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 ISMRA2016-14

Augmented auralization: Complementing auralizations with immersive virtual reality technologies

David Poirier-Quinot, Barteld NJ Postma, Brian FG Katz

Audio Acoustics Group, LIMSI, CNRS, Université Paris-Saclay, Orsay, France
{first.lastname}@limsi.fr

Abstract

A framework conceived and designed to enable ecological assessments of theater, concert hall, and auditorium acoustics is presented. Coupling real-time convolution based auralization and 3D visualization with the inclusion of a 3D audio-visual (3D-AV) recorded performance provides for an animated visual anchor, allowing for appropriate orientation and distance variations perception, improving the sense of presence in the simulation. Auralizations are rendered either via an Ambisonic speaker array or binaural headphones using a Max/MSP audio engine. The 3D visualization is handled by BlenderVR, an open-source framework for Virtual Reality (VR). The framework is designed to facilitate multi-platform operation and seamless porting of the rendering over Head-Mounted Display (HMD), a portable one-projector CAVE system, or any other VR architecture. A use-case is presented with the historic Théâtre de l'Athénée in Paris. The acoustics from various seats are auralized using High-Order Ambisonic room impulse responses calculated using the Geometrical Acoustics (GA) software CATT-Acoustic on a GA model of the theater validated in a previous study. A 3D-AV theatrical performance is filmed with a depth sensor (Kinect 2), the resulting volumetric video is encrusted on the virtual stage allowing the same performance to be played in different room configurations without the need to construct and animate CGI avatars. Diverse application scenarios are discussed. The tools developed to support this framework are Open Source to promote research and development in augmented auralization.

Keywords: auralization, room acoustic, virtual reality, point-cloud

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

Augmented auralization: Complimenting auralizations with immersive virtual reality technologies

1 Introduction

Room acoustic simulations and auralizations are more and more accepted as a component of acoustic analysis and architectural design, obviously more so when the project targets acoustic-oriented buildings (concert halls, theaters, etc.). Kleiner et al. [1] defined auralization as: "*the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space*". Auralization makes the results of room acoustic design more tangible and accessible than descriptions of acoustic properties by abstract numerical quantities, making results accessible to all, from acousticians and architects to end-users.

The framework presented below is part of a research project employing real-time auralization on the assessment of room acoustic qualities by non-acousticians in an historical context. This paper focuses on the creation of CGI (Computer Graphic Image) to support auralization. The remaining of the Introduction will discuss the current state of auralization techniques and detail the recent works in integrating CGI to support room exploration in VR. The paper's organization is exposed along with the framework overview in Section 2.

Room acoustic assessment tools can roughly be sorted in two categories: physical models and computer simulations. Physical modeling was at first based on the use of high-frequency waves propagating in 2D sections of scaled models, coupled with Schlieren photography to observe local variations in wave front density [2]. A similar method used ripple tanks [3] where vibrating sources were used to create wavelets in shallow water to observe the acoustic properties of a given room geometry. Other phenomenological approaches were designed such as the use of light beams to simulate reflections behavior [4] or to investigate the distribution of acoustic energy [5], until the development of microphones accurate enough to enable scale model based techniques as used in modern acoustic design. Scale models allow for full 3D study of room acoustics [6, 7], scaling all physical dimensions of the room, including sound wavelengths. Techniques based on Room Impulse Response (RIR) recordings progressively replaced in-situ listening tests to assess rooms acoustical quality [8]. A continuous effort was since made at developing scale model techniques that would allow work with smaller scale factors [9].

The presented framework, as much of today's room acoustic studies, is based on computer simulation [1, 10]. The most frequent method consists in creating a Geometrical Acoustics (GA) model of the room, defining its dimensions and the materials it is composed of, then computing the RIR associated to sound propagation from emitter to listener throughout the space, and using this RIR for auralization [11]. Compared to scale model, computer simulation suffers no scale constraints and relies on numerical models, easily modified to assess architectural design proposals. Latest research and development has targeted navigable auralizations, allowing end-users to navigate in the model to assess the acoustics at different positions in the room [12]. Interactive auralization also received some attention [13], allowing end-users

to emit or trigger sound events to assess the acoustics as they navigate in the virtual room. The presented framework currently implements navigable auralization. Future developments (discussed in Section 6) will focus on the implementation of interactive auralization.

