Walk-through auralization framework for virtual reality environments powered by game engine architectures, Part II

Daniel Castro (1), André Verstappen (2), Samuel Platt (3)

(1) WGE / Stantec, Melbourne, Australia
(2) WGE / Stantec, Melbourne, Australia
(3) WGE / Stantec, Melbourne, Australia

ABSTRACT
This paper aims to present a Virtual Reality (VR) Acoustic application that allows a user to experience the effect of changing the acoustic properties of surface materials in a room, while freely moving in a VR scenario. In a previously issued paper (Part I), the background and description of the software were presented in detail. While a brief introduction is also described in this paper, the focus of this work is to compare a real case study of an existing classroom with the same simulated room in VR. A comparison between the Ambisonics Room Impulse Response (ARIR) measurements and 3D simulated ARIRs using CATT is presented, while providing an overview of the potential use for such application. In addition to the integration of the VR acoustic application in Unity3D, a link to an online video to provide a subjective listening experience of the same room is also shared.

1 INTRODUCTION
Auralizations are often used for prospective acoustic design of rooms as they convey the intrinsic characteristics of the space and makes them audible in a familiar way. They are a strong front-end tool allowing a hearing description of the generally back-end and highly theoretical computer prediction results. They allow a more reality-friendly approach in the design of any kind of virtual space; may they be already existing, yet to be built or destroyed through time. Traditionally, auralization simulations have been static, providing the response of a specific room between a source and a receiver with no option for the user to freely move within the environment.

Whilst some dynamic auralization solutions have been developed in the recent years (Southern, Wells, and Murphy 2009; James 2011; Hodgson et al. 2008; Lindebrink and Nätterlund 2015; Ballestero, Robinson, and Dance 2017), the application herein described presents the methodology to provide an upgraded VR-friendly experience with an additional degree of freedom allowing the user to freely move within a virtual space while, not only experiencing the acoustics of the room, but being able to change the acoustic properties of the materials of the room on the spot, and hence changing the overall acoustic experience.

Audio rendering is usually associated with audiovisual activities such as cinematic or gaming content. Whilst video rendering has benefited from significant advances for the past decades, the standard of audio realism has remained relatively low. This is mostly due to the inherent nature of both audiovisual contents, i.e. a screen based medium. Advanced audio rendering technologies have been relatively minimized during this period due to their conflicting potential with the main medium – the video screen – as they could bring inconsistent or confusing auditory cues to the spectator, leading to a general disruption of the audiovisual immersion, e.g. strong sound localization effects for rear sound reproduction in cinematic or gaming configurations instead of ambient fill effects. However, a few improvements to 3D audio rendering where made in first-person game environments where the latter could provide tactical assistance to the player, usually being headphone-based. Such potential applications nurtured the development of better spatial sound technology.

Whilst VR has been resolutely more focused on its 'video-ready' aspect, actual audio rendering capabilities for VR are still not at the highest level they could be. Signal processing improvements have enabled the incorporation of real-time HRTF's convolution of dry acoustic signals within game engine architectures (Rungta et al. 2017)(Rungta and Hill 2017) (Aspöck et al. 2019) (Thery and Katz 2019) but solutions that provide accurate acoustic characteristics of spaces, in plugin format, are not commercially available.

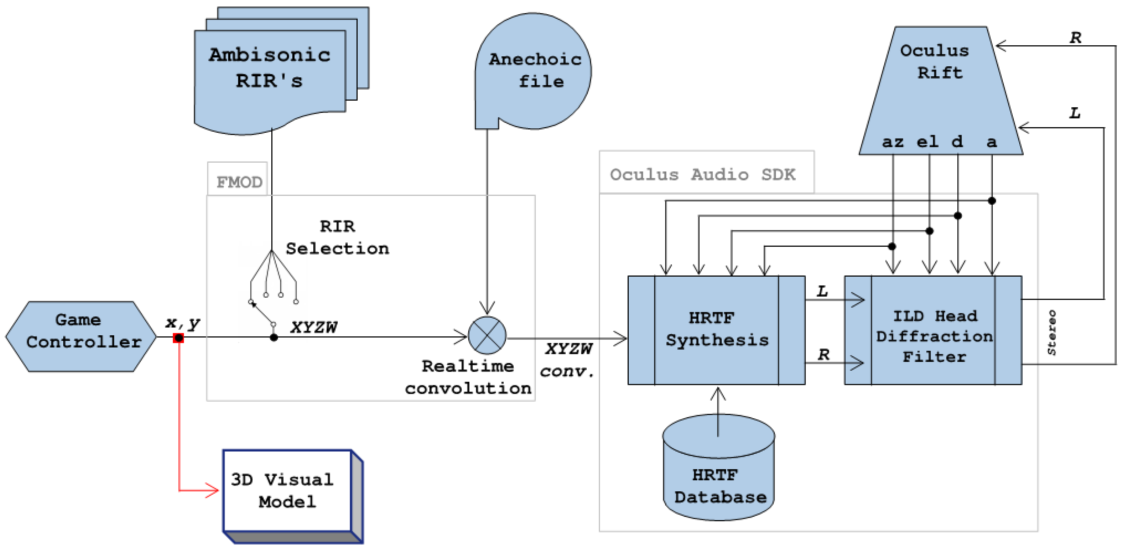
To achieve a user satisfactory immersion in VR environments, it is therefore of the upmost importance to create spatial sound that matches the visual impression – convergence of at least two senses, vision and audition.

