Hybrid method for room acoustic simulation in real-time

PACS: 43.55.Ka

Schröder, Dirk; Vorländer, Michael
Institute of Technical Acoustics, RWTH Aachen University, Neustrasse 50, 52066 Aachen, Germany, dsc(mvo)@akustik.rwth-aachen.de

ABSTRACT
In real-time virtual reality systems used today, room acoustics is often added just as an effect without plausible reference to the virtual environment. Thus, these approaches are unsuitable for applications such as virtual prototyping of rooms, for instance, concert halls and complex of buildings, as distinct acoustical effects, e.g., (flutter) echoes and curved decays could remain undetected. The approach presented is a real-time hybrid method for a dynamic room acoustic simulation of virtual environments. It combines deterministic image sources to ensure an exact localization of sound sources with a stochastic ray tracing algorithm for determining the late reverberant sound field. In addition, the concept of scene graphs has been adapted for an efficient and flexible linkage of autonomous-operating entities, i.e., rooms connected by so-called portals, where room neighborhoods can be changed online, for instance, doors may be opened or closed. Using this framework enables an interactive auralization of an arbitrary number of coupled rooms, where freely movable sound sources, a moving listener and geometry manipulation are supported. Furthermore, portals and their respective states (open/closed) are used to efficiently include sound transmission effects to the real-time simulation.

INTRODUCTION
Virtual Reality (VR) is the representation and simultaneous perception of reality and its physical attributes in an interactive virtual environment which is computer-generated in real-time. VR-Technology is used in numerous applications, for instance computational engineering science, mechanical engineering, medicine and architecture. The focus of these applications lies usually on the visual component of the VR environment and acoustics is often added just as an effect without any plausible reference to the physical aspects of the presented scenery. However, it is important to address at least the hearing as well to enforce the feeling of actual presence in the simulated scene (degree of immersion). Especially in the case of architectural/room acoustic applications such as an immersive virtual walk through rooms, e.g., complex of buildings and concert halls, physical models of room- and building acoustics are demanded in order to detect in real-time distinct acoustic effects, e.g., flutter echoes, curved decays and deficient airborne sound insulation.

The determination of the spatial sound field of virtual environments is a difficult task, especially under real-time constraints. Even rather simple situations require quite complex acoustic models and the more the user is allowed to interact with the scenery, the more this complexity increases, e.g., in the case of movable sound sources, movable objects and coupling of rooms. For simulating sound propagation in rooms, a model is required which describes sound emitting sources, physical characteristics of rooms, transmission effects in the case of coupled rooms, and the receiver. A human receiver evaluates these events on attributes which are characterized for example by loudness, coloration, source localization, spaciousness, and the perceived distance to the source. From a system theoretical point of view the perceived sound signal can be represented by the binaural room impulse response (BRIR) convolved with a dry recorded signal of the emitted sound. Thus, the key for a high quality auralization of virtual environments is the computation of BRIRs in real-time in order to produce a plausible sound field at the receiver’s position. In this context, the term plausible means that a number of approximations have to be made due to limited computing power and real time constraints,
respectively. However, the resulting sound is not intended to be physically absolutely correct, but perceptively plausible. Therefore, knowledge of the human sound perception and the field of psychoacoustics are essential to find the optimal balance between the needed accuracy on the one hand and the available computation power on the other hand.

HYBRID ROOM ACOUSTIC SIMULATION
The room impulse response can be divided into three parts: the direct sound, the early reflections and the late reverberation. From a perceptive point of view, the direct sound and early reflections affect the localization of sound (precedence effect). A human receiver separately processes the individual reflections (level and angle of impact) and gets an impression for the position and distance to the source. At later times the single reflections are more and more overlapping and the human hearing starts to perform an energetic integration over a certain time slot and angle field [1]. Thus, the early part of the impulse response has to be modeled as accurately as possible whereas a lower time resolution is allowed for the late reverberant sound field.

The benefit of the image source model [2, 3] is that it exactly captures every possible specular sound path up to a pre-defined reflection order. This deterministic approach is perfectly qualified for simulating especially the early reflections due to the exact time resolution. Unfortunately, the image source model misses a proper representation of a very important aspect of room acoustics, namely surface and obstacle scattering. It was pointed out by Kuttruff in [4] that scattering becomes even a dominant effect in the temporal development of the room impulse response, already from reflections of order two or three (see Figure 1). This implies that pure image source modeling would be too much of an approximation of the late part of the room response, which has also been attested by intercomparisons of room acoustics simulation programs [5].

