DOI: 10.61782/fa.2023.0401

MODULAR ROOM SIMULATION FOR THE IVES 3D ENGINE

Damian Dziwis[1,2*] **Tim Lübeck**[1,2] **Christoph Pörschmann**[1]

[1] Institute of Computer and Communication Technology,
TH Köln - University of Applied Sciences, Cologne, Germany
[2] Audio Communication Group, Technische Universität Berlin, Germany

ABSTRACT

IVES (Interactive Virtual Environment System) is a modular open-source 3D engine for creating virtual worlds in the Max programming environment. It is intended both for artists to create 3D audiovisual compositions and for researchers to use as a basis for prototyping new algorithms or experimental setups. IVES is based on Max's OpenGL implementation Jitter and IRCAM's spatial audio library Spat. This paper presents the latest development, a room simulation for IVES's parametric rendering. It consists of Spat's Feedback-Delay Network (FDN) for late reflections, accompanied by an Image-Source Model (ISM) module for early reflections and a generator for designing rooms. Here, the geometric parameters of the desired rooms are determined, and material properties are assigned to reflective surfaces. This results in parameters for reflective surfaces, filters, and attenuation for the ISM module and frequency band-specific RT_{60} times for the reverberation module using the FDN. The mirror image sources generated by the ISM are fed into the subsequent parametric rendering chain as additional sound sources, and late reverberation from the FDN reverberation module is added to the encoded sound field as spherical harmonic coefficients. We describe the underlying concept and technical implementation, and the possibilities resulting from the open and modular implementation.

Keywords: *spatial audio, virtual acoustics, interactive virtual environments, room simulation, parametric rendering*

1. INTRODUCTION

Creating and simulating convincing virtual environments requires not only the visual representation of 3D models, light, and shadows, but also the consideration of the acoustic environment [1]. The virtual acoustic environment (VAE) should match the visual environment to create audiovisual coherence [2]. To achieve this, the virtual sound sources must not only be rendered in the correct position and distance from the listener, but also in relation to their surroundings, particularly if it is an enclosed room. Here, the reflections of the sound sources are attenuated by objects and walls of different materials, and the geometry of the room provides different levels of late reverberation [3]. These acoustic properties lead to different acoustic perceptions of rooms, which creates a corresponding expectation when perceiving their virtual simulations. When simulating virtual environments in real-time, e.g. with arbitrary 6 degrees of freedom (6-DoF) for the user, depending on the target purpose and available computing resources, a balance between the physical accuracy of such a simulation and the required computational effort may be necessary. For real-world applications on the largest possible range of hardware, it may be reasonable to replace a physically accurate simulation with a perceptually convincing approximation model.
The modular 3D engine IVES (Interactive Virtual Environment System) is such an application for real-time virtual environments [4]. Intended for artistic work such as audiovisual composition, but also for research, e.g. for prototyping new algorithms, IVES provides an open toolkit for the Max development environment [5]. Max is

