INTERNATIONAL SYMPOSIUM on MUSICAL and ROOM ACOUSTICS
ISMRA 2016
September 11-13, 2016
La Plata, Buenos Aires, Argentina

Simulation and Auralization of Concert Halls / Opera Houses: Paper ISMRA2016-70

Interactive multi-source sound propagation and auralization for dynamic scenes

Carl Schissler [(a)], Dinesh Manocha[(b)]

[(a)] University of North Carolina at Chapel Hill, USA, schissle@cs.unc.edu
[(b)] University of North Carolina at Chapel Hill, USA, dm@cs.unc.edu

Abstract

We present a sound propagation and auralization system designed for interactive simulation of complex dynamic environments with many sound sources. Our approach builds on previous work in geometric acoustics using ray tracing. A key component of our technique is the use of temporal coherence in the computation of the sound on each simulation update. Previous ray-tracing methods recompute the impulse response for each source and listener pair on every update. However, this can lead to auralization artifacts because the result at each time step is slightly different. We leverage this variation using a cache of results from previous time steps, including early reflection paths and late impulse responses, to improve the quality and performance of the simulation. We also apply a psychoacoustic metric based on the human threshold of hearing to the resulting impulse responses in order to determine how far to propagate rays on the next frame. This feedback mechanism allows dynamically-changing reverb times and both quiet and loud sound sources without expending computation for inaudible parts of the impulse response. To handle the case of large environments with tens or hundreds of sound sources, we introduce a method for clustering sources into representative clusters based on their relative visibility and distance from the listener. The result is that the number of simulated source and listener pairs is reduced. We demonstrate the results of our technique on a variety of indoor and outdoor benchmarks with up to 200 sound sources and show that it can achieve interactive performance on consumer computer hardware.

Keywords: interactive sound propagation ray tracing

INTERNATIONAL SYMPOSIUM on MUSICAL and ROOM ACOUSTICS
ISMRA 2016
September 11-13, 2016
La Plata, Buenos Aires, Argentina

Interactive multi-source sound propagation and auralization for dynamic scenes

1 Introduction

The simulation of sound propagation within virtual environments is an important topic that has many potential applications in the fields of acoustics and computer science. Much of the previous work in this area has focused on the prediction of acoustics in concert halls. Several commercial software products such as ODEON and CATT are widely used in the architectural acoustics community for this purpose. These programs take as input a 3D CAD model of the environment to be simulated, the surface boundary conditions (e.g. absorption/scattering coefficients), and the positions of the source(s) and listener(s). The result of a simulation is an impulse response (IR) for each source and listener pair that describes the acoustic transfer function of the environment. Once computed, the impulse response can be convolved with anechoic source audio to generate the audio heard at the listener's position. However, these systems are designed primarily for offline acoustic simulation and therefore cannot be used in applications that require real-time interaction with the user such as games and virtual reality. In those applications, dynamic objects in the environment such as doors, cars, and people can have a strong effect on the acoustics of virtual spaces.

In this work, we focus in particular on the challenging problem of dynamic interactive simulations of multi-source scenes. To maintain full interactivity in this case, the impulse response between every source and listener in the scene must be updated at a rate of at least 10Hz. In contrast, existing commercial sound propagation systems can take a few seconds or minutes to compute a single impulse response. If the sound can be updated interactively, games and virtual reality become more responsive. In addition, the workflow for architectural acoustics can be streamlined by allowing changes to the material properties and acoustic treatment in real time with auralization of the resulting impulse response(s).

To achieve the goal of interactive sound rendering of complex virtual environments, we present a sound propagation system that utilizes several recently published techniques in order to accelerate the sound-propagation and rendering computation. In particular, we use backward ray tracing from the listener [1], sound source clustering [1], temporal coherence [2], and an adaptive impulse response length [2]. We have evaluated this system on several complex benchmarks containing dynamic objects and tens of sound sources and notice an improvement of at least an order of magnitude over traditional sound propagation approaches. As a result, our system is able to meet interactive performance goals for a variety of complex scenes on consumer computer hardware.