The incorporation of acoustically simulated auralizations into VR thus results in a two-fold advantage, (i) a more accurate capture and reproduction of the different 3D sound fields than standard audio engine capabilities (i.e. acoustic simulations vs. intuitive audio mixing), and (ii) the ability to modify the acoustic field of a room by virtue of selecting different material finishes and hence creating a different auditory experience by choosing different materials.

The remainder of the paper is divided in two parts, the first provides an overview of the VR acoustic application previously developed, and the second presents an example of a real classroom which is used to compare the results of measured versus simulated ARIRs. It is noted that the main purpose of this paper is not provide evidence on the quality or potential of reproducing a real ambisonics measurement in a virtual scenario, but to try to link the subjective experience with the results obtained through objective measurements.

2 SOFTWARE OVERVIEW

Figure 1 displays a block diagram consisting of the most relevant elements involved in the application implementation.



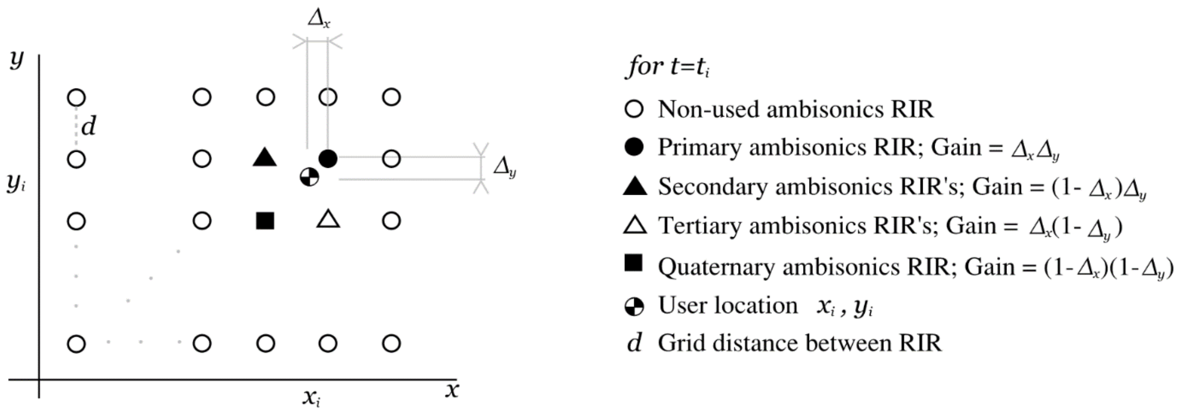
Source (Castro, Ballestero, and Platt 2018)

Figure 1: Block diagram of the application architecture

- The game/VR controller is used to move the user through the space, effectively defining the x and y coordinates in a constant plane, i.e. the user position within the virtual model.
- The application requires a set of ARIRs that must be either previously measured or calculated.
- A proprietary algorithm (refer to Section 2.1) is issued to choose which ARIR is to be used depending on the (x, y) location of the user.
- The application then provides the ability of performing 16 real-time convolutions with a preloaded anechoic file.
- The last step is then to port the real-time audio data to the head mounted display SDK. In there, the predefined HRTFs will take care of the head orientation of the user providing the 3D spatialization component, providing the user with the aural information corresponding to its location in relation to the source.
- There is a feedback loop constantly happening where the head orientation feeds back the information while the convolutions keep on happening.

2.1 ARIR selection

The location of the user in the application provides the x and y coordinates within the room grid, where the ARIR's have been previously measured or simulated. The specific pair of coordinates (x, y) of the user in relation to the ARIR coordinates will determine which impulse responses are to be used at any moment in time. The application will always consider four ARIRs corresponding to the four nearest ARIRs with respect to the instantaneous user location. Each of these four ARIRs is assigned with a gain weighting – equivalent to a fader value in a mixing desk – and simultaneously convolved with the anechoic file. It should be noted that, in fact, there are actually 16 files being simultaneously processed at any given time, as each of the active ARIR will consist of the X, Y, Z and W ambisonics channels.
Figure 2 displays the RIR selection process. At any moment, the distances between a user location and the predefined grid locations are used to apply the appropriate gain to each ARIR.