The addition of visual feedback to an auralization, even based on static images, has been proven to impact the final perception or *feel* of the room [14]. The closed feedback loop that exists between audio and visual information can be exploited to further improve the ecological validity of auralizations [15], producing results and sensations close to what the physical counterpart of the room being simulated would. The presented implementation proposes the addition of animated avatars of the sound sources (actors, performers, etc.), positioned in a virtual visual model of the room being assessed. The avatars are created from a video recording of a performance, avoiding the need for any lengthy 3D animation process while presenting a plausible and dynamic embodiment of the acoustic source.

In the remaining sections, the Théâtre de l'Athénée (Paris) will serve as a case-study of the presented framework. A short theatrical performance with two actors was filmed and recorded to serve the auralization and avatar creation. The paper organisation matches the framework's, as presented in the next section. It concludes with a discussion of how the framework will be used to support further research on the development of augmented auralization.

2 Framework overview

Figure 1 illustrates the augmented auralization framework, from the creation of a GA model to the real-time auralization of the VR scene.

The GA model of the theater was created and calibrated in CATT-Acoustic [16], as detailed in Section 3. Geometry and materials were defined based on blueprints and in-situ photos of the existing building. After acoustic calibration of the model, RIRs were simulated for a given source position and set of listener positions to enable end-user motion during the auralization.

Section 4 discusses how the visual model of the theater was added to the virtual scene. Initially designed in 3ds Max [17], said model was then exported to Blender [18] to be rendered in real-time on a VR architecture using the BlenderVR extension [19].

Finally, the virtual avatar creation method is detailed in Section 5. Actors were recorded with both microphones and a Kinect 2 sensor, the latter producing a coupled pair of RGB/Depth videos. These videos were used to generate a point-cloud of the actors in the rendered scene while the audio was convolved in real-time with the RIRs corresponding to the actual end-user listener position in the virtual environment for the auralization[1].

[1] Example videos of the virtual Théâtre de l'Athénée case-study, see `http://www.youtube.com/watch?v=arFU8yFe73Q` (single actor, single screen version) and `https://youtu.be/6hTfTvBH5WE` (two actor, three-screen version)