2 Related work

The most prominent methods for computing sound propagation in virtual environments can be divided into two main categories: wave-based and geometric techniques. Wave-based sound

propagation is the most accurate and involves directly solving the Helmhotz wave equation using numerical techniques. These include the finite-difference time-domain method [3], the finite-element method [4], the boundary-element method [5], adaptive rectangular decomposition [6], and the equivalent source method [7]. However, these techniques scale very poorly when considering high frequencies or large simulation domains and so are not suitable for dynamic interactive applications. As a result, precomputation approaches are commonly used but are limited to static scenes and can require a large amount of memory.

On the other hand, geometric sound propagation makes the simplifying assumption that the wavelength of sound is much smaller than the size of the geometric primitives (e.g. triangles) in the environment. As a result, fast algorithms from computer graphics like ray tracing can be used to approximate sound transport. The most prominent approaches for handling specular reflections are the image source method [8], beam tracing [9] and frustum tracing [10]. Diffuse reflections are usually handled using Monte Carlo path tracing [10,11,12,13]. In order to model wave effects such as diffraction, additional algorithms such as the uniform theory of diffraction [14], or the slower but more accurate Biot-Tolstoy-Medwin (BTM) formulation [15] can be used. While they are much faster than wave-based methods, most existing geometric sound propagation approaches are limited to a few sound sources and a few reflection bounces when the computation is required to update at an interactive rate.

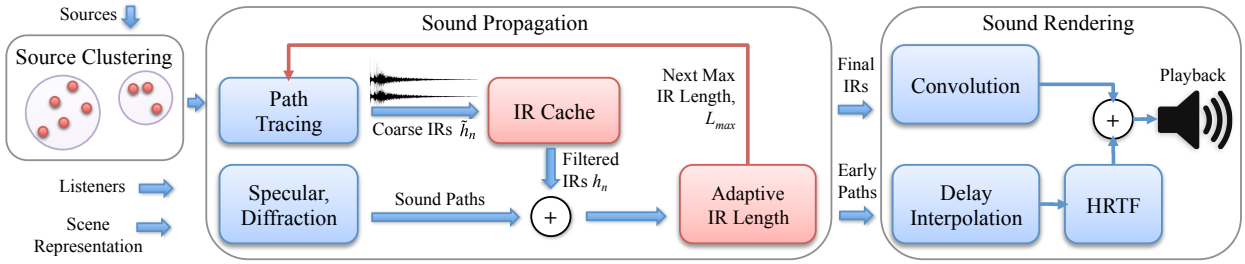


Figure 1: An overview of our auralization system showing the major components.

3 Overview

The main components of our sound propagation and rendering system are shown in Figure 1. As input, the system takes the source(s), listener(s), and a triangle mesh representing the scene geometry. Sources are first clustered to reduce the amount of computation in many-source scenes. The resulting clusters are provided as input to the sound propagation system, where backward path tracing from the listener is used to efficiently compute impulse response(s). These IRs are then filtered by the IR cache, then a perceptually-driven technique is applied to determine the IR length for the next frame. The output are the final IRs and a set of important discrete direct and early reflection paths. These are provided to the sound rendering subsystem which performs convolution of the anechoic source audio with the IRs, as well as fractional delay interpolation and then convolution with the users's HRTF for the early paths. The final

audio is reproduced for the user over headphones. The remainder of this section describes each of the imporant components of our system in more detail.

3.1 Multi-source sound propagation

Realistic auralization of complex environments often requires the simultaneous modelling of many sound sources. A busy city street may consist of hundreds of audible sources from car tires, engines, pedestrians, wildlife, and other ambient sounds. When applied to these scenarios, traditional geometric sound propagation techniques like forward path tracing [10,11,12,13] become slow because the computation time is linearly related to the number of sound sources. Here, we introduce two approaches that improve the performance in the multi-source case: backward path tracing from the listener and source clustering.

3.1.1 Backward sound propagation

In forward path tracing, a large number of rays are traced from each sound source and then propagated through the scene up to the IR length L_{max}. These are the *primary rays*. At each intersection of a primary ray with a surface, a ray is traced toward the listener's position to detect if the listener is visible to that point. If so, a path from the source to the listener has been found that is then added to the output impulse response. This is the so-called *diffuse rain* sampling approach [16], also known as *next-event estimation* in the field of computer graphics. When this approach is applied to multi-source scenes, the computation time increases quickly with the number of sources because a different set of primary rays must be traced for each sound source.