Source (Castro, Ballestero, and Platt 2018)

Figure 2: Example of the ARIR selection process

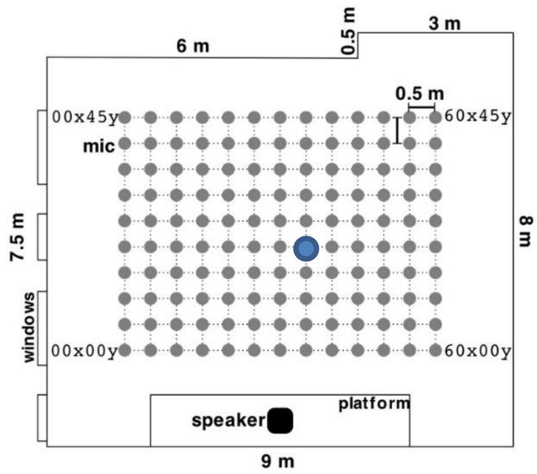
3 CASE STUDY

As a way of validation, the following example was created to test the viability of the proposed application. The aim of this particular exercise was to demonstrate that the application could be used to measure the acoustic properties of an existing room, then create a virtual model of the room, and finally apply acoustic treatment to the room to experience potential benefits or changes due to addition of acoustic material. Once the virtual model is created, it is rather straightforward to change the acoustic properties of the materials in the room providing different acoustic scenarios. This way, the user can switch between the scenarios with the press of a button (of the VR/game controller) changing the acoustic properties of the room and hence the acoustic experience while freely moving in the room.
An existing room of the Queen Mary University of London was chosen for this purpose. The main reason for such selection was that a series of ARIR's measurements were already available (Stewart and Sandler 2010) as well as the accessibility of enough information about the room to be able to generate the 3D visual model. The chosen classroom is within the School of Electronic Engineering and Computer Science.

3.1 Measurements

The classroom object of study measures roughly 7.5 x 9 x 3.5 m, with a floor area of 159 m^2 and a volume of 236 m^3. The room consists of mainly reflective surfaces with a linoleum floor, painted plaster walls and ceiling, and a large whiteboard. When in use for lectures the room is filled with desks and chairs. These were stacked and moved to the side against the windows during the measurements. ARIR measurement locations are 0.5 m apart arranged in 10 rows and 13 columns relative to the speaker, with the 7th column directly on axis with the speaker. These ARIRs are released under the Creative Commons Attribution-Noncommercial-Share-Alike license with attribution to the Centre for Digital Music, Queen Mary University of London. The set of ARIRs were taken within the classroom using the sine sweep technique with a Genelec 8250A loudspeaker and a Soundfield SPS422B microphone (Farina 2000).

Figure 3 displays the nomenclature for the filenames for each position. The filenames indicate the microphone/channel and position. For example, Z05x10y.wav is the Z channel, up-down bidirectional, from the Soundfield microphone at the 2nd column from the right of the 3rd row from the front when facing the loudspeaker. 35x20y.wav is the DPA microphone at the 8th column from the left of the 5th row from the front. See Figure 3 for more details.



Source (Stewart and Sandler 2010)

Figure 3: Example of the measured RIR grid for the classroom

3.2 Visual model

A virtual reproduction of the classroom was generated using Unity3D, a multipurpose game engine that can be used to create both 2D and 3D virtual environments for computers, mobile devices, and video game consoles. Unity3D possesses a wide range of tools for both visual and audio features involved in today's gaming industry. Its scripting capabilities (C# and Javascript) also allow endless custom developments. The following Figure 4 displays a visual representation of the room as well as sample of the measurements set up.

Source (Authors / Stewart 2018 / 2010)

Figure 4: 3D visual representation of the room (Left) and classroom during ARIR measurements (right)

3.3 Acoustic computer modelling

The acoustic simulation of the virtual space is usually conducted with dedicated software that can calculate and export ARIRs in ambisonic format, such as CATT-Acoustic, ODEON, RAMSETE or EASE. These software packages are able to create or to import of a virtual 3D CAD model of a space, and the ability to assign specific acoustic properties to the model boundaries, such as absorption and scattering/diffusion. It is to be noted that most of these software packages work under the statistical and geometrical formulations of acoustics, therefore limiting the extent of acoustic simulations to a certain frequency range and to airborne sound. Although it is obvious that the virtual environment would be as good as the simulations that are generated, there is enough evidence that, when used appropriately, the acoustic software can accurately reproduce the acoustics at a given point (Savioja and Svensson 2015; Georgiou et al. 2019). In this case study CATT-Acoustic v9.1c (Dalenback: 2018) acoustics was used.

CATT-Acoustic is a geometric acoustic modelling software package which uses a hybrid cone-tracing and image-source approach, called The Universal Cone Tracer (TUCT). Room surfaces can be modelled with frequency-dependent (octave-band) absorption and scattering coefficients and cone-tracing is performed independently for each band. Edge scattering is also able to be applied where in-plane impedance changes occur within the model.