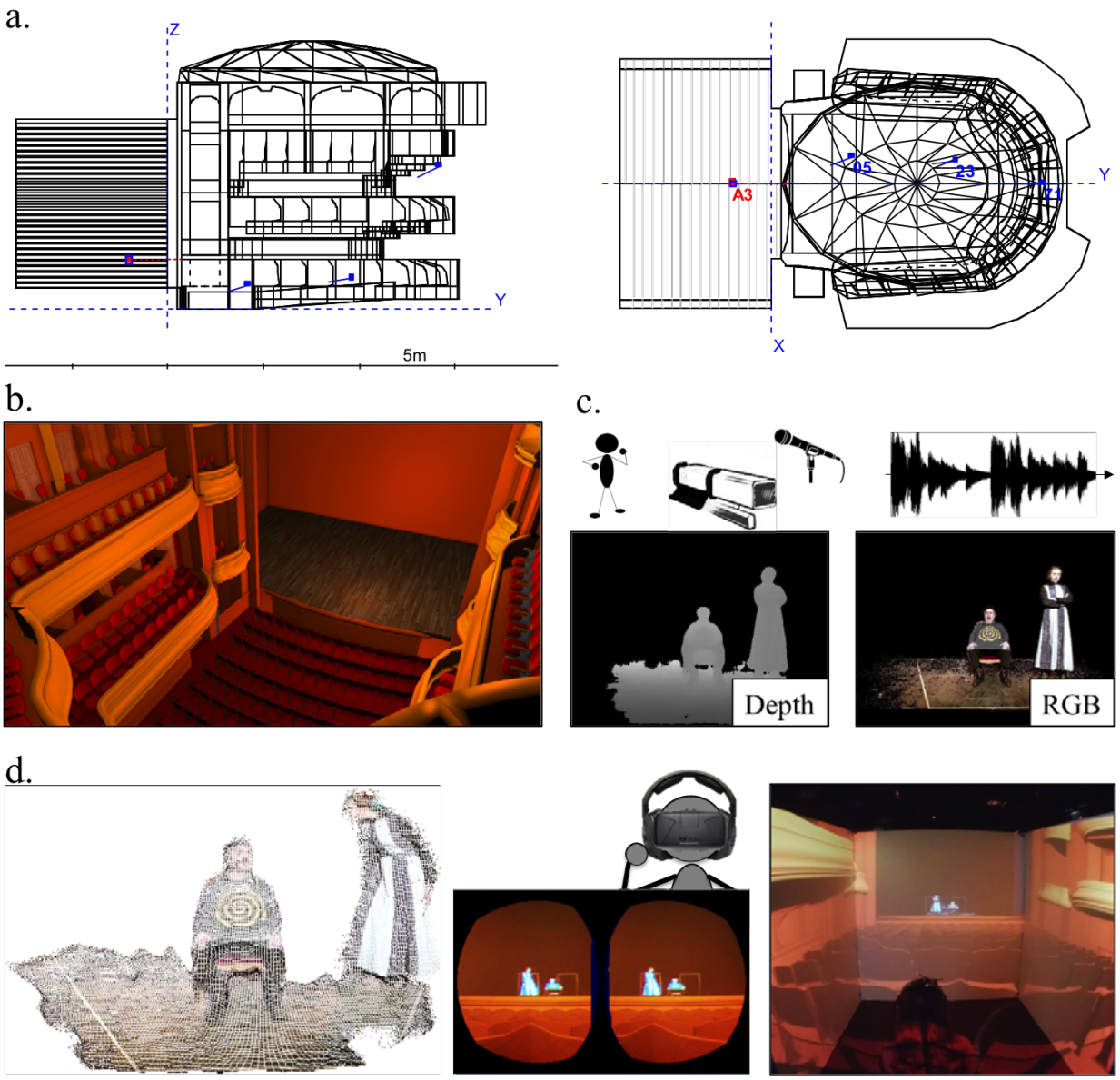


Figure 1: Conceptual overview of the Augmented Auralization framework. (a) Creation of the Théâtre de l'Athénée GA model and RIRs simulation for source-receiver positions. (b) Creation of the visual model. (c) Audio (dry) and Visual (RGB and Depth) recording of the performance. (d) Rendering the performers' avatar as a point-cloud, created from RGB and Depth recordings, which is integrated in the virtual environment for real-time augmented auralizations.

3 Room acoustic rendering

This section details the framework components related to the auralization: the creation and calibration of the GA model and the use of the exported RIRs for real-time convolution with

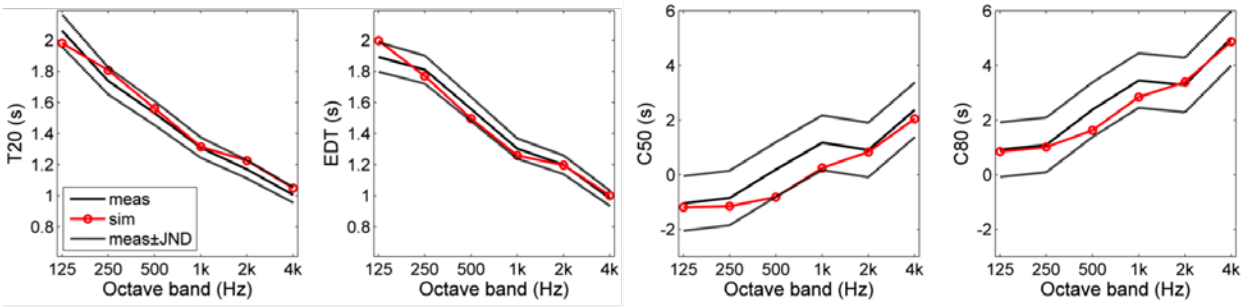


Figure 2: Comparison between simulated and measured means along with ±1 JND values for reference. From left to right: T20, EDT, C50, and C80.

anechoic/dry recordings of the performance.