In order to improve the performance for multi-source scenes, we take advantage of the well-known principle of acoustic reciprocity that states that the sound heard at a listener from a source is the same as if the source and listener exchanged positions [17]. Rather than trace the primary rays from each source independently, we instead emit the primary rays from the listener's position [1]. As a result, the same primary rays are shared for all sound sources. The difference between forward and backward path tracing is illustrated in Figure 2.

Similar to the forward diffuse rain sampling technique, at every intersection of a primary ray with the scene a ray is traced toward each sound source to determine if the source is visible. If so, a path from the listener to the source has been found and can be added to the output IR. Since the primary rays represent a large portion of the cost for each additional sound source, this backward ray tracing provides a significant performance benefit. The only additional cost for each source is incurred by the diffuse rain sampling/next-event estimation that is performed at the primary ray intersections.

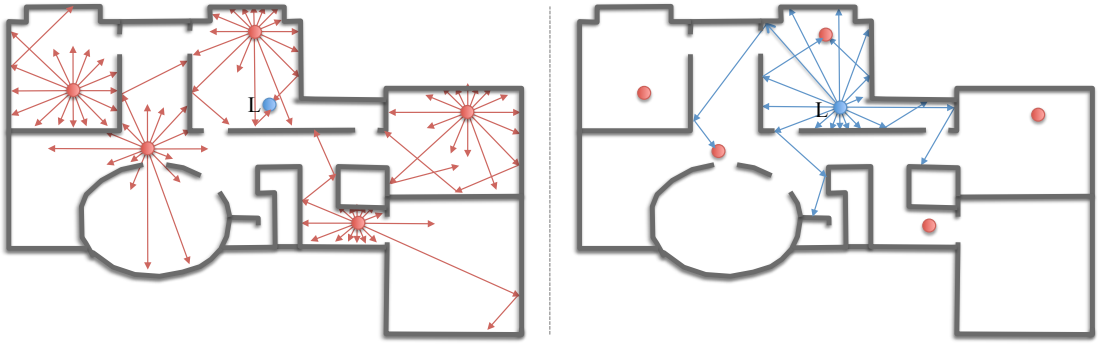


Figure 2: Forward vs. backward path tracing. Backward ray tracing significantly reduces the cost of tracing primary rays for multi-source scenes.

3.1.2 Source clustering

While this backward ray tracing approach is much faster than forward ray tracing, large scenes with tens or hundreds of sound sources may still be too slow to compute at an interactive rate. If interactive performance is required for these environments, one possibility to reduce the computation is to cluster certain sound sources together. Then, the clustered sources can be simulated as if they were a single larger source. If the clustering is performed in an intelligent manner, it is likely that the listener will be unable to detect the difference versus the full simulation.

We propose a source clustering technique [1] that uses the distance from the listener as well as the relative visibility between sources to determine when the sources should be clustered. The simulation domain is first segmented into non-overlapping grid cells of varying sizes, as shown in Figure 3.

Source: (Schissler, 2016 [1])

Figure 3: An example octree subdivision of a large tradeshow scene with 200 sound sources. Sources are considered for clustering when they are in the same octree cell.

This data structure is also known as an octree and is used to efficiently group sources into *potential* clusters. The size of grid cells is chosen to be smallest near the listener and then increases in proportion to the squared distance between the cell's center and the listener. At the beginning of each simulation update, the sources in the scene are inserted into the grid cells.

Then, the sources in each grid cell are evaluated to determine what clusters are present. At the start, all sources in a cell are marked as unclustered. A source is then picked at random from the unclustered sources and the other sources in the cell are tested to see if they could be clustered with the first source. For each other source, a ray is traced to the first source to see if the sources are mutually visible. If so, the sources are considered to be part of the same cluster. The mutual visibility requirement prevents the case where sources that are in different rooms but nearby in 3D space might be incorrectly clustered. This process repeats until all sources within a given grid cell are clustered or are treated as individual sources because they could not be clustered. The end result is a set of clusters that can be used as proxies for the sound sources in the scene. Since the number of clusters will usually be less than the number of sources, this provides a performance speedup that can enable the simulation of more complex scenes.