![Figure 1: Distribution of specular and diffuse reflected energy (after Kuttruff [4]).](image)

Better results are achieved by combining image sources with stochastic models for the simulation of the room’s reverberant sound field. In contrast to deterministic image sources, stochastic approaches take into account energetic models for the simulation of sound scattering, e.g., Lambert's cosine law. Especially the late part of the impulse response is important for a plausible representation of the room’s sound characteristics. Short reverberation times are usually associated to smaller rooms, or vice versa. In addition, spaciousness or spatial impression is strongly related to the listener envelopment and the apparent source width. Bradley showed that especially the level, direction of arrival, and temporal distribution of late arriving reflections have a strong influence on these two important psychoacoustic parameters [6]. Therefore, temporal and spectral directivity information of late arriving sound are to be included to the binaural synthesis in order to create a sound field with a certain performance of realistic physical behavior.

The hybrid room acoustics simulation software RAVEN (Room Acoustics for Virtual Environments) is currently being developed at the Institute of Technical Acoustics, RWTH Aachen University, Germany, which takes into account all criteria mentioned above. RAVEN combines deterministic image source method [7] with a stochastic ray tracing algorithm [8] in order to compute impulse responses in real-time which reach state-of-the-art room acoustics
simulation standards. RAVEN is incorporated into the 3D sound rendering system of RWTH Aachen University’s CAVE [9] and is used for any room acoustics simulation as a networked service (see Figure 2).

![Diagram of VR Application](image)

**Figure 2:** Brief system overview (further details can be found in [9]).

**BUILDING ACOUSTIC SIMULATION**

To maintain an immersive simulation in the case of complex sceneries, e.g., a complex of buildings, the online auralization of airborne sound from neighboring rooms is mandatory, as it secures the believability of the virtual scene and, thus, the user’s feeling of immersion. For instance, the opening and closing of a door to a neighboring room, where a source emits sound at high volume, would be perceived unnaturally if the source is simply switched on and off, respectively. Instead, at least the dominant sound characteristics of airborne sound, i.e., the level (low) and coloration (dull), have to be auralized.

In the case of coupled rooms, the process of sound transmission between two rooms includes the room acoustic simulation of the sound field in the source room, the modeling of the excitation and radiation of bending waves in the walls, and a room acoustic simulation of the sound field in the receiver room. Fortunately, the airborne sound transmitted from an adjacent room into the receiver room through direct and flanking room components can well be auralized by using secondary point sources located in the centre of each transmitting structural element of the receiver room [10, 11]. The benefit of this kind of building acoustics auralization method is that it requires only the sound reduction index, $R$, of each construction element as additional input data, where the single indices can be measured after ISO 140, while their overall performance in the building structure, including direct and flanking transmission, can be computed according to standardized simulation models, e.g., the European Standard EN 12354. The auralization concept of sound transmission is shown in Figure 3.

In summary, in an acoustic linear and time invariant system, the transfer function $H$ between the source and the receiver is described by

$$H = \sum_{j=0}^{N} \left( \sum_{i=0}^{M} H_{S,i} \cdot H_{i,j} \right) \cdot H_{R,j},$$

where $H_{S,i}$ denotes one of $M$ room transfer functions between the source and the source room’s structural elements $i$. $H_{R,j}$ describes one of $N$ room transfer functions between the centre of each transmitting element $j$ of the receiver room and the receiver itself. $H_{i,j}$ is the transfer function of the transmission path between the structural elements $i$ and $j$, which can be described by a standardized filter function. This filter is built from the interpolated transmission coefficients of the respective structural elements, where the transmission coefficient $\tau_{i,j}$ relates to the standardized sound reduction index $R_{i,j}$ with

$$\tau_{i,j} = 10^{-R_{i,j}/10}.$$

From this it follows that the airborne sound from a source $S$, which is located in an adjacent room, can well be auralized by using secondary point sources $S_j$ with

$$S_j = S \cdot \sum_{i=0}^{M} H_{S,i} \cdot H_{i,j} \; , \; 0 < j < N$$
where each point source is positioned at the centre of the respective radiating structural element of the receiver room (see Figure 3(b)). It was shown in [12], that no improvement can be achieved if surface sources are used instead.