---

**Corresponding author*: *damian.dziwis@th-koeln.de.*

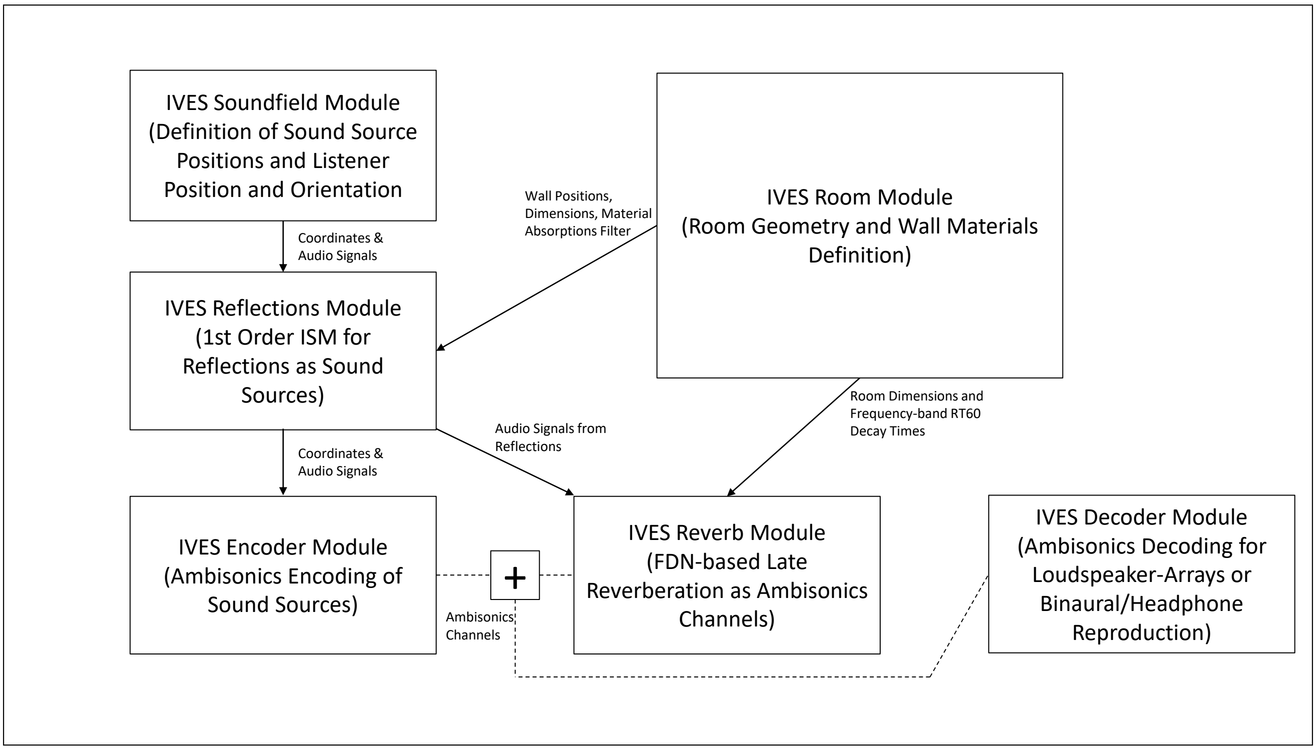

Figure 1. The IVES room simulation architecture with the ISM-based ”Reflections” module and the FDN-based ”Reverb” module together with the rest of the processing chain.

a visual programming language, domain-specific for signal processing, music and sound programming, and media arts. Inside Max, IVES allows the simple creation of virtual audiovisual environments with visual rendering modules based on Max’s OpenGL [6] implementation within the Jitter library. The results can be displayed on screens and projections, as well as in common virtual reality (VR) systems with head-mounted displays (HMDs) and 6-DoF head and room tracking. The spatial audio rendering modules were implemented using IRCAM’s Spat library [7,8]. Based on Higher-Order Ambisonics (HOA) [9], spatial audio can be reproduced on arbitrary loudspeaker arrays or headphones using binaural rendering. In addition to the existing anechoic spatial audio implementation in IVES, a parametric room simulation was implemented in the form of modules for a computationally efficient hybrid approach consisting of a module for early reflections using the Image-Source Method (ISM) and a module for late reverberation based on the Feedback Delay Network (FDN) algorithm in the Spat library [10].

In the following, we describe the basic concept of the hybrid room simulation approach and its implementation in the modular IVES engine. Furthermore, we give a brief evaluation of the system and discuss the resulting possibilities of the modular architecture.

2. BACKGROUND

The ISM [11] proposes a room simulation model in which the reflections of a sound source in a room are simulated as image sources. A sound source is mirrored on the room walls as an axis, and the position and distance of the mirrored image sources result in corresponding delays and attenuations with respect to the receiver. The attenuation is also affected by the material properties of the walls. The resulting image sources can in turn be mirrored on the room walls to model reflections of higher order. Image sources from higher-order reflections in arbitrary room geometries must be validated for visibility in the receiver path. In empty, symmetrical ”shoebox” geometries, this is not necessary, since the image sources always remain visible. However, even without validation checks, the com-

plexity and thus the computational cost increases significantly with each reflection order.
For computational efficiency, early, lower-order specular reflections from an ISM can be combined in a hybrid approach with another reverberation model for diffuse, late reflections. As shown by Wendt et al. [12], an FDN for diffuse late reflections can be a reasonable extension of an ISM for computationally efficient and perceptually plausible room simulations.
FDNs are a computationally efficient approach to creating reflections based on parallel connected and feed-backed delay lines combining comb and all-pass filters [13]. Further advancements allow multi-channel input as a feedback matrix [14] and control of the reverberation time depending on the frequency [10]. This allows to adapt the decay behavior to the desired room and to use the delayed, attenuated and material-filtered output from the ISM as input for the FDN [12].