3.2 Impulse response cache

The runtime complexity of geometric sound propagation algorithms is dominated by the computation of IRs using ray tracing. Monte-Carlo path tracing methods are frequently used for diffuse reflections and account for a significant fraction of the total time spent in IR computation. The runtime complexity is a linear function of the number of rays being traced, and it is important to minimize the number of rays in order to achieve interactive performance.

One promising area of research in the field of interactive sound propagation is the use of persistent data structures that allow incremental computation of the sound field by exploiting temporal coherence. For many interactive applications, there is little change in the sound field over short timescales, and so these approaches take advantage of that property to reduce the computations per frame.

This idea has been applied to diffuse sound [13]. However, that approach does not work well for late reverberation (i.e. 50-100 orders of reflections) because the memory usage grows to the tens of gigabytes and the performance for interactive applications suffers as a result.

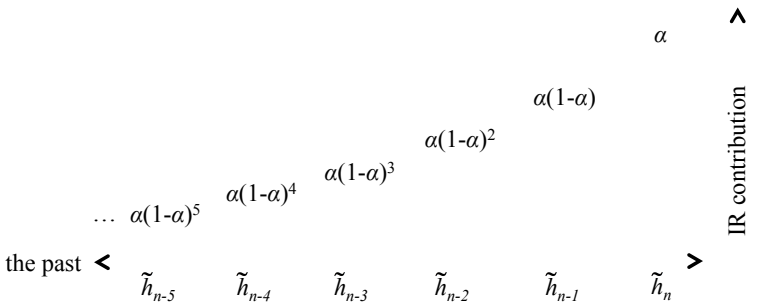
To address these issues, we introduce the notion of the impulse response cache [2], H_{n-1}, a copy of the previous frame's impulse response that is used to filter the output of a path-tracing based sound propagation algorithm. The IR cache H_{n-1} stores the accumulated weighted sum of past impulse responses, and so uses multiple frames of path tracing to compute the resulting final IR for frame n, h_n. During each frame, a different set of uniform random rays is traced, producing a slightly different impulse response. The weighted sum of many of these impulse responses is a better estimate of the actual sound field than the IR computed only for that frame alone, since it contains the contributions of many more sound paths.

Our IR caching module takes as input a coarse impulse response from path tracing, $\tilde{h}_n$, that contains the contributions from a small number of rays traced on the current frame for each

sound source. It produces a filtered impulse response utilizing the history information stored in the IR cache H_{n-1}. We introduce a parameter $\alpha \in [0,1]$ that controls how responsive the IR cache is to changes in the impulse responses. The jth impulse response sample on frame n is computed by the recursive relationship in equation (1).

$$h_n^j = H_n^j = \alpha \tilde{h}_n^j + (1-\alpha) H_{n-1}^j \quad (1)$$

The final IR sample h_n^j is a linear combination of the current frame's path tracing output, $\tilde{h}_n^j$, and the contents of the IR cache for the sample, H_{n-1}^j. In essence, the IR cache applies a 1st-order recursive low-pass filter to each sample in the impulse response, thereby using the history stored in H_{n-1} to produce a higher-quality impulse response with fewer undersampling artifacts. The parameter α determines the weight of the coarse IR in the final output IR. A value of α close to 1 means that the system is more responsive to dynamic changes in the scene, but also gains less benefit from the IR cache. A value of α closer to 0 indicates that more weight is given to the IR cache history stored in H_{n-1}, and thus the simulation will benefit more from the cache but be less responsive. The contribution of past IRs to the current frame's IR for α is illustrated by Figure 4.



Source: (Schissler, 2016 [2])

Figure 4: The contribution of the impulse responses from previous time steps decreases for those are further in the past.

3.3 Adaptive impulse response length

The length of the impulse response that is computed during ray tracing is an important parameter that influences the computation time required for sound propagation. Longer impulse responses tend to require more ray reflections and therefore are more expensive to compute, while if an impulse response is too short it can abruptly truncate the reverberation decay. To achieve the best combination of quality and performance the IR length should be chosen such that only audible parts of the IR are computed. Rays that are traced past the audible end of the IR are wasted when the sound is reproduced for a human listener. Existing sound propagation systems usually require the user to enter a fixed IR length for the entire simulation. However, it is difficult in practice to know a priori how long the IR should be for an arbitrary environment. Many factors determine the length of the IR including the power of the sound sources, the

surface acoustic material properties, the size/shape of the environment, air absorption, and the presence of dynamic objects like doors. Some models such as the Sabine and Eyring reverb equations [18,19] can be used to predict the reverberation decay rate of rectangular rooms, but they do not generalize well to more complex geometries. Therefore, the IR length cannot be easily predicted in general without running a simulation to determine the required length.