TUCT calculates the full-length impulse response using a mix of image source modelling for first-order specular reflections, and cone and ray tracing first-order diffuse and later reflections. The cone and ray tracing calculation method uses two different methods to generate diffuse reflections, deterministic split-ray scattering and random scattering. Random scattering produces a single reflected ray for each incident ray, the reflected ray will either be specular or scattered, the probability of which is determined by the scattering coefficient. Split-ray scattering generates a specular reflection ray and multiple new rays (of lower energy) for each incident ray. Consequently, a very large number of diffuse rays (each with relatively low energy) are generated after a few orders of reflection, resulting in increased reflection density as the sound decays.

The room-prediction algorithm includes phase information in order to predict impulse responses. The calculation method used for the reflection-path phase depends on the reflection order. Direct sound and first-order specular reflections employ a phase response based on magnitude using Hilbert-transform relations. For first-order diffuse reflection and all second- and higher-order reflections, the phase becomes that of the linear-phase, octave-band, finite impulse responses involved.

While measurements of absorption coefficients are widely available and generally able to be adapted to match specific model conditions, scattering coefficients can present an area of ambiguity due to the lack of test information available for commonly used materials. Generally, scattering coefficients are approximated based on the geometric ‘roughness’ or features of the surface in question, and relative to the wavelength of incident sound. Scattering is a particularly important consideration for spaces which lack the shape or surface orientations to achieve good geometrical mixing or have an uneven distribution of absorption properties. Appropriate selection of these model inputs is critical in the production of accurate ARIRs.

3.3.1 Absorption and scattering/diffusion coefficients used in the computer model

Table 1 displays the absorption and diffusion coefficients for each surface type used in the acoustic 3D model. The absorption coefficients were chosen based on published data and default values provided with CATT database, while diffusion values where used based on recommendations by (Rees, I; James, A. 2016).

Table 1: Absorption coefficients (α) and diffusion (d) of the materials in the classroom.

Frequency (Hz)	Linoleum (α/d)	Plasterboard wall (α/d)	Plasterboard ceiling (α/d)	Glazing (α/d)	Whiteboard (α/d)	MFT (acoustic ceiling treatment)
125	0.04/0.2	0.08/0.2	0.08/0.2	0.35/0.2	0.02/0.2	0.55/0.2
250	0.04/0.2	0.11/0.2	0.11/0.2	0.25/0.2	0.02/0.2	0.85/0.2
500	0.04/0.2	0.08/0.2	0.05/0.2	0.18/0.2	0.03/0.2	1.00/0.2
1000	0.08/0.2	0.09/0.2	0.03/0.2	0.12/0.2	0.04/0.2	0.95/0.2
2000	0.08/0.2	0.09/0.2	0.02/0.2	0.07/0.2	0.04/0.2	1.00/0.2
40000	0.08/0.2	0.08/0.2	0.03/0.2	0.04/0.2	0.05/0.2	1.00/0.2

3.3.2 Hardware

Table 1 below displays both the minimum requirements for a VR computer (“VR Ready Equipment” 2019) as well the specifications of the hardware used in the development of this application:

Table 2: VR minimum requirements / used hardware

Component	Minimum required	Used in this application
Video Card	NVIDIA GTX 1060 AMD Radeon RX 480 or greater	NVIDIA GTX 1070 8GB GDDR5
CPU	Intel i5-4590 equivalent or greater	Intel i7 8750H Processor,
Memory	8GB RAM or more	24GB RAM or more
Video Output	Compatible HDMI 1.3 video output	✓
USB Ports	3 x USB 3.0 ports plus 1x USB 2.0 port	4 x USB 3.0 ports
OS	Windows 10	Windows 10
Head-mounted-display	Any	Oculus Rift S

The acoustic simulation using CATT was also run in this same computer. It is relevant to highlight that the application (the *.exe file) has been built in such way that is able to run both in the VR ready computer as well as in a standard laptop/PC. This means that the application will detect if a head-mounted-display (e.g. Oculus Rift) is connected and will direct the graphics output to the VR headset or to the standard computer screen. This way the very same application could be used in VR or in a standard laptop/PC.

3.4 VR integration

The process of exporting the ARIR’s from CATT into Unity3D is rather simple, however the only step that is delicate is the orientation of the impulse responses. Due to the ambisonics nature of the measured/simulated data, the impulse response requires certain orientation so the user can be correctly collocated in the 3D virtual space. In the case here presented, special attention was required to the orientation of the ARIR’s so the virtual world matches the configuration that was used during the measurements in the real classroom. For this reason, the classroom application has an in-built scenario that represents an anechoic chamber with the same dimensions of the classroom. The purpose of this virtual anechoic chamber is to facilitate the validation steps to ensure that the physical and virtual locations match. With the aid of the audio-visual help of the anechoic room, it is easier to detect errors in the import data, and to ensure that the 3D spatialization works as expected.