3. IMPLEMENTATION

For the IVES modular 3D engine, a hybrid room simulation approach consisting of an ISM and an FDN-based module has been implemented (see Fig. 1). The spatial audio rendering in IVES is Ambisonics-based and has been implemented using Spat5 (v.5.3.0). The newly developed modules ”Reflections”, an ISM-based module for early reflections, and ”Reverb”, an FDN-based module for late reverberation, can be integrated accordingly into the rendering chain of Ambisonics-based en-/decoding. The modules also interact with the visual rendering in IVES, explicitly the ”Room” module, which allows the user to design a six-wall symmetric shoebox room, defining the position, dimension, and material properties of walls. The data about the position coordinates, dimensions, and material filters of the walls, as well as calculated reverberation decay times (RT_{60}, RT_{30}) of the room for specific frequency bands, are sent from the Room module to the Reflections and Reverb modules as Open Sound Control (OSC) messages.

3.1 Reflections

The Reflections module calculates first-order reflections using the ISM for each virtual sound source defined in the IVES Soundfield module. Based on the defined position and dimensions of the room, the positions of the mirrored image sources are calculated. The corresponding signal is processed with the filters according to the material properties of the walls.

Table 1. Available material properties in the IVES Room module.

Material filter:
neutral
acoustic-ceiling-tiles
brick-bare
brick-painted
concrete-block-coarse
concrete-block-painted
curtain-heavy
fiber-glass-insulation
glass-thin
glass-thick
linoleum-on-concrete
marble
parquet-on-concrete
plaster-rough
plaster-smooth
plywood-panel
polished-concrete-or-tile
sheet-rock
water-or-ice-surface
wood-ceiling
wood-panel

A JavaScript-based ISM has been developed for use with the Reflections module in conjunction with an established FDN object from Spat. An alternative could be the EVERTims [15] auralization engine implemented in Spat, which also provides an ISM in conjunction with an FDN, but is an unfinished prototype.
The ISM module calculates the positions of the image sources by mirroring the original sound source as a 3D point on each wall as a plane. As a compromise between computational efficiency and a perceptually sufficient degree of spatial details, only first-order reflections are computed [16, 17]. The positions of the image sources result in corresponding delays and attenuations of the original signal in relation to the position and distance to the receiver/listener.
The signal processing for material absorption is done with an infinite impulse response (IIR) biquad filter in second-order section (SOS) structure, using filter coefficients derived from material absorption coefficients of the Physikalisch-Technische Bundesanstalt (PTB) [18]. The first selection of suitable candidates of materials was inspired by the Resonance Audio Spatializer [19]. As a re-

sult, the materials available in the version at the time of this writing are as listed in Table 1. The direct sound signal and its position, together with the processed signals of the image sources and their positions are fed into the IVES Encoder module (using the "spat5.hoa.enconder∼" object) and encoded into Ambisonics signals. As in Wendt et al. [12], the early reflections from the ISM-based Reflections module serve as input to the FDN Reverb module to generate the late reverberation.

3.2 Reverb

The Reverb module has been developed using the FDN object "spat5.reverb∼" from the Spat library. It is based on the algorithms of Jean-Marc Jot [20] and generates late reverberation that can be parameterized with frequency-band specific (with arbitrary low, mid, and high-frequency bands) reverberation times. Band-specific cut-off frequencies were defined as follows: $f_{low} < 250Hz < f_{mid} < 5kHz < f_{high}$. The RT_{60} decay times for these frequency bands are calculated in the Room module based on the room geometry using Eyring's equation [21]

$$\mathrm{T}_{60}(f) = \frac{0.161V}{-ln(1 - \overline{\alpha}(f))S} \tag{1}$$

with the room volume V, the room surface S, and the average room absorptivity $\overline{\alpha}(f)$ for each frequency band. The frequency depended absorption coefficients are calculated based on the selected material properties of the walls, again the absorption coefficient information is derived from the PTB database [18].

