Immersive audio rendering for interactive virtual architectural environments

Article *in* Journal of the Audio Engineering Society · September 2016

2 authors:

Muhammad Imran
RWTH Aachen University

Jin Yong Jeon
Hanyang University

Audio Engineering Society

Conference Paper

Presented at the Conference on
Audio for Virtual and Augmented Reality
2016 September 30–October 1, Los Angeles, CA, USA

Immersive audio rendering for interactive virtual architectural environments

Muhammad Imran[1], Jin Yong Jeon[1]

[1] *Department of Architectural Engineering, Hanyang University, Seoul, Korea*

Correspondence should be addressed to Jin Yong Jeon (jyjeon@hanyang.ac.kr)

ABSTRACT

In this study we investigate sound propagation methods in virtual architectural environments for spatialized audio rendering to use in immersive virtual reality (VR) scenarios. During the last few decades, sound propagation models have been designed and investigated for virtual architectural structures, using geometrical approach (GA) and hybrid techniques. For sound propagation, it is required to design fast simulation tools to incorporate a sufficient number of dynamically moving sound sources and receivers, room acoustical properties, and reflections and diffractions from interactively changing surface elements in VR environments. Using physically based models, we achieve a reasonable trade-off between sound quality and system performance. Furthermore, we describe the sound rendering pipeline into a virtual scene in order to simulate virtual scenarios.

1 Introduction

Recent investigations and studies in virtual reality (VR) technology have played a vital role in science and technology and become focus of interest, as a powerful tool, for numerous potential usages, encompassing different research areas (e.g., in architecture designing, spatial audio for virtual environments and games). In addition, the applications of sound rendering are extended towards representing the complex architectural structures in virtual reality environments for its accurate acoustical perception within the dynamically varying building scenes. However, the modern VR-systems are only equipped with and limited to visual rendering systems. Occasionally, these systems incorporate the vital role of room acoustical characteristics that explains the behaviours and the responses of architectural structural elements in VR systems and are always complementary to visual cues. Moreover, there are certain discrepancies in synchronization of audio-video during rendering process that create disparity among different audio and visual cues and hence produces a poor virtual perception [1]. In order to emulate a truly immersive VR environment, a high-tech systems is required to imitate the virtual environments as realistic as its corresponding real environments to maximize the user's feeling of multimodal immersion and presence, and enable them to act interactively with their surrounding spaces.

Recent, up-to-date auralization methods for virtual environments, that describe the sound propagation in enclosed spaces, commonly use geometrical acoustics (GA) and are mostly used for achieving a better quality in synthesis of aural stimuli, while incorporating realistic acoustic behaviour of the structures [2]. A recently introduced approach that achieves convincing results is modelled by combining deterministic methods for the computation of early specular sound reflections with stochastic approaches for the computation of the reverberant sound field [3]. In this hybrid approach, acceleration algorithms from computer graphics make the VR system capable to manage the computational cost of dynamically moving sound sources and receivers in real-time in complex architectural scenarios [4].

In this paper, we propose an efficient pipeline for the implementation of a room acoustics simulation for immersive VR system. These simulation techniques enable us to manipulate realistic auralization of sound propagation in building structures for VR environments. Furthermore, certain wave phenomenon; i.e. sound scattering, airborne sound transmission between building elements and sound diffraction are considered. In one hand, the spatial distribution of sound energies and dynamic movement of sound sources as well as receiver are supported at runtime, while, on the other hand modifications and manipulations of the environment are updated during the real time. To achieve the spatial audio rendering of rooms; (i.e. offices, small rooms) wave-based techniques are suggested to resolve certain frequency responses that produces modal structures. As a high-tech spatial sound rendering model a combination of finite difference time domain method (FDTD) or finite element method (FEM) with geometrics approach (GA) is proposed to serve well in this respect using the information of the room boundaries and sound diffraction to achieve improvements in accuracy of sound rendering. The performance of these auralization models is evaluated for the accuracy of simulations.

2 Auralization

There are three stages involved in an accurate auralization process for building structures in graphical environments; i.e. sound generation, its propagation through building spaces (doors, windows, corridors etc.), its transmission through building structural elements (i.e. walls, partitioning) and finally computation of transfer functions for sound rendering using fast convolution methods. The perceptual authenticity of whole auralization chain critically depends on the accuracy of computing transfer function between source and receiver that might be located anywhere in the building elements. Figure 1 shows the auralization processing chain and further explanation is given in following paragraphs.