3.1 Creation and calibration of the GA model, RIR simulation

The room acoustic model of the Théâtre de l'Athénée (see Figure 1a) was created using the GA software *CATT-Acoustic* (v.9.0.c, TUCT v1.1a) [16], composed of approx. 1,300 faces. Calibration was performed following the 7-step procedure presented in [20], including on-site measurements for calibration reference.

The absorption distribution in the Théâtre de l'Athénée is not uniform. As such, simulations were performed using CATT's "*Algorithm 2: Longer calculation, detailed auralization*" with 100,000 rays. Figure 2 presents a comparison of averaged measured acoustic parameters (T20, EDT, C50, and C80) with those estimated from the simulated RIRs. Simulated reverberation parameters EDT and T20 as well as clarity parameters C50 and C80 are within 1 Just Noticeable Difference (JND) of the measured values across all frequency bands for 2 omnidirectional source $\times$ 44 omnidirectional receiver configurations.

3.2 Real-time auralization

A Max/MSP patch was designed to handle the real-time auralization (see Figure 3) using the 3rd order Ambisonic (16 channels) RIRs from the calibrated GA model. The current RIR was selected according to the user position and convolved with the input audio stream (performance dry sound file). Decoding was based on either (1) a virtual speaker array for binaural rendering (see e.g. [21]) or (2) on a physical speaker array for Ambisonic rendering. When available, real-time head-tracking of user orientation was applied as a rotation to the resulting Ambisonic stream prior to binaural decoding. Current user position and orientation in the virtual environment, or more precisely virtual camera position and orientation, were determined in BlenderVR (see Section 4) and sent via OSC (Open Sound Control) protocol to the Max/MSP patch.