The multi-channel input to the FDN consists of the processed signals from the image sources as early reflections calculated by the ISM. Using this input, a suitable late reverberation tail is generated according to the RT_{60} decay times calculated in the Room module. As the ISM calculates six reflections for each virtual sound source, six input channels per sound source are fed into the FDN. Although the channel number of the FDN for a single virtual sound source is thus below the recommended lower limit of 8 [22], this is accepted here for reasons of computational efficiency, so that the required channel number does not become too high with multiple sound sources.

Appropriate delay values for the FDN channels are calculated with

$$\tau_{j} = \frac{1}{c}(d_{i} + \overline{d}\epsilon_{j}) \tag{2}$$

as suggested by Wendt et al. [12]. Resulting in delay times τj for each channel. With the sound velocity c, the room dimension $d, i \in \{x, y, z\}$, and a random value ϵj between -0.1 and 0.1 multiplied by the mean of the room dimensions $\overline{d}$.

The output signals of the FDN are then encoded into HOA signal channels using the Spat5 encoder "spat5.pan∼" in "diffuse" mode as described in the Spat5 documentation. The HOA signals with the diffuse reverberation part can then be added to the signals with the direct sound and early reflections. Since it is a modular architecture, it is rendered independently of the other signals.

Figure 2. Image of the reference room and measurement set-up.

4. EVALUATION

Because the parametric rendering in IVES is implemented using established signal processing algorithms from the Spat library, most of the evaluation was focused on the parameters generated by the algorithms and modules for room simulation. These parameters (position coordinates, delay and attenuation values, room parameters, and decay times) were validated and traced along the rendering chain (encoder, FDN).

In order to technically evaluate the implemented hybrid room simulation and to analyze the obtained results in terms of just-noticeable differences (JNDs) in RT_{60} times and direct-to-reverberant ratio (DRR), a binaural room impulse response (BRIR) was measured. The used

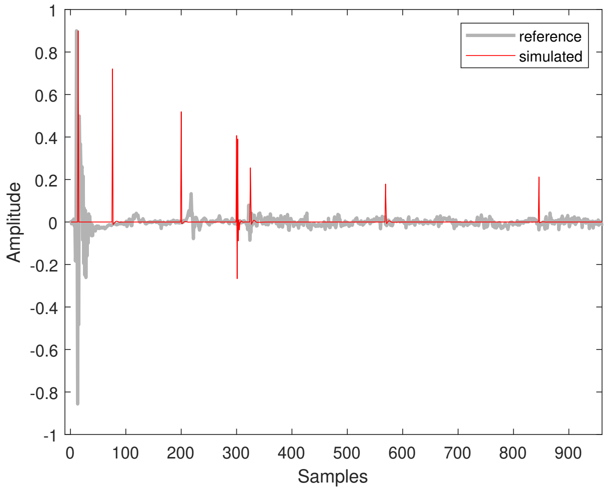

Figure 3. Simulated impulse response using the ISM, showing the reflection pattern compared to the omnidirectional measurement of the reference room.

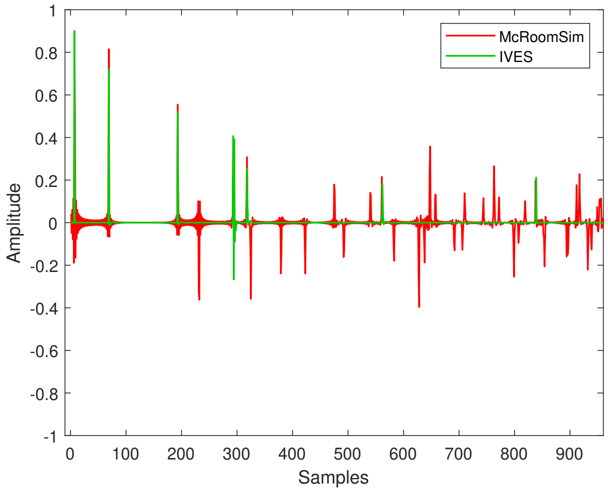

Figure 4. A comparison between a McRoomSim simulation using same room conditions and the simulation resulting from the IVES Reflections module.

