Real-time filtering structures for interactive geometry modification in acoustic virtual reality

Frank Wefers, Dirk Schröder

Institute of Technical Acoustics,
RWTH Aachen University,
52066 Aachen, Germany
Email: {fwe, dsc}@akustik.rwth-aachen.de

Abstract

Interactive modification of scene geometry is an important factor to improve the user's immersion in a virtual acoustic environment. However, modifications such as changes of surface properties and the scene geometry itself can result in strong sound field variations within very small time intervals. The opening and closing of a door to a neighbouring room which contains a high volume sound source illustrates such an event. In the sense of physical consistency, it is not sufficient to simulate these events by just adapting the volume of the sound source, as the process of opening and closing of a door usually results in a significant change of reverberation time, sound coloration and source localization.

Real-time auralization by means of finite impulse response filters (FIR filters) is usually performed by non-uniformly partitioned frequency-domain convolution. Slight changes in room acoustics can easily be incorporated into these algorithms, but fast changes require special filter exchange strategies. The low-pass characteristic of a door being closed will immediately influence the whole room impulse response (RIR). Therefore, the entire filter must be exchanged instantly to maintain physical consistency.

In this contribution, efficient strategies for fast exchange of filters are proposed for a non-uniform partitioned convolution approach. Their relations to filter partitions are discussed and the computational costs are analyzed.

Introduction

In todays virtual reality (VR) systems an appropriate simulation of the acoustics becomes more and more important. Considering a virtual walk through a complex building, the user will typically be faced with a multitude of sounds originating from sound sources inside the user's room, but also from sound sources located in adjacent rooms. Architectural applications like this demand for physical-based auralization methods in order to reproduce sound pressure signals with a certain realistic behaviour. As a static environment would mostly mismatch with real world experiences, the system should also allow a high level of user interactivity to further improve the simulation in terms of immersion. This can include the dynamic modification of scene elements, e.g. the opening and closing of doors and windows, the change of acoustic material properties of surfaces, and even a modification of the geometry itself, e.g. adding new objects or moving walls. Many virtual acoustic systems exist that provide a sophisticated simulation of room acoustics [3][4][5]. Most of them auralize scenes that consist of a single room only.

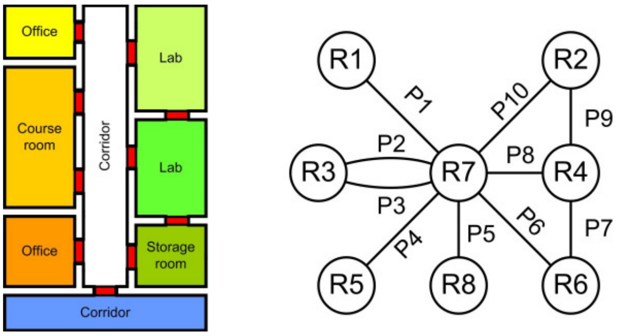

Figure 1: *Floor plan and acoustic scene graph (ASG) of an example scene: A part of an office complex. Rooms are labeled R1, ..., R8. Portals (in this case doors) are labeled P1..., P10.*

Auralization of room acoustics using RIR is often realized by FIR filtering [1]. In scenes that only consist of a single room, the sound propagation can be described by a single RIR per (source,listener)-pair. Consequently, the real-time filtering is structurally simple. Each sound source's signal is filtered with one corresponding RIR. For this application fast convolution techniques exist that allow to filter a great number of channels with long impulse responses at a short input-to-output latency, required for interactivity. The standard method for this filtering application is non-uniformly partitioned convolution in the frequency-domain (NUPFDC) [6][7].

In this article we focus on the real-time filtering necessary to real-time auralize complex sceneries of coupled rooms. We show that the auralization of scenes that consist of a multitude of coupled rooms result in networks of filters. It is discussed if and how existing filtering techniques can be used to simulate these filtering networks and present solutions. We then point out difficulties concering the real-time implementation and introduce a method that enables the real-time auralization also for large-scale sceneries. This is achieved by rendering filtering networks into equivalent single filters. The performance of the method is discussed and finally, we describe how the handling of scene dynamics can be realized by our approach.

Sound propagation modelling

Sound propagation inside buildings is highly complex. In [2] we present a method that allows an efficient real-time simulation of the sound propagation between and across rooms. In this paper we pick up the results from [2] and focus on the aspect, how to perform the real-time filtering necesarry for the real-time auralization. As they build the foundation for the considerations in this work, we firstly give a brief summary of the concepts presented in [2].