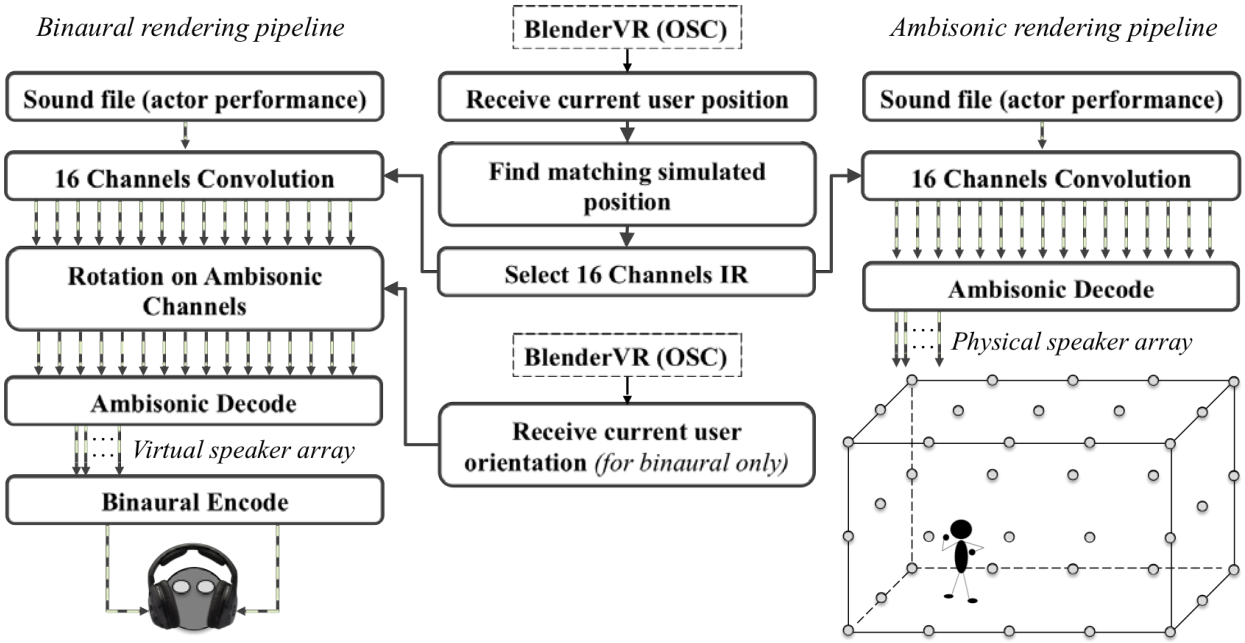


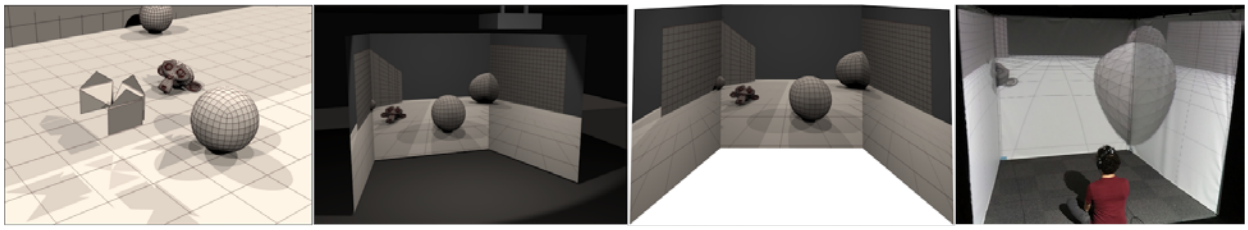
Figure 3: Conceptual schematic detail of the real-time auralization implemented in Max/MSP. Left and right hands of the figure respectively illustrate binaural and Ambisonic rendering pipelines.

4 Visual room rendering

This section details the creation of the visual model of the Théâtre de l'Athénée in Paris and its rendering on a VR architecture.

The initial mesh creation and texturing of the model was performed in 3dsMax, based on the materials collected for the creation of the GA model in CATT-Acoustic. It was then imported in Blender for real-time rendering in the Blender Game Engine. The whole scene was ported to BlenderVR [19] to be rendered on either a 3 screens video-wall architecture or an Oculus Rift DK2 HMD. Lighting was adapted to the targeted VR architecture, based on purely aesthetic considerations.

The HMD rendering made use of an Image Distortion GLSL shader and a user tracking Python script already integrated in BlenderVR. A lightweight system architecture was conceived to allow for the creation of a 3-wall CAVE-like system as illustrated in Figure 4. The objective was to project the virtual scene on the 3 screens using a single wide-angle lens projector, based on a technique similar to standard homography. A set of 3 virtual cameras was positioned in the theater scene, each camera rendering its image on a virtual screen in an intermediate layer "*projection*" scene. The relative positions and dimensions of the virtual screens in this scene matched those of the physical screens of the CAVE. A virtual camera sharing the extrinsic and intrinsic parameters of the physical projector rendered the image that the latter projected on the actual CAVE. Finally, head-tracking was accomplished using a set of OptiTrack infrared



a. b. c. d.

Figure 4: Illustration of the single-projector based rendering on the 3 screen CAVE. (a) Virtual scene with the set of 3 virtual cameras. (b) "Projection" scene with the 3 virtual screens and the virtual camera (cube-like shape on the top of the image), whose position, extrinsic, and intrinsic parameters match those of the screens and projector of the real world. (c) Pre-distorted image of the 3 screen CAVE architecture, as seen from the virtual camera of the "projection" scene, to be projected by the projector. Adaptive rendering is achieved by synchronizing the position of the 3 virtual cameras with the position of the user tracked in the CAVE. (d) Final projection on the 3 screen CAVE.

cameras to adapt the current rendered viewpoint to the user's actual position, providing a stable virtual environment (with respect to the real world).