rendering chain contained the IVES Soundfield module for generating virtual sound sources, the Reflections and Reverb module as the hybrid room simulation, the IVES Encoder for encoding the virtual sound field into 7th-order Ambisonics channels (ACN, SN3D), and the SPARTA ambiBin plug-in [23] with default settings (MagLS, Diffuse-field EQ, default HRIR set and MaxRE-Weights) for the binaural rendering of the Ambisonics signals (see Fig. 5). A real and measured reference room was approximately re-created virtually with the Room module for the simulation. The reference room has dimensions of 6.53 x 3 x 3.17 m (width×height×depth), and the receiver position was at a distance of 2.75 m to the left, 1.75 m to the back, and 1.29 m from the floor. Using the exponential sine sweep method, a BRIR measurement was made with a Neumann KU100 dummy head and an omnidirectional impulse response (IR) measurement with a Beyerdynamic MM1 microphone. The Genelec 8020 loudspeaker was placed at the same height and 1.84 m in front of the receiver position (see Fig. 2). This setup was recreated with the IVES Room module, with an approximation of the material properties of the walls using ”concrete-block-painted” for the floor, front and right wall, ”curtain-heavy” for the left and back wall, and ”brick-bare” for the ceiling.

For a successive evaluation of the IVES rendering, we determined the transfer functions of the system once after the Reflections module and once after the binauralization. Therefore, we employed Max's ”click∼” object and measured the impulse response. The direct sound and the first-order reflections from the Reflections module produce a reflection pattern with comparable delay times to the omnidirectional measurement (see Fig. 3), with expected deviations due to the differences in the only approximated wall materials and lack of source directivity of the loudspeaker. This is evident especially in the first two reflections in the mentioned plot. While in the simulation, the sound source's directivity is omnidirectional, a front-facing loudspeaker was used in the measurement. Accordingly, the first reflection from the wall directly behind the sound source has a much higher amplitude in the simulation compared to the real measurement. In the real measurement, the absorber behind the loudspeaker and the speaker's directivity result in higher attenuation and a low-pass behavior. The second reflection from the