Our method is to restrict the sound transmission between rooms to discrete-space coupling joints, called *portals*. This discretization reduces the number of individual sound propagation path to numbers that are manageable in real-time simulations. A portal is a receiver and sender in one object. Sender and receiver are interconnected by a *portal filter* (*PF*) that simulates the sound transmission through the portal by a transfer function that is derived from the portal's sound reduction index R [1]. Portals can be used to simulate static couplings between rooms, like rigid walls, but also dynamic couplings, like for instance doors that can be opened and closed.

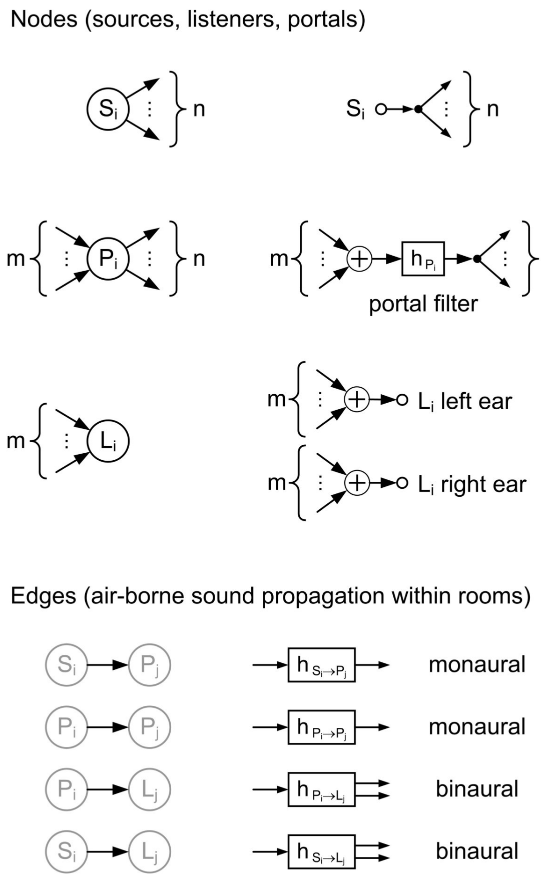

Figure 2: *Transformation rules that allow to transform a TP-DAG describing sound transfer paths into a filtering network that simulates the sound propagation from one sound source to one listener.*

We consider a virtual acoustic scene to consist of a finite number of *rooms* $R_1, ..., R_k$, which are interconnected by a finite number of portals $P_1, ..., P_l$. The topological structure of a scene is described by an *acoustic scene graph* (*ASG*), in which nodes present rooms and edges represent portals. Figure 1 shows an example scene, a part of an office, in which rooms are interconnected by doors. Structural coupling by walls is left unconsidered here. To the right of the floor plan the corresponding ASG is depicted.

A scene furthermore contains a number of sound sources $S_1, ..., S_m$ and listeners $L_1, ..., L_n$ which have a position and orientation in three-dimensional space. Determined by their position, sources and listeners are assigned to the rooms of the scene.

Portals have an individual state. They can be fully opened, closed or opened to a certain angle. Changes of these states make it necessary to introduce a logical layer on top of the rooms: *Room groups* (*RGs*). Room groups are *acoustically separated* volumes, meaning that the acoustics in these volumes can be simulated individually, without a degradation of quality. This e.g. holds for two rooms interconnected by a closed door, but it is not applicable for the case of two rooms interconnected by a fully opened door. Rooms sharing one or more open portals are merged into room groups, of which each is then simulated as a whole. For the sake of complexity we leave room groups unattended here and consider each room to be acoustically separated.

Each sound propagation path from a (primary) sound source to a listener trespassing structural elements is divided into sub paths. Four different types of sub paths exist: a) Source-to-Listener (S2L), b) Source-to-Portal (S2P), c) Portal-to-Portal (P2P) and d) Portal-to-Listener (S2L). The airborne sound propagation within rooms is simulated using room impulse responses (RIRs), which are calculated by hybrid simulation [2] using image sound sources and acoustic stochastical ray-tracing. Within the listener room binaural room impulse responses (BRIRs) are used. Elsewhere monaural RIRs are employed, because spatial localization is not possible here anymore (e.g. a sound source behind a door). The structure-borne sound transmission is described by the standardized sound reduction index R of the structural element [1]. It can be simulated using either low order FIR or IIR filters.