4 DISCUSSION

This link (Castro 2018) displays a screen recording of a user using the classroom application. In this example there are four different acoustic scenarios: 1) the measurement, where the measured ARIR’s are convolved with an speech recorded in an anechoic chamber, 2) the simulation, where the ARIR’s generated from CATT are convolved with the same anechoic speech recording, 3) the treatment scenario, where the entire ceiling of the room is changed from plasterboard to mineral fibre tiles (see Table 2 for details). Finally, in the 4th scenario, the example of the anechoic room is also shown to demonstrate the 3D spatialization, i.e. moving your head 180° sideways in an anechoic environment, allows the used to perceive a clear auditory clue as to where is the sound source is located.

4.1 Objective vs subjective assessment

It is pertinent to point out that one of the potential uses of this application is aiming to replicate existing acoustic spaces. For this purpose, the simulation of a space needs to sound as natural and as close as possible to the measured real space. In the classroom example, switching between scenario 1 and scenario 2 should not provide any noticeable difference, i.e. one should have the same acoustic experience in both. However, the subjective experiences (perceptible differences between 1 and 2) seems to depend on the user ear training / experience in music or acoustics. The goal or success of this simulation exercise would be clear when the user cannot tell the difference between the two (measured vs simulated). A non-scientific conclusion based on the feedback of the number of people that have tried the application (in the order of a few hundred) is that, generally, any user with no formal acoustic training is not able to tell the difference between the two scenarios.

However, those individuals with experience in audio and acoustics are more likely to detect minor differences in the audio signal. In order to provide some details on these discrepancies, the following graphs aim to provide an objective comparison of the measured vs simulated data, as well as compare those with the simulated and treated scenario.

It should be noted that the purpose of this section is not to provide details on the validity or accuracy of how to use computer simulations to recreate real acoustic scenarios; there are extensive examples and (Savioja and Xiang 2019; Lokki and Savioja 2005; Pelzer et al. 2014; Katz et al. 2018) on the limitations and capabilities of such techniques. The idea behind this comparison is to put emphasis in how the objective data matches the subjective experience, i.e. how well close the user perceives the space regardless of the objective differences in the data. Hence, a series of comparisons between the measurements and the VR simulations are presented to provide context to the simulations; in particular the reverberation time (RT20); similar conclusions could be extracted from the Early Decay Time (EDT) but have been omitted for brevity. It is noted that due to the ambisonics nature of the recordings, each of the four channels has been presented (X, Y, Z and W).