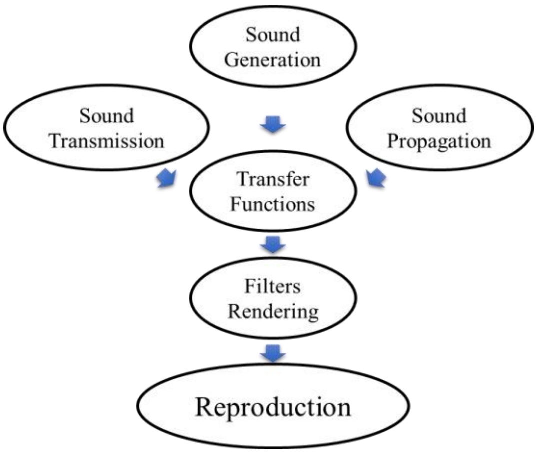

Figure 1. Auralization processing chain (A conceptual flow).

For sound generation, a mono dry sound signal is desired that can be recorded in anechoic chamber (ideally having zero reverberation time). An effective approach proposed for recording such sound source is given in [2]. The sound source is placed at a particular distance and direction from the recording device. The source directivity patterns are very important to be considered and can be integrated into recordings through simulation algorithm. The next process is to compute all the possible sound transmission paths between source and receiver positions and to predict the sound propagation through the spaces between the rooms. The resulting signals are termed as transfer functions (i.e. impulse responses (IRs)) between the rooms and are subjected to convolution process for rendering the sound sources. The convolution process is normally preferred in frequency domain to make the process capable for real time rendering applications. There are many convolution methods available. The commonly used method is discrete convolution with overlap-add (OAM) or overlap-save (OSM) methods. In the final process of sound reproduction, binauralization is one of the effective method for adding 3D effects to the processed sound source in order to achieve spatial impression according to the source and receiver position. The binaural transfer functions, known as head related transfer functions (HRTFs) in frequency domain and head-related room impulse responses (HRIRs) in time domain, are set of convolution filters. These transfer functions are computed by using dummy head recording of the impulse responses at different angles in an anechoic chamber hall. A large variety of HRTFs data-base is available, however, these HRTFs are person specific [5]. Therefore, the use of personalized HRTFs are proposed for an accurate sound rendering for virtual reality systems to achieve immersive 3D audio. The commonly used approaches are based on either measurements or numerical modelling of individual's head and torus.

3 Related work and background

In the past few decades, most of the work has been done on propagation and auralization of airborne sound. Numerical methods precisely model the sound waves propagation phenomenon by numerically solving the wave equation [6]. These methods are pretty accurate, however, they offers high computational cost. Existing approaches using numerical methods (i.e. finite element methods (FEM), boundary element methods (BEM) and finite difference time domain methods (FDTD)) reasonably predict the sound transmission paths in building elements for general purpose auralization and are limited to simple scenes, however, quite slow for modelling interactive sound transmission in complex buildings. On the other hand, geometric approaches use sound models based on rectilinear propagation of waves and can accurately model the early reflections. Furthermore, these methods were improved based on interactive ray tracing techniques that are reasonably applicable to dynamically varying scenario [7]. The image source methods synthesize virtual sound sources to compute specular reflections and further used to combine the early reflections with diffusive reflections. These methods precisely calculate the propagation paths from the sound source to the receiver considering certain assumption in boundary conditions of surface elements. Geometric methods also use beam tracing techniques that recursively traces pyramidal polyhedral from the source to the listener [8] and can be useful for modelling sound propagation at runtime in complex environments for moving sources.

In addition, each method faces certain limitations in terms of time, frequency and perceptual domains that a single method is unable to resolve all constraints in all domains. A combination of the FDTD, FEM and/or GA is very well applicable as far as the source-receiver interactivity and perceptual accuracy for auralization is concerned. However, existing techniques require huge pre-processing time and therefore are restricted to static scenes with stationary sources and listeners and are more concerned with the airborne sound propagation phenomenon rather than deeply considering structural borne transmission of sound through structural elements. Moreover, further investigations regarding the boundary and source representation, and the phenomenon of accurate structural born sound are necessary to improve the accuracy. These shortcomings are a common problem in all acoustic simulation methods based on existing models [9].

