, 2008.

[4] M. Noisternig, B. Katz, S. Siltanen, and L. Savioja, "Framework for real-time auralization in architectural acoustics," *Acta Acustica United with Acustica*, vol. 94(6), pp. 1000–1015, 2008.

[5] [Online]. Available: www.vrca.rwth-aachen.de

[6] C. Cruz-Neira, D. Sandin, T. DeFanti, R. Kenyon, and J. Hart, "The CAVE: audio visual experience automatic virtual environment," *Communications of the ACM*, vol. 35, pp. 64–72, 1992.

[7] M. Vorländer, "Simulation of the transient and steady state sound propagation in rooms using a new combined sound particle - image source algorithm," *The Journal of the Acoustical Society of America*, vol. 86, pp. 172–178, 1989.

[8] J. Vian and D. van Maercke, "Calculation of the room impulse response using a ray-tracing method," in *Proceedings of the ICA Symposium on Acoustics and Theatre Planning for the Performing Arts, Vancouver, Canada*, 1986.

[9] J.-H. Rindel, "Auralisation of a symphony orchestra - the chain from musical instruments to the eardrums," in *EAA Symposium on Auralization, Espoo, Finland*, 2009.

[10] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," *Journal of the Acoustical Society of America*, vol. 115, pp. 739–756, 2004.

[11] B.-I. Dalenbäck, "Engineering principles and techniques in room acoustics prediction," in *Baltic-Nordic Acoustics Meeting, Bergen, Norway*, 2010.

[12] U. Stephenson, "Quantized Pryamidal Beam Tracing - a New Algorithm for Room Acoustics and Noise Immission Prognosis," *ACTA ACUSTICA united with ACUSTICA*, vol. 82, pp. 517–525, 1996.

[13] M. Vorländer, "International Round Robin on Room Acoustical Computer Simulations," in *Proceedings of 15th International Congress on Acoustics, Trondheim, Norway*, 1995.

[14] T. J. Cox, B.-I. L. Dalenbäck, P. D. Antonio, J. J. Embrechts, J. Y. Jeon, E. Mommertz, and M. Vorländer, "A tutorial on scattering and diffusion coefficients for room acoustic surfaces," *Acta Acustica united with ACUSTICA*, vol. 92, pp. 1–15, 2006.

[15] D. Hammershøi and H. Møller, "Methods for binaural recording and reproduction," *Acustica united with Acta Acustica*, vol. 88, p. 303, 2002.

[16] B.-I. Dalenbäck, "A new model for room acoustic prediction and auralization," Ph.D. dissertation, Chalmers University, Gothenburg, Sweden, 1995.

[17] DIN EN12354-1: 2000, "Building Acoustics - Estimation of acoustic performance of buildings from the performance of elements, Part 1: Airborne sound insulation between rooms."

[18] R. Thaden, "Auralisation in building acoustics," Ph.D. dissertation, RWTH Aachen University, 2005.

[19] D. Schröder and M. Vorländer, "Hybrid method for room acoustic simulation in real-time," in *Proceedings of the 20th International Congress on Acoustics (ICA)*, Madrid, Spain, 2007.

[20] 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, pp. 2331–2344, 1999.

[21] U. M. Stephenson and U. P. Svensson, "An improved energetic approach to diffraction based on the uncertainty principle," in *Proceedings of the 19th International Congress on Acoustics, Madrid, Spain*, 2007.

[22] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom, "Modeling Acoustics in Virtual Environments using the Uniform Theory of Diffraction," *ACM Computer Graphics, SIGGRAPH'01 Proceedings*, pp. 545–552, 2001.

[23] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer-Verlag Berlin, 2005.

[24] M. R. Schroeder, B. S. Atal, and C. Bird, "Digital computers in room acoustics," in *Proceedings of the 4th International. Congress on Acoustics, Copenhagen, Denmark*, 1962.

[25] W. G. Gardner, "Efficient convolution without input-output delay," *Journal of the Audio Engineering Society (JAES)*, vol. 43, pp. 127–136, 1995.