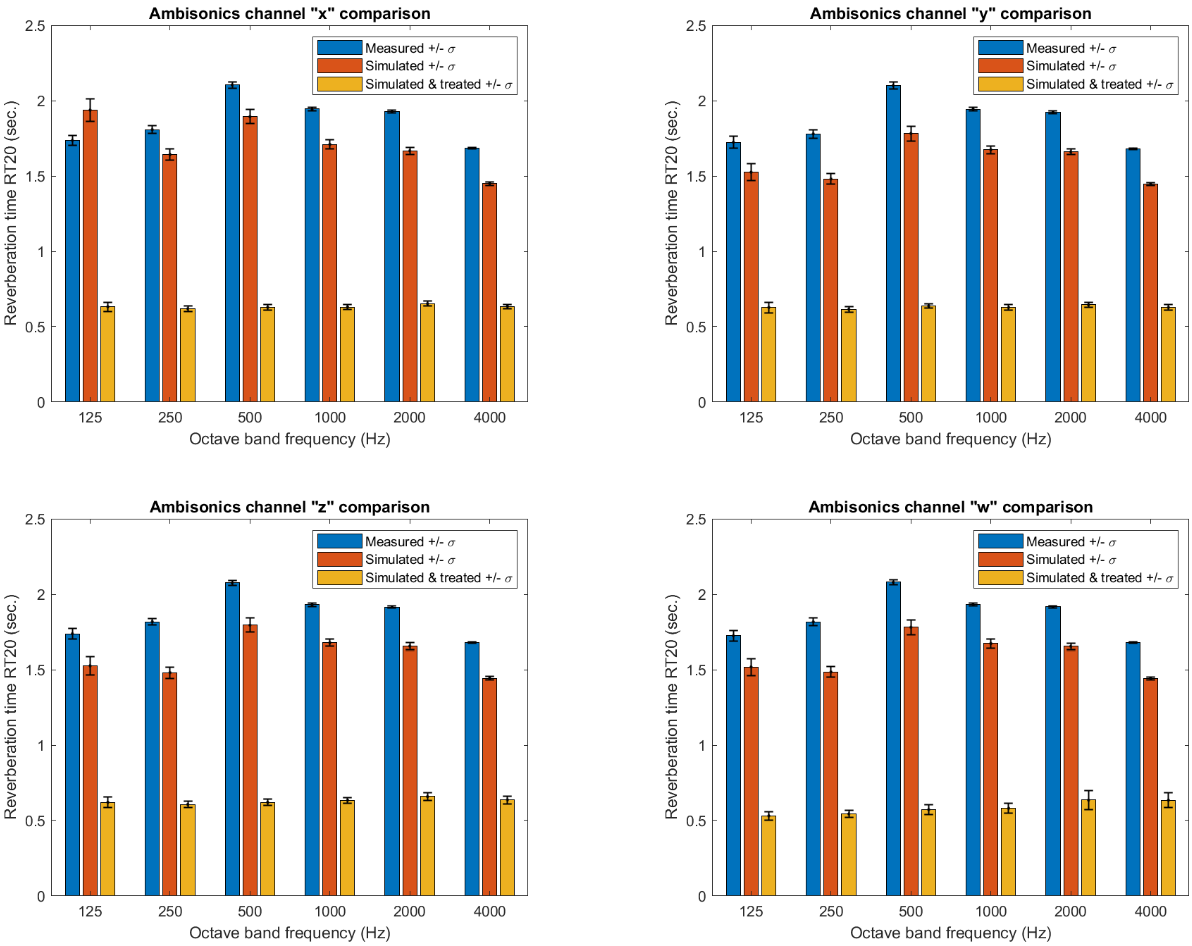


Figure 5: 3D visual representation of the room (Left) and classroom during ARIR measurements (right)

As it can be seen, there are some differences between the measured and simulated RT20 derived from the ARIR's. Interestingly, although the difference at individual octave bands can be up to 20%, the subjective experience seems to indicate that these differences are not clearly perceived by the general audience. Notwithstanding that, this classroom example has not been fine-tuned, i.e. the simulated absorption coefficients where not modified to match the measurements, and a deeper assessment of appropriate scattering coefficients hasn't been undertaken. What is certainly noticeable during the application, however, is the significant change in the acoustics of the space once the ceiling treatment is applied, resulting in a drop in the reverberation time from ~1.8s to ~0.5s.

Precisely these discrepancies between the objective and the subjective impression suggested that there could be other graphical interpretations of the objective data that could match the subjective impression. Figure 6 displays the waveform of three wav files representing the measured, the simulated, and the simulated & treated scenarios. In this case, only the channel W (omnidirectional) was chosen in the middle of the room (highlighted grid point in Figure 2). These wav files were also analysed using a Short-time Fourier Transform (STFT) to display a spectrogram; this way, the time, frequency and amplitude level can be seen in one graph.

The wav files in Figure 6 correspond to the already convolved files with the same anechoic recording file used in the application. In this case, the differences in the spectrogram are visible but less noticeable than in the RT20 graphs. However, it is looking at the bare waveform that the differences between the three cases are, again, rather obvious.