4 Computational framework

In this paper we designed an auralization pipeline for complex building structures as shown in Figure 2. The pipeline is based on hybrid methods using geometric acoustics (GA) proposed by many researchers during their recent investigations. The processing chain computes the transfer functions for sound propagation and transmission through the building elements between source and receiver, located in different rooms. These transfer functions are then convolved with the sound source and finally reproduction block is activated for synthesizing binaural signals. These binaural signal can be presented through headphone or loudspeakers. In case of loudspeakers the additional set of filters are required, termed as crosstalk cancellation filters. Then the visual cues are integrated with the synthesized audio in VR environment to present the complete auralization chain.

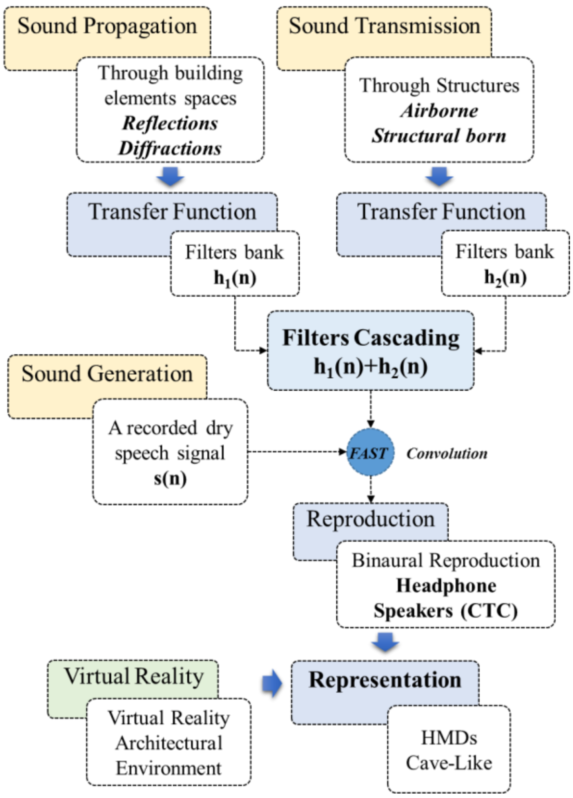

Figure 2. Auralization processing chain and corresponding signal processing.

To achieve an interactive auralization of a virtual scene, the prerequisite are; development of efficient techniques for estimating sound transmission paths through structures, modelling remotely coupled rooms, scene representation, handling interactive changes in structure and corresponding filters rendering and updating, and finally real time implementation of whole processing chain within the given time constraints using fast convolution techniques. In this paper, our computational framework adopts a hybrid method of room acoustics simulations of complex architectural environments, based on Image Source Method (ISM) and Ray Tracing (RT) algorithm. This hybrid method use implemented by [10] and enables a physically based computations of room impulse responses (RIRs) in real time. For synthesis of RIRs many factors are incorporated during simulations, such as; specular and diffusive sound scattering, sound transmission and edge diffraction phenomenon. The sound transmission for complex structure requires.

For the real-time auralization of virtual environments a software package named “Room Acoustics for Virtual ENviornmanets” (RAVEN) [10] is developed by “Institute of Technical Acoustics” (ITA), RWTH Aachen University and provides a good quality room acoustics simulations. For sound transmission certain real-time filtering state-of-the-art techniques i.e., “non-uniformly partitioned frequency-domain convolution (NUPFDC)” have been proposed [11] and modifications for its improvement are recommended reporting that this technique offers very low latency [12]. Wefers in [13] proposed a high-fidelity auralization procedure using room impulse responses (RIRs) and computing clusters to achieve a trade-off between algorithmic complexity and sound quality. In this method, authors claim that their approach is independent of any precomputation requirements and performs in interactive mode with exchange of filters during runtime. A challenging task of exchange of filters during runtime is achieved by splitting the binaural impulse responses (BRIRs) into an early reflections part and a reverberation part. We have adopted the RAVEN software methods for computing the Transfer function between sources and receivers. One of the most important runtime phenomenon is interactivity during sound rendering and this phenomenon is also incorporated into rendering process for the true representation and perception of architectural environments in VR systems.