**Figure 3:** (a) Simple room-to-room situation, where a source $S$ is located in room 1 and a receiver $R$ is located in room 0, with $H_{S,i}$: source room transfer functions between source and the structural elements $i$, $H_{R,j}$: receiver room transfer function between the centre of the element $j$ and receiver, $H_{ij}$: transfer function of the transmission paths between the structural elements $i$ and $j$. For better visibility some transmission paths are omitted. (b) As an Example: The source room is acoustically substituted for 7 secondary point sources located in the receiver's room, i.e., direct and first order junctions are taken into account.

However, a complete simulation of all transmitting parts between two rooms is infeasible under real-time constraints. Thus, the first focus lies on the auralization of airborne sound, which is directly transmitted through structural elements with a low sound level difference such as windows or doors and therefore dominates the overall level of the transmitted sound. This implies that the whole source room can be represented with sufficient accuracy by only one point source located in the centre of, for instance, a door.

**INTERACTIVE REAL-TIME AURALIZATION**

In contrast to one-room situations, the real-time auralization of a complex of buildings requires a very fast data handling and convenient interaction management. Imagine a concert hall with adjacent foyer and rehearsal room, where sound emitting sources are located in any 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, the dynamic elements from the scene, such as doors and windows, are used as logical and physical separators between rooms. Information about these separators, which will be called portals in the following, and their respective states, e.g., “opened” and “closed” can be exploited to significantly speed up the simulation. Parts of the geometric model can be excluded from the room acoustics simulation, while other parts can be replaced by means of the secondary source model for sound transmission (see previous chapter). In other words, the concept of portals allows a very controlled and efficient simulation of sound propagation and sound transmission (see Figure 4).

**Figure 4:** Inclusion, exclusion and substitution of rooms depending on portal states
Therefore, RAVEN uses the concept of scene graphs, which is a general data-structure for the logical concatenation of entities, so-called nodes. Each node contains the spatial representation of a single room, while the scene graph’s edges represent polyhedral portals which connect adjacent rooms. The connectivity between two nodes is steered by the state of the respective portal. The state “opened” connects the certain nodes. The state “closed” either cuts off the specific link (Source $S_A$, see Figure 3) or substitutes the node for a secondary source, which represents the airborne sound at the portal’s position (Source $S_B$, see Figure 4).

In order to save computation time, the secondary sources are pre-processed. As long as the portal is outside the critical distance of the primary source, no updates are required, as no relevant changes of the sound level and sound coloration will appear. However, if the primary sound source is moved towards the portal, the sound level of the airborne sound will be too low if the portal is inside the critical distance of the source (see Figure 5). Therefore, the impulse response between the source S and the portal, i.e., $H_S$ will be updated by testing the source’s valid image sources (in this example image sources of room 1 only) on audibility in order to replace the level-dominating early reflections of the impulse response and, thus, correctly adjust the level of the secondary source $S_{Portal}$ to the changed primary source’s position.

This reveals another advantage of the applied scene graph data structure, as it allows the pre-computation of subsets of potentially audible image sources. This is done by sorting the entire set of image sources dependent on the room(s) they originate from (see Figure 5(b)). Using this subsets, portal states can be exploited to reduce remarkably the overall number of image sources, which have to be processed during the simulation (compare Figure 4 and Figure 5(b)).

**Figure 5:** (a) As soon as the portal lies within the critical distance around source S, the secondary source $S_{Portal}$ is adjusted to the new position of S. (b) Image sources/room-combination table. All image sources are sorted into room-combination-dependent encapsulated containers. In this example, only image sources of room 1 have to be taken into account for the current operation- the update of $H_S$ (compare Figure 4, source $S_A$-configuration, portal closed).

**SUMMARY & OUTLOOK**

A hybrid method for room acoustics and building acoustics simulation of virtual environments is presented which is able to cope with real-time requirements even in the case of complex sceneries. This is achieved by the introduction of a very efficient portal-driven framework for the simulation control. Portals are extracted from the building geometry and require no manual overhead. A physically justified model for simulating the relevant portions of sound transmission between neighboring rooms is introduced, where updates are efficiently controlled by taking portal states and portal/source-distances into account. In this contribution a model for the auralization of airborne sound insulation by structural elements with a low sound level difference such as doors is described. However, the model can easily be extended to auralize direct and first order flanking transmission paths of arbitrary structural elements.

**References:**