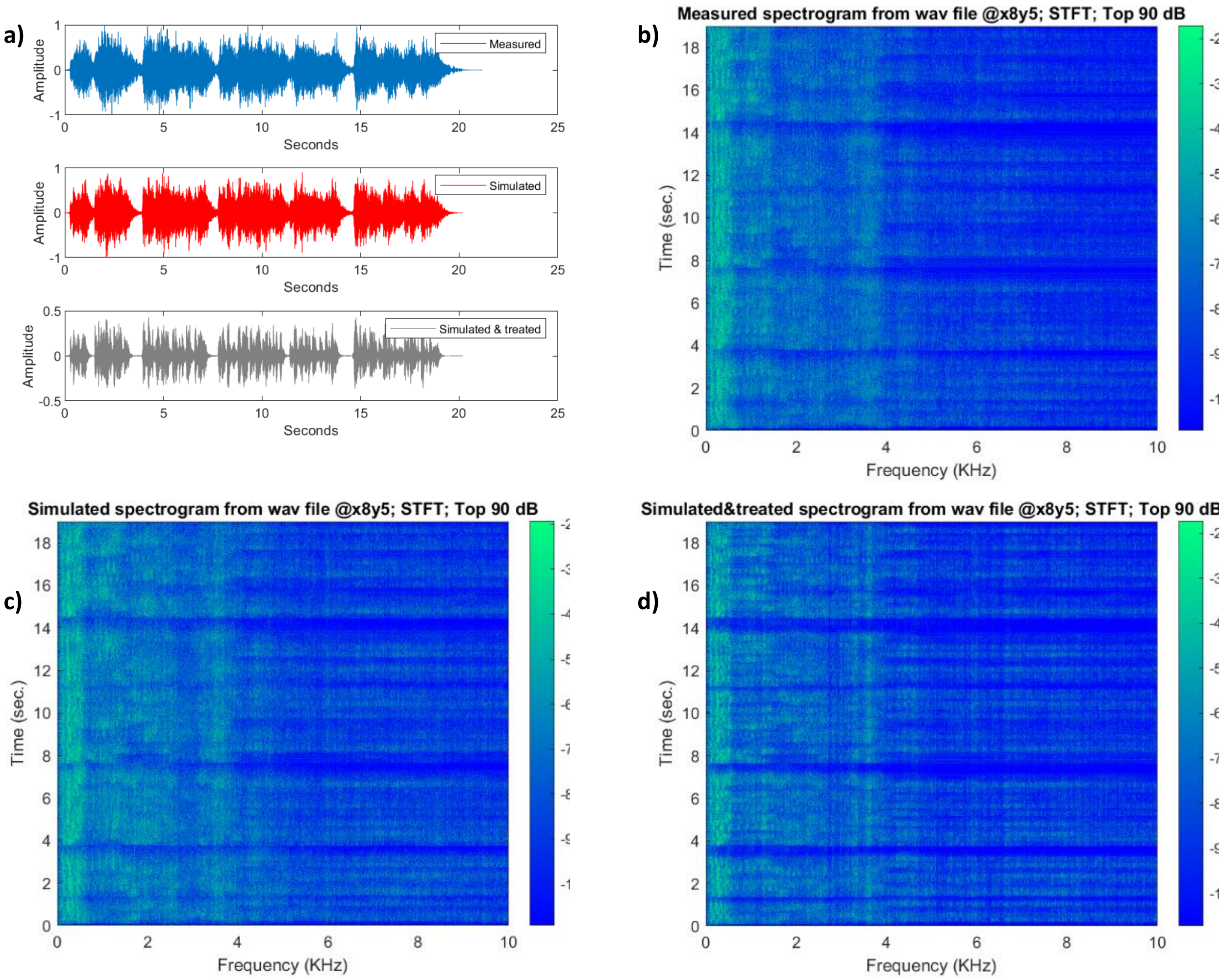


Figure 6: a) Comparison of wav files recorded in the centre of the classroom showing three scenarios: measured, simulated and simulated with acoustic treatment; b) spectrogram of the measured wav file; c) spectrogram of the simulated wav file, and d) spectrogram of the simulated and acoustically treated wav file.

Finally, an additional acoustical parameter was chosen to further illustrate the similitudes/differences between the three scenarios: the direct-to-reverberant ratio (DRR). Although this parameter could be better analysed using binaural recordings, due to nature of the ambisonics source signal from the microphones, it was decided to perform the analysis using the impulse response on the channel W instead of converting the quadruplicate XYZW, to avoid further signal manipulation. The objective of this analysis is to present, in a graphical format, how different the DRR is in a fairly reverberant scenario (RT20 > ~1.8sec @1000Hz) in comparison with the treated room at RT20 > ~0.5sec @1000Hz. To do so a comparison of all the channels against one another was conducted for the three cases (measured, simulated and treated).

Each subgroup was formed with the 130 individual impulse response signals from the grid (Figure 3). As expected, all the channels combinations display a common behaviour, i.e. the DRR is higher for the treated room, as an acoustic absorbent ceiling has been added to the room and hence reduce the reverberant field. This seems to indicate that there is less booming effect, echoes and overall reverb and the signal perceived by the user is a lot clearer and sharp. This plot exemplifies the relation between what the user can hear and what can be objectively measured.