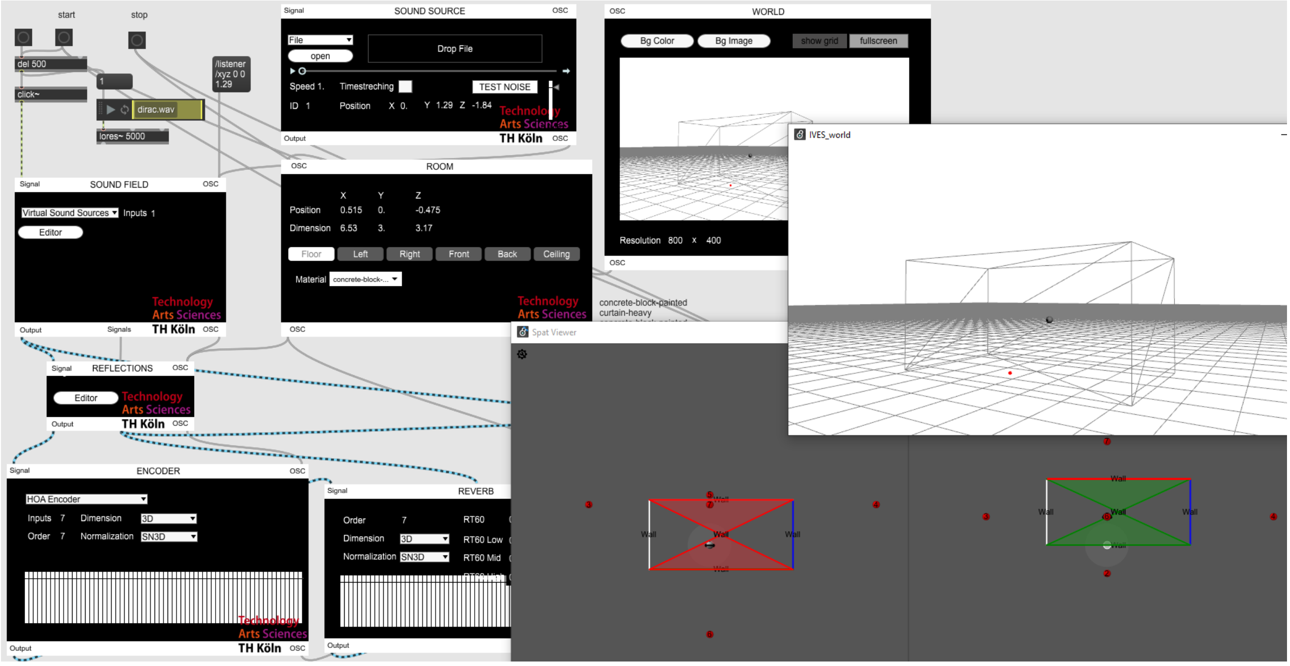

Figure 5. Screen capture showing an implementation of a virtual room simulation in IVES, including the Reflections and Reverb modules (as used for the evaluation).

floor is being additionally absorbed by a carpet in the real room, which is not considered in the simulation. The later reflections also show a corresponding deviation, due to the non-ideal shoebox room condition, e.g. caused by additional objects like chairs in the real room. In general, deviations can be explained by the very limited choice of wall materials, their different absorption behavior, measurement inaccuracies, and non-ideal conditions in real rooms. However, a comparison with a similar simulation of the room using the well-established software McRoomSim [24], shows matching first-order reflections with slight deviations in the attenuation due to a different technical realization of the material filters in McRoomSim (see Fig. 4).

The synthesized BRIR of the entire rendering chain, including the late reverberation from the FDN in the Reverb module, was compared to the measurement of the reference room. Mean RT_{30} decay times of the mid-frequency range from 315 Hz - 4 kHz were analyzed using the ITA toolbox [25]. In this range, the RT_{30} decay time of the simulated room is 0.24 s and approximates the RT_{30} time of 0.27 s of the reference room. This difference in reverberation time is about the JND range, with up to ~8% for T = 0.3 s according to Seraphim [26] and other studies also report a much higher JND (e.g. 24% in Blevins et al. [27]). The DRR was determined as the ratio of the direct sound energy (which we defined as the first 2.3 ms) to that of the remaining impulse response. The DRR of the simulated room at 2.1 dB is compared to the DRR of the reference room at 5.1 dB approximately within the JND (2-3 dB) for this category of rooms [28].

5. CONCLUSION & FUTURE WORK

With the described implementation of a hybrid room simulation based on the ISM for early reflections and an FDN for late reverberation, modules for the modular IVES 3D engine could be developed to simulate virtual rooms also acoustically. For the present reference room, a perceptually comparable auditory room could be simulated using the developed modules, as shown in Sec. 4. Since the simulation only allows a rough approximation of a real room, deviations, as described above, are to be expected. Apart from the dependency on ideal shoebox rooms, the material filters in particular make deviations visible. This is because in real rooms, in addition to other objects such as furniture, the walls themselves are often made of several different materials, making it difficult to approximate such compositions with the model presented here. Also, the absorption properties of similar materials can differ from each other. The reverberation properties (DRR and RT_{60}) thus depend strongly on the materials chosen. Therefore, the selection of suitable materials requires an appropriate consideration, but also the largest possible choice of materials. In order to provide a more accurate design of rooms in terms of wall materials, the number of materials is being successively expanded.

The modular architecture of IVES allows to separate the visual rendering from the spatial audio rendering to different devices and to reproduce direct sound and reflections separately from late reverberation on different output systems (e.g. different loudspeaker arrays and/or binaural for headphones). The open source code allows customization of the parameter determination algorithms, e.g. to implement deviating distance laws, reflection behavior for non-Euclidean geometries, and much more. In future developments, adjusting such parameters directly in the modules is aimed. Other planned developments include higher-order reflections, support for arbitrary, non-shoebox rooms, and optimization of the Eyring-based RT determination for critical room geometries [29].

The developed modules are part of the IVES engine and can be accessed from the repository: `https://github.com/AudioGroupCologne/IVES`

6. REFERENCES

[1] M. Vorländer, *Auralization : fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality; 1. ed.* RWTHedition, Berlin [u.a.]: Springer, 2008.

[2] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, pp. 1–6, IEEE, 2016.