As a result, we propose a technique for interactive sound propagation that adaptively determines the IR length based on the simulation results from the previous time step [2]. Since we are only interested in the audible parts of the IR, we use the absolute threshold of hearing for an average adult human to determine the required IR length. An analytical expression for the threshold $T_q(f)$ as function of frequency can be computed using an empirical model [20].

$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}(f/1000)^4. \text{ (dB SPL)} \quad (2)$$

The threshold for each sound propagation frequency band is computed and then applied to the IR that was computed on the current time step. Starting at the end of the IR, we find the last sample in the IR that is above the threshold $T_q(f)$ for band center frequency f. The delay time of this sample within the IR is the *perceptual IR length*, L_p. This information about the length of the IR is computed for each sound source and then stored so that it can be used on the next update time step. On the next update, it would not be necessary to propagate rays further than a delay time of L_p. However, this simple approach does not allow the length of the IR to grow, only shrink. To rectify this problem, we always trace rays slightly farther, up to $L_{max} = L_p + \Delta L$, where ΔL is the most that the IR can grow/shrink on each time step. In our implementation, ΔL is chosen to be proportional to the simulation time step Δt.

This approach allows the length of the IR to adapt to the current configuration of the virtual environment. For example, if the listener moves from a room with short reverberation time to one with a longer reverberation time, the IR length will automatically increase over several frames until the length is suitable for the new room. This technique also enables the efficient handling of sound sources of high dynamic range within the same simulation. Distant or quiet sound sources automatically use less computation, while louder sound sources are ensured to produce a complete reverberation decay. In all cases, rays are traced only as far as necessary to compute the parts of the IR that are audible to a human listener. This provides a significant savings in computation time and reduces the parameters that must be tuned for a simulation.

Table 1: The main results of our auralization system for various benchmarks.

Scene	# Triangles	# Sources	# Clusters	Time (ms)
Cathedral	75,273	18	14	49.2
City	206,976	50	24	47.2
Hangar	71,461	18	13	72.6
Office	82,125	24	19	49.6
Space Station	35,581	21	15	75.0
Tradeshow	28,070	200	95	182.7

4 Results & conclusions

We evaluated the performance of our sound propagation system on six different benchmarks that contain a variety of scene types (indoor, outdoor), dynamic sources (cars, people), and dynamic objects (doors, cars). The scenes contain from 18 to 200 sound sources and have simulation domains that range from a small office to a very large city. The results for all scenes are presented in Table 1. A 4-core Intel i7 4770k CPU was used to measure the timings. For these results, 1,000 primary rays were traced from the listener's position on each frame. Overall, our system is able to achieve interactive performance (10Hz or better update rate) on all scenes except for the tradeshow. By clustering sources, the computation time is reduced by 30-50%, depending on the spatial arrangement of the sources relative to the listener.

When the sound quality of our approach is compared to a traditional forward path tracing approach without temporal coherence, about 20,000 primary rays are needed for the traditional algorithm to achieve the same level of subjective quality as our approach with just 1,000 rays. Our IR cache approach allows the path tracer to trace about 20x fewer rays and maintain the same level of simulation quality. Therefore our system can simulate much more complex environments interactively. The adaptive impulse response length reduces the amount of ray tracing by 20-40% because quiet sources have shorter IRs. In scenes like the hangar, office, and space station, it optimizes the ray tracing for dynamic changes in the reverberation time due to opening doors.

In conclusion, our system's combination of backward ray tracing [1], source clustering [1], the IR cache [2], and an adaptive IR length [2] provide a significant improvement over the previous state of the art in interactive sound propagation. For further details, results, and evaluation of each method, please refer to the original publications. Our system can simulate complex dynamic scenes with dozens of sources in real time on consumer computer hardware and has many applications in architectural acoustics, real-estate walkthroughs, games, and virtual reality. However, there is some room for improvement. Due to the limited diffraction modeling employed, our approach cannot accurately simulate all acoustic wave effects at low frequencies. In the future we would like to develop new techniques for integration of diffraction models into the Monte Carlo path-tracing framework.