5 Dynamic performance recording, point-cloud rendering in the VR world

A 5 min extract of the play *Ubu Roi*, by Alfred Jarry, was performed by two actors and recorded in the Théâtre de la Reine Blanche, a 140 seat theater in Paris, using two headset microphones and a Kinect 2 sensor. As the direct-to-reverberant ratio is high for close mic recordings, these were employed as approaching anechoic recordings. The video stream of the Kinect 2 sensor was handled by a script based on the *libfreenect2* library [22], recording current time stamp and both RGB and Depth images to disk. RGB and Depth videos were created from these images with a Matlab script checking for frame-per-second (fps) regularity of the image recording. Both videos were then combined during the real-time rendering in BlenderVR to produce a 512×424 point-cloud of the actors. The term point-cloud here refers to a GLSL texture rather than a 2D deformable mesh to reduce CPU consumption, projected in the VR world from a point in the virtual environment corresponding to the Kinect camera's position. The Depth video was used to define the spatial position of the point-cloud pixels, the RGB to define their color. The work of Pagliari et al. [23] was used to define the mapping between the hue of the Depth video gray-scale and the pixels depth position, along with the X/Y scaling coefficients of the 3D volumetric pyramid projection. The global scale of the point-cloud was defined to produce life-sized avatars in the VR scene's. Noise in the captured Depth video was removed frame-by-frame using filters for pepper-noise removal, forcing consistency amongst neighboring pixel values (*medfilt2*, Matlab Image Processing Toolbox).

6 Future work

To support future research on augmented auralization, further developments of the framework are planned, focusing primarily on real-time point-cloud capture and integration as well as performance voice directivity simulation.

Streaming the Kinect 2 Depth and RGB images directly into BlenderVR to generate the point-cloud in real-time will allow further diversification of applications of the presented framework. The live audio stream of the performers would likewise be used directly for the auralization. With the final framework, performers should be able to interact with the room space and acoustics in real-time, interacting with virtual avatars or other performers in different physical spaces, or to record themselves for latter assessment. Preliminary testing with multiple Kinect cameras further showed promise for the capture of more complex scenes and increased variety in possible viewpoints.

To reinforce the realism of the simulated room acoustic, voice directivity will be integrated in the framework, generated based on tracked user orientation and directivity patterns integrated into the simulated RIRs. This feature will support future studies on the impact of voice directivity on perceived performance and room characteristics. Based on these results, an investigation on the need of phoneme-specific directivity patterns in real-time auralization is planned.

7 Conclusions

This paper presented a framework to design real-time augmented auralization in VR environments, supporting the creation of immersive auralization experiences with little CGI skills and a low-cost VR architecture. The auralization itself is based on standard convolution with RIRs (in Max/MSP), simulated in a calibrated GA model (CATT-Acoustics), and rendered via either a headset (binaural) or a set of loudspeakers (Ambisonic). The creation of virtual avatars is described, based on recorded RGB/Depth videos of the sound sources involved in the auralization (e.g. actor performance). These avatars are rendered as point-clouds in the VR environment, completing the auralization with a lifelike visual anchor without the need for further CGI animation developments. To provide visual as well as audio immersion, the final scene is rendered on a set of 3 screens positioned around the user (U-shaped). A single wide-angle lens projector is used to display a pre-corrected image on the 3 screens around the user. Homography and adaptive rendering based on tracked user position are handled in the BlenderVR software (Blender extension for VR).

Acknowledgements

Part of the framework developments have been carried out by Dalai Felinto[2] and Martins Upitis[3] (Oculus DK2 integration in BlenderVR, point-cloud shader, homography implementation). We we would like to acknowledge the actors Philippe Fretun and Pascale Caemerbeke for their performance of *Ubu Roi*. This work was funded in part by the ANR-ECHO project (ANR-13-

[2] http://www.dalaifelinto.com
[3] http://devlog-martinsh.blogspot.fr

CULT-0004, echo-projet.limsi.fr).