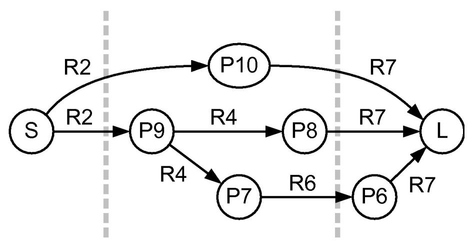

Figure 3: Example of a *resulting TP-DAG for a sound source located in room R2 and the listener residing in room R7.*

We consider the sound propagation through a scene for pairs of sources and listeners. The sound propagation for a (source, listener)-pair is represented by a directed acyclic graph (DAG), called the *transfer-path DAG* (*TP-DAG*). The TP-DAG nodes correspond to the scene objects (the source, the portals and the listener). The directed edges state the airborne sound propagation within the rooms. For a given scene configuration (portal states and position of source and listener) the TP-DAG can be constructed using a search algorithm, described in [2]. The algorithm unrolls the TP-DAG starting from the source room node and builds up and continues sound transfer paths successively. Before each continuation, audibility checks are performed for several frequency bands, guaranteeing that just audible sound propagation paths are considered.

Filtering networks

Within scenes that consist of a single room only, the sound propagation from one sound source to one listener can be simulated by using a single BRIR [5]. The auralization of the sound propagation across various coupled rooms is more complex and results in networks of filters. Each TP-DAG can be translated into a corresponding filtering network using the set of transformation rules, which is listed in figure 2.

Each source node S_i translates into a mono input of the sound source's signal. Edges state sound propagation inside rooms and are substituted with RIR or BRIR filters. S2L-edges $S_i \rightarrow L_j$ as well as P2L-edges $P_i \rightarrow L_j$ are translated into binaural filters $h_{S_i \rightarrow L_j}$ and $h_{P_i \rightarrow L_j}$, respectively. S2P-edges $S_i \rightarrow P_j$ and P2P-edges $P_i \rightarrow P_j$ result in monaural RIR filters $h_{S_i \rightarrow P_j}$ and $h_{P_i \rightarrow P_j}$, respectively. All portals, as well as the listener, act as summation points where several sound waves superpose. The listener node L_j is transformed into two separate summations, one for each ear signal. In case a node's number of input/output edges equals one, the summation is dropped in the filter network. Sound transmission at a portal P_i is realized by the portal filter h_{P_i}, which is placed after the summation. The transformation of the example TP-DAG depicted in figure 3 results in the filtering network that is shown in figure 4.

Real-time implementation

In order to auralize scenes, the filtering networks for each (source, listener)-pair must be implemented in real-time efficiently. Todays standard method to perform real-time FIR filtering of several channels with long filters (e.g. room impulse responses) by ensuring a short input-to-output latency is *non-uniformly partitioned convolution in the frequency domain* (*NUPFDC*). In this algorithm, filter impulse responses are divided into several subfilters of different lengths, starting with a short subfilter to realize a short input-to-output latency and then increasing the subfilter lengths to lower the computational effort [6][7]. There are many publications on this technique.

A logical step is trying to utilize this matured technology to realize the filtering networks. Therefore, every computationally intensive RIR in a filtering network is implemented by a dedicated real-time convolver unit. Due to the network structure, several convolver units are cascaded, whereas compared to single room scenaries, just a single multi-channel convolver unit is required. The portal filters are uncritical in terms of complexity and are realized by low-order FIR or IIR filters. Lining up several convolver units causes problems in real-time implementations, when many sound sources shall be auralized and the filtering networks become complex (in terms of branching and path depths). This is because the first subfilter in every NUPFDC convolver must be processed as soon as the input data is present, within a limited-time callback from the audio hardware [7]. Processing networks of linked convolvers increases the computation within the callback significantly and eventually limits the possible maximum number of sound sources and scene complexity, even if the system is not fully utilized.

Since we want to build a high-quality auralization system also for large-scale applications, we come up with another solution: Alternatively, *whole* filtering networks can be *rendered* into an equivalent single BRIR. This rendering consumes extra computations, but has the advantage that a single convolver unit can be used again and only one direct calculation must be performed.