[26] G. García, "Optimal filter partition for efficient convolution with short input/output delay," in *Proceedings of 113th AES convention*, 2002.

[27] T. Lentz, "Binaural technology for virtual reality," Ph.D. dissertation, RWTH Aachen University, 2008.

[28] I. Assenmacher and T. Kuhlen, "The vista virtual reality toolkit," in *Proceedings of the IEEE VR SEARIS*, 2008, pp. 23–26.

[29] I. Assenmacher, D. Rausch, and T. Kuhlen, "On device driver architectures for virtual reality toolkits," *Pres-*

ence: Teleoperators and Virtual Environments, vol. 23, pp. 83–95, 2010.

[30] D. Rausch and I. Assenmacher, "A sketch-based interface for architectural modification in virtual environments," in *5th Workshop VR/AR, Magdeburg, Germany*, 2008.

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

[32] 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.

[33] F. Wefers and D. Schröder, "Real-time auralization of coupled rooms," in *Proceedings of the EAA Auralization Symposium, Espoo, Finland*, Espoo, Finland, 2009.

[34] D. Schröder, P. Svensson, and M. Vorländer, "Open measurements of edge diffraction from a noise barrier scale model," in *Proceedings of the International Symposium on Room Acoustics (ISRA)*, Melbourne, Australia, 2010.

[35] D. Schröder and A. Pohl, "Real-time hybrid simulation method including edge diffraction," in *Proceedings of the EAA Auralization Symposium*, Espoo, Finland, 2009.

[36] D. Schröder and I. Assenmacher, "Real-time auralization of modifiable rooms," in *2nd ASA-EAA joint conference Acoustics, Paris, France*, 2008.

[37] D. Schröder, A. Ryba, and M. Vorländer, "Spatial data structures for dynamic acoustic virtual reality," in *Proceedings of the 20th International Congress on Acoustics (ICA)*, Sydney, Australia, 2010.

[38] M. Teschner, B. Heidelberger, M. Müller, D. Pomeranets, and M. Gross, *Optimized Spatial Hashing for Collision Detection of Deformable Objects*. VMV '03, 2003.

[39] D. Schröder, A. Ryba, and M. Vorländer, "Real-time auralization of dynamically changing environments," *submitted to Acta Acustica united with Acustica*, 2010.

[40] N. Raghuvanshi and M. Lin, "Interactive sound synthesis for large scale environments," in *Proceedings of the Symposium on Interactive 3D Graphics and Games, Redwood City, USA*, 2006.

[41] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments," *Journal of the Audio Engineering Society (JAES)*, vol. 54(4), p. 283–293, 2006.

[42] T. Lentz, "Performance of spatial audio using dynamic cross-talk cancellation," in *119th AES Convention, New York, NY, USA*, 2005.

[43] G. P. M. Egelmeers and P. Sommen, "A new method for efficient convolution in frequency domain by non-uniform partitioning for adaptive filtering," *IEEE Transactions on signal processing*, vol. 44, 1996.

[44] C. Müller-Tomfelde, "Time-varying filter in non-uniform block convolution," in *Proceedings of the Conference on Digital Audio Effects (DAFX-01)*, 2001.

[45] A. Gerndt, B. Hentschel, M. Wolter, T. Kuhlen, and C. Bischof, "Viracocha: An efficient parallelization framework for large-scale CFD post-processing in virtual environments," in *Proceedings of the 2004 ACM/IEEE Conference on Supercomputing, Magdeburg, Germany*, 2004.

[46] S. Pelzer and M. Vorländer, "Frequency- and time-dependent geometry for real-time auralizations," in *20th International Congress on Acoustics (ICA), Sydney, Australia*, 2010.

[47] T. Ikedo and W. L. Martens, "Multimedia processor architecture," in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Austin, TX, USA*, 1998.

[48] N. Tsingos, "Using programmable graphics hardware for auralization," in *Proceedings of the EAA Symposium on Auralization, Espoo, Finland*, 2009.

[49] F. Wefers and J. Berg, "High-performance real-time FIR-filtering using fast convolution on graphics hardware," in *Conference on Digital Audio Effects (DaFX)*, 2010.