References

[1] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," *J. Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[2] W. C. Sabine, "Theatre acoustics," *American Architect, Incorporated*, p. 104:257, 1913.

[3] A. H. Davis and G. W. C. Kaye, *The Acoustics of Buildings*. G. Bell and sons, ltd., 1927.

[4] V. O. Knudsen, *Architectural Acoustics*. Wiley, 1932.

[5] R. Vermeulen and J. de Boer, *Philips Techn. Review*, vol. 1. 1936.

[6] F. Spandöck, "Akustische modellversuche," *Annalen der Physik*, vol. 412, no. 4, pp. 345–360, 1934.

[7] B. F. Katz, Y. Jurkiewicz, T. Wulfrank, G. Parseihian, T. Scélo, and H. Marshall, "La Philharmonie de Paris - Acoustic scale model study," in *Intl. Conf. on Auditorium Acoustics*, vol. 37, (Paris), pp. 431–438, Institute of Acoustics, Oct. 2015.

[8] V. L. Jordan, *Elektroakustiske undersøgelser af materialer og modeller*. PhD thesis, Reitzels Forlag, 1941.

[9] M. Barron and C. Chinoy, "1:50 Scale acoustic models for objective testing of auditoria," *Applied Acoustics*, vol. 12, no. 5, pp. 361–375, 1979.

[10] M. Vorländer, *Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media, 2007.

[11] M. Kleiner, P. Svensson, and B.-I. Dalenbäck, "Auralization: Experiments in acoustical CAD," in *Audio Engineering Soc. Conv. 89*, vol. 38, pp. 874–895, 1990.

[12] B. N. Postma and B. F. G. Katz, "Acoustics of Notre-Dame Cathedral de Paris," in *Intl. Cong. on Acoustics (ICA)*, (Buenos Aires), 2016.

[13] L. Picinali, A. Afonso, M. Denis, and B. F. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *Intl. J. Human-Computer Studies*, vol. 72, no. 4, pp. 393–407, 2014.

[14] J. Y. Jeon, Y. H. Kim, S. Y. Kim, D. Cabrera, and J. Bassett, "The effects of visual input on the evaluation of the acoustics in an opera house," in *Forum Acusticum*, pp. 2285–2289, 2005.

[15] C. Suied, G. Drettakis, O. Warusfel, and I. Viaud-Delmon, "Auditory-visual virtual reality as a diagnostic and therapeutic tool for cynophobia," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 2, pp. 145–152, 2013.

[16] B. Dalenbäck, "CATT-Acoustic v9 powered by TUCT use manuals," *Computer Aided Theatre Technique, Gothenburg, Sweden*, 2011.

[17] 3ds Max: 3D modelization software. `http://www.autodesk.fr/products/3ds-max`.

[18] Blender: 3D creation software. `https://www.blender.org`.

[19] B. F. Katz, D. Q. Felinto, D. Touraine, D. Poirier-Quinot, and P. Bourdot, "BlenderVR: Open-source framework for interactive and immersive VR," in *IEEE VR Proceedings*, pp. 203–204, 2015.

[20] B. N. Postma and B. F. Katz, "Creation and calibration method of acoustical models for historic virtual reality auralizations," *Virtual Reality*, vol. 19, no. 3-4, pp. 161–180, 2015.

[21] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, "A 3D real time rendering engine for binaural sound reproduction," in *Proceedings of the 9th International Conference on Auditory Display (ICAD)*, pp. 107–110, 2003.

[22] Libfreenect2: Open source drivers for the Kinect for Windows v2 device, `https://github.com/OpenKinect/libfreenect2`.

[23] D. Pagliari and L. Pinto, "Calibration of Kinect for Xbox One and comparison between the two generations of Microsoft sensors," *Sensors*, vol. 15, no. 11, pp. 27569–27589, 2015.