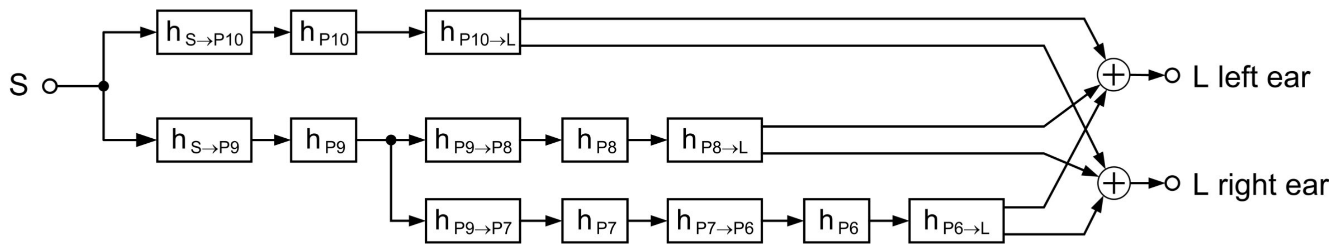

Figure 4: Resulting filtering network of the transformed example TP-DAG shown in figure 3.

For the rendering, we attributed each node N of the according TP-DAG with a partial FIR filter $h^N(n)$. The rendering starts with the source node S that is assigned $h^S(n)=\delta(n)$. Then each attribute $h^N(n)$ of a successor node *N* is assigned by the attributes of its predecessor nodes *pred(M)* as follows:

$$h^N(n) = \sum_{M\in pred(N)} h^M(n) * h_{N\rightarrow M}(n)$$

The rendering ends in the listener node L which is then attributed with a BRIR equivalent to the whole filtering network. For the rendering (combining of impulse responses) an efficient *offline* convolution algorithm is required. Clearly, the rendering should be as fast as possible, but there is no input-to-output latency requirement during the rendering. We therefore employ *uniformly-partitioned frequency-domain convolution* (UPFDC) with an optimized blocklength. This algorithm is most efficient for offline convolution and the whole rendering of a filter network becomes a matter of very few milliseconds on a current desktop system.

Handling scene changes

Scene changes always manifest in modifications to the filtering networks and their components. Here we list the most important changes, ordered by ascending complexity of their handling by the system and describe how the system handles them.

1. A listener/source changes position/orientation, but stays within its room
2. A listener/source changes its room
3. A portal is opened/closed
4. The state of a room (geometry, surfaces) is changed

All of the changes require the rerendering of filtering networks (at least to some extent). Change 1 requires to recalculate the RIR/BRIR of the sound propagation within the sender or receiver room, respectively. Change 2 influences the transfer-paths structurally and requires to reconstruct corresponding TP-DAGs. Change 3 influences the room groups, which are divided or merged, respectively. This will mostly lead to structural changes of transfer-paths. Therefore the TP-DAGs must be reconstructed. Change 4 requires to recalculate the sound propagation within inner rooms. Filter networks have to be rerendered entirely in this case, which remarks the worst-case. The computational effort of the filter rerendering can be lowered significantly, when intermediate evaluation results are cached in previous renderings and can be reused.

Conclusions

In this paper we have presented a real-time filtering method for the real-time auralization of room acoustics in scenes that consists of coupled rooms. We introduced a method to describe the sound propagation in a scene structurally using TP-DAGs and showed how these can be transformed into filtering networks, that describe the signal processing necessary for the auralization of the scene. Established real-time filtering methods can be used to realize these networks, but only for scenes of lower complexity. For large-scale sceneries we presented a method to render filter networks into single equivalent filters and discussed its performance. Finally, we turned our attention to scene dynamics and described their handling by our method. Video presentations of our technique can be downloaded from our website: `http://www.akustik.rwth-aachen.de/raven`

References

[1] M. Vorländer, Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, Springer, 2007.

[2] D. Schröder and M. Vorländer, “Hybrid method for room acoustic simulation in real-time,” in Proceedings of the 19th International Congress on Acoustics, Madrid, Spain, 2007.

[3] 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.

[4] P. Lundén, “Uni-verse acoustic simulation system: interactive real-time room acoustic simulation in dynamic 3d environments,” The Journal of the Acoustical Society of America, vol. 123(5), pp. 3937–3937, 2008.

[5] 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, 2007.

[6] G. P. M. Egelmeers and P. C. W. Sommen, “A new method for efficient convolution in frequency domain by nonuniform partitioning for adaptive filtering,” IEEE Transactions on signal processing, vol. 44, 1996.

[7] W. G. Gardner, “Efficient convolution without input-output delay,” Journal of the Audio Engineering Society, vol. 43, pp. 127–136, 1995.