[3] H. Kuttruff, *Room Acoustics*. CRC Press, 2016.

[4] D. Dziwis, J. M. Arend, T. Lübeck, and C. Pörschmann, "IVES - Interactive Virtual Environment System: a Modular Toolkit for 3D Audiovisual Composition in Max," *Proceedings of the Sound and Music Computing Conferences*, pp. 330–337, 2021.

[5] "Cycling'74 Max." `https://cycling74.com/`. Accessed: 2023-06-07.

[6] "OpenGL." `https://www.opengl.org/`. Accessed: 2023-06-07.

[7] T. Carpentier, "A new implementation of Spat in Max," in *Proceedings of the 15th Sound and Music Computing Conference: Sonic Crossings, SMC 2018*, no. Umr 9912, pp. 184–191, 2018.

[8] "Spat." `https://forum.ircam.fr/projects/detail/spat/`. Accessed: 2023-04-19.

[9] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording*. Springer, 2019.

[10] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Convention*, 1991.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[12] T. Wendt, S. Van de Par, and S. D. Ewert, "A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation," *Journal of the Audio Engineering Society*, vol. 62, no. 11, 2014.

[13] M. R. Schroeder, "Natural Sounding Artificial Reverberation," *Journal of the Audio Engineering Society Audio Eng. Soc*, vol. 10, no. 3, pp. 219–223, 1962.

[14] J. Stautner and M. Puckette, "Designing Multi-Channel Reverberators," *Computer Music Journal*, vol. 6, no. 1, p. 52, 1982.

[15] D. Poirier-Quinot, B. F. Katz, and M. Noisternig, "EVERTims: Open source framework for real-time auralization in architectural acoustics and virtual reality," *Proceedings of the International Conference on Digital Audio Effects, DAFx*, pp. 323–328, 2017.

[16] H. Steffens, S. van de Par, and S. D. Ewert, "The role of early and late reflections on perception of source orientation," *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2255–2269, 2021.

[17] S. Fichna, C. Kirsch, B. U. Seeber, and S. D. Ewert, "Perceptual evaluation of simulated and real acoustic scenes with different acoustic level of detail," in *Proceedings of the International Congress on Acoustics*, no. 12, 2022.

[18] Physikalisch-Technische Bundesanstalt (PTB), "Absorption Coefficient Database." `https://www.ptb.de/cms/ptb/fachabteilungen/abt1/fb-16/ag-163/absorption-coefficient-database.html`. Accessed: 2023-06-07.

[19] Resonance Audio, "Git Homepage." `https://resonance-audio.github.io/resonance-audio/`. Accessed: 2023-06-07.

[20] J. Jot, *Etude et Realisation d'un Spatialisateur De Sons Par Modeles Physiques et Perceptifs*. Lille thèses, 1992.

[21] C. F. Eyring, "REVERBERATION TIME IN "DEAD" ROOMS," *The Journal of the Acoustical Society of America*, vol. 1, pp. 217–241, 1930.

[22] J.-M. Jot, "Efficient models for reverberation and distance rendering in computer music and virtual audio reality," *International Computer Music Conference*, pp. 236–243, 1997.

[23] L. McCormack and A. Politis, "SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, 2019.

[24] A. Wabnitz, N. Epain, C. T. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics (ISRA)*, 2010.

[25] P. Dietrich, M. Guski, J. Klein, M. Müller-Trapet, M. Pollow, R. Scharrer, and M. Vorländer, "Measurements and room acoustic analysis with the ita-toolbox for matlab," in *40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, 2013.

[26] H. Seraphim, "Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen," *Acta Acustica united with Acustica*, vol. 8, no. 4, pp. 280–284, 1958.

[27] M. G. Blevins, A. T.Buck, Z. Peng, and L. M.Wang, "Quantifying the just noticeable difference of reverberation time with band-limited noise centered around 1000 Hz using a transformed up-down adaptive method," *International Symposium on Room Acoustics, ISRA*, vol. 1, no. June 9-11, pp. 1–6, 2013.

[28] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.

[29] K. Prawda, S. J. Schlecht, and V. Välimäki, "Calibrating the Sabine and Eyring formulas," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 1158–1169, 2022.