5 Simulations

To validate the computational work flow, we have organized the room acoustical simulations in the following steps. At the first step we have constructed a computer model of an example architectural structure and its corresponding polygon representation as show in the Figure 3a and 3b respectively.

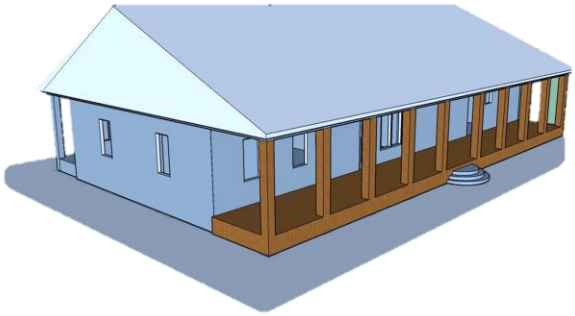

Figure 3a. An example architectural Structure.

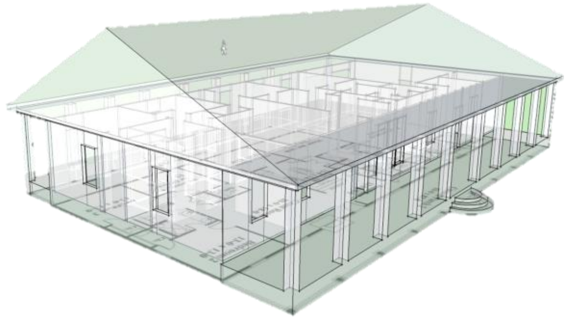

Figure 3b. Polygon representation (wire frame) of example architectural structure.

Secondly, the scene decomposition is performed to compute the sound propagation paths for computations of the corresponding transfer functions between source and listener. As a further step, hybrid method (combination of ISM and RT) is used for computing the impulse response at the location of listener. One source S and two receivers (R1 and R2) are located in different room, separated by structural elements i.e. doors and walls. The binaural impulse responses were synthesized for listener positions (R1 and R2) as shown in Figure 4. The results for receiver R1 are shown in Figure 5 and corresponding room acoustical parameters in Table 1.

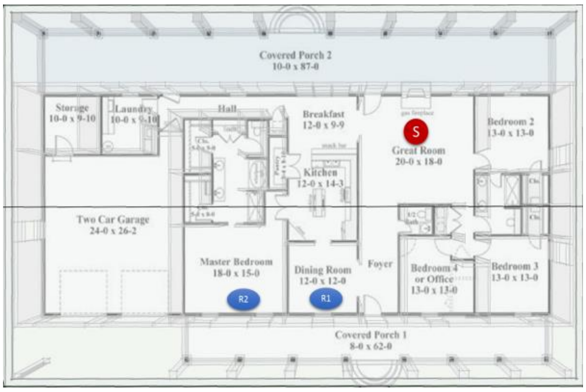

Figure 4. The location of source and the two receivers placed in different positions inside the example architectural structure.

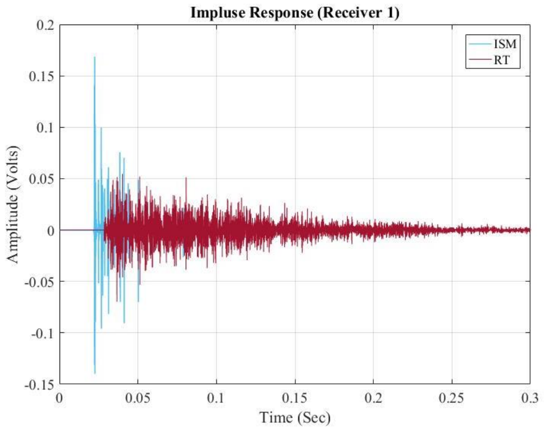

Figure 5a. RIR computed using ISM and RT methods for receiver R1.