Acknowledgments

This research was supported in part by the Link Foundation Fellowship in Advanced Simulation and Training, ARO Contracts W911NF-10-1-0506, W911NF-12-1-0430, W911NF-13-C-0037, and the National Science Foundation (NSF awards 0917040, 1320644, 1456299).

References

[1] Schissler, C.; Manocha, D. Interactive sound propagation and rendering for large multi-source scenes. *ACM Transactions on Graphics*, Vol. 36 (1), 2016.

[2] Schissler, C.; Manocha, D. Adaptive Impulse Response Modelling for Interactive Sound Propagation. *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2016, pp 71-78.

[3] Savioja, L. Real-time 3D finite-difference time-domain simulation of mid-frequency room acoustics. *13th International Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010, pp 77-84.

[4] Thompson, L. L. A review of finite-element methods for time-harmonic acoustics. *J. Acoustical Society of America*, Vol. 119 (3), 2006, pp 1315-1330.

[5] Gumerov, N. A.; Duraiswami, R. A broadband fast multipole accelerated boundary element method for the three-dimensional Helmholtz equation. *J. Acoustical Society of America*, Vol. 125 (1), 2009, pp 191-205.

[6] Raghuvanshi, N.; Narain, R.; Lin, M. C. Efficient and accurate sound propagation using adaptive rectangular decomposition. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15 (5), 2009, pp 789-801.

[7] Mehra, R.; Raghuvanshi, N.; Antani, L.; Chandak, A; Curtis, S.; Manocha, D. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics*, Vol. 32 (2), 2013, pp 19:1 – 19:13.

[8] Borish, J. Extension to the image model to arbitrary polyhedra. *J. Acoustical Society of America*, Vol. 75 (6), 1984, pp 1827-1836.

[9] Funkhouser, T.; Carlbom, I.; Elko, G.; Pingali, G.; Sondhi, M.; West, J. A beam tracing approach to acoustic modelling for interactive virtual environments. *Proceedings of ACM SIGGRAPH*, 1998, pp 21-32.

[10] Taylor, M.; Chandak, A.; Antani, L.; Manocha, D. RESound: interactive sound rendering for dynamic virtual environments. *Proceedings of the seventeenth ACM international conference on Multimedia*, Beijing, China, 2009. pp 271-280.

[11] Vorländer, M. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *J. Acoustical Society of America*, Vol. 86 (1), 1989, pp 172-178.

[12] Embrechts, J. J.; Broad spectrum diffusion model for room acoustics ray-tracing algorithms. *J. Acoustical Society of America*, Vol. 107 (4), 2000, pp 2068-2081.

[13] Schissler, C.; Mehra, R.; Manocha, D. High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *Proceedings of ACM SIGGRAPH*, Vol. 33 (4), 2014, pp 1-12.

[14] Tsingos, N.; Funkhouser, T.; Ngan, A.; Carlbom, I. Modeling acoustics in virtual environments using the uniform theory of diffraction. *Proceedings of ACM SIGGRAPH*, 2001, pp 545-552.

[15] Svensson, U. P.; Fred R. I.; Vanderkooy J. An analytic secondary source model of edge diffraction impulse responses. *J. Acoustical Society of America*, Vol. 106 (5), 1999, pp 2331-2344.

[16] Schröder, D. *Physically based real-time auralization of interactive virtual environments*, Logos Verlag, Berlin (DE), 2011.

[17] Case, K. Structural acoustics: A general form of reciprocity principles in acoustics. *The MITRE Corporation*, JSR-92-193. 1993.

[18] Schroeder, M. R. Natural sounding artificial reverberation. J. of the Audio Engineering Society. Vol. 10 (3), 1962, pp 219-223.

[19] Eyring, C. F. Reverberation time in “dead” rooms. *J. Acoustical Society of America*, Vol. 1 (2a), 1930, pp 217-241.

[20] Terhardt, E. Calculating virtual pitch. *Hearing research*, Vol. 1 (2), 1979, pp 155-182.