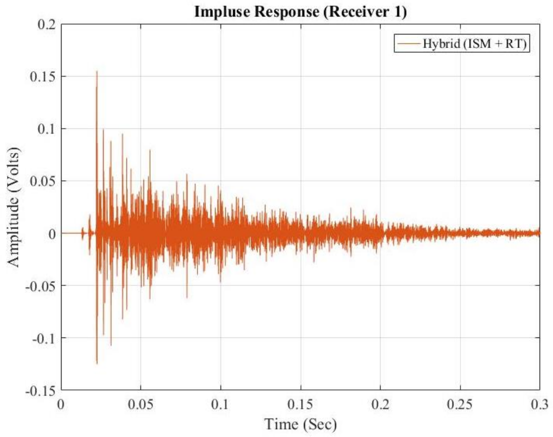

Figure 5b. RIR computed using hybrid (ISM + RT) method for receiver R1.

Parameter		**Receiver R1**
Reverberation Time (sec)	T60	0.81
	T30	0.76
	EDT	0.83
Gain (dB)	G	18.71
Definition (%)	D50	60
	D80	69
Clarity (dB)	C50	1.3
	C80	4.27
Central Time (sec)	Ts	0.072

Table 1. Acoustical Parameters for Receiver R1.

6 Summary and future work

This paper has described and investigated the techniques and methodology for sound propagation in virtual architectural environments. A hybrid technique based on combination of ISM and RT was adopted for generating room impulse responses considering the reflection and sound scattering phenomenon in multi-coupled rooms. The sound propagation paths were computed between the listener and the speaker, located in different rooms, for example architectural structure. The results are presented to show the computation of RIR from ISM, RT and combination of both and finally binaural reproduction was employed for auralization process.

The interactivity during the auralization process is achieved by updating the filter bank during the user's movement during the runtime. However, there are still require certain methodologies to increase the performance

Further, as described the wave base solutions produce more realistic result as compared to ray tracing, therefore, in further work wave based solution will be adopted. The filter rendering and fast convolution still require to be upgraded and full interactivity will be considered, such as; opening and closing of door in runtime, placement and movement of room components (chairs, tables etc.) and last but not least the surface texture of walls would be considered.

Acknowledgements

This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government (MSIP) (2013M3A6A3079356).

The authors are thankful to Institute of technical Acoustics (ITA) RWTH Aachen University Germany especially, "Virtual Acoustics Group ITA" for arranging a couple of visits to their Cave-like virtual reality environment during a couple of years. These visits inspired the authors to formulate the applications "immersive spatial sound rendering for virtual architectural environment" for interactive video games technology.

References

[1] M. Imran and J. Y. Jeon, "Feature Based Impact Sound Synthesis of Rigid Bodies Using Linear Modal Analysis for Virtual Reality Applications," in *Audio Engineering Society Conference: 61st International Conference: Audio for Games*, 2016.

[2] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*: Springer, 2008.

[3] M. Vorländer, D. Schröder, S. Pelzer, and F. Wefers, "Virtual reality for architectural acoustics," *Journal of Building Performance s*

[4] D. Schröder, *Physically based real-time auralization of interactive virtual environments* vol. 11: Logos Verlag Berlin GmbH, 2011.

[5] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models* vol. 22: Springer Science & Business Media, 2013.
[6] K. S. Kunz and R. J. Luebbers, "The finite difference time domain method for electromagnetism," 1993.
[7] I. Wald, S. Boulos, and P. Shirley, "Ray tracing deformable scenes using dynamic bounding volume hierarchies," *ACM Transactions on Graphics (TOG),* vol. 26, p. 6, 2007.
[8] I. Drumm, "The development and application of an adaptive beam tracing algorithm to predict the acoustics of auditoria," University of Salford, 1998.
[9] M. Vorländer, S. Pelzer, and F. Wefers, "Virtual Room Acoustics," in *Sound-Perception-Performance*, ed: Springer, 2013, pp. 219-242.
[10] D. Schröder and M. Vorländer, "RAVEN: A real-time framework for the auralization of interactive virtual environments," in *Forum Acusticum*, 2011.
[11] G. P. Egelmeers and P. C. Sommen, "A new method for efficient convolution in frequency domain by nonuniform partitioning for adaptive filtering," *IEEE Transactions on signal processing,* vol. 44, pp. 3123-3129, 1996.
[12] W. G. Gardner, "Efficient convolution without input/output delay," in *Audio Engineering Society Convention 97*, 1994.
[13] F. Wefers and D. Schröder, "Real-time auralization of coupled rooms," in *Proceedings of the EAA auralization symposium, Espoo, Finland*, 2009.