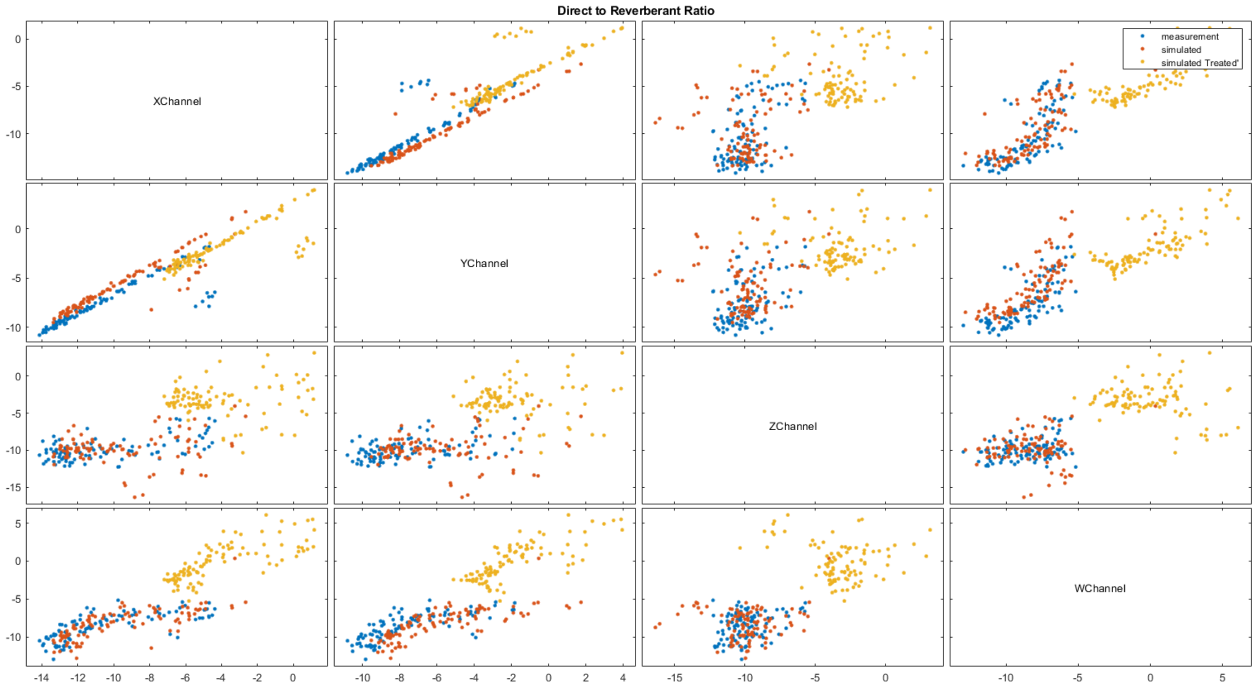


Figure 7: Matrix of scatter plots of all ambisonics channels against each other showing the DRR grouped in Measured, Simulated and Simulated and treated.

5 REMARKS AND FUTURE WORKS

A VR application that allows a user to freely move in VR while changing the acoustic properties of the room in real time has been developed. This application has demonstrated the potential for the use of VR acoustics in both a commercial environment (for design purposes) as well as for more artistic endeavours (recreating acoustic spaces that no longer exist).

In order to correlate the subjective and objective experiences, the actual output of the acoustic simulation has been compared with the measured data. While the differences in RT, waveform and spectrogram are rather noticeable, it is also perceived that most users are not able to distinguish between the two first scenarios (measured vs simulated). It is relevant to point out that this outcome occurs without tuning the simulated model to more closely match the measured data. This conclusion results in more questions than answers in regard to interpretation of what the user hears. Hence, additional investigations would be required to further understand if this mismatch between what the users hear and the objective data can be understood. In summary, two potential lines of work are opened 1) the relation between subjective impressions and objective data when listening to convolutions in VR, and 2) the use of multiple sources in VR, i.e. being able to use two or more noise sources concurrently in VR.

ACKNOWLEDGEMENTS

The authors would like to thank Eric Ballestero (Department of The Built Environment, London South Bank University) for his contribution and help in the preparation of the Part I of this paper, which set the framework for this Part II.

REFERENCES

Aspöck, Lukas, Lucas Mösch, Jonas Stienen, and Michael Vorländer. 2019. "Auralization of Interactive Virtual Scenes Containing Numerous Sound Sources." In *Proceedings of the 23rd International Congress on Acoustics*, 6401–8.

Ballestero, Eric, Philip Robinson, and Stephen Dance. 2017. "Head-Tracked Auralisations for a Dynamic Audio Experience in Virtual Reality Sceneries." *International Congress on Sound Vibration*, 1–8. https://www.iiav.org/archives_icsv_last/2017_icsv24/content/papers/papers/full_paper_386_20170329170651697.pdf.

Castro, Daniel. 2018. "Virtual Reality Acoustic Plugin v1.3." 2018. https://rb.gy/af6a04.

Castro, Daniel., Eric. Ballestero, and Sam. Platt. 2018. "Walk-Through Auralization Framework for Virtual Reality Environments Powered By Game Engine Architectures." In *Proceedings of the Institute of Acoustics Vol. 40. Pt.4. 2018*. Vol. 40.

Dalenback:, B. I. 2018. "CATT-Acoustic v9.1c User's Manual." www.catt.se.

Farina, Angela. 2000. "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique." *Proc. AES 108th Conv, Paris, France*, no. I: 1–15.

Georgiou, Fotis, Baltazar Briere, De La Hosseraye, and Maarten Hornikx. 2019. "Design and Simulation of a Benchmark Room for Room Acoustic Auralizations Design and Simulation of a Benchmark Room for Room Acoustic Auralizations." In *Proceedings of the 23rd International Congress on Acoustics*.

Hodgson, Murray, Natalie York, Wonyoung Yang, and Mark Bliss. 2008. "Comparison of Predicted, Measured and Auralized Sound Fields with Respect to Speech Intelligibility in Classrooms Using CATT-Acoustic and ODEON." *Acta Acustica United with Acustica* 94 (6): 883–90.

James, Adrian. 2011. "Demonstration of CATT-Acoustic Walker Module." 2011. https://goo.gl/4xqyu1.

Katz, Brian F G, David Poirier-quinot, Sorbonne Université, Institut Jean, Le Rond, F- Paris, Université Paris-saclay, and Paul Luizard. 2018. "Objective and Perceptive Evaluations of High-Resolution Room Acoustic Simulations and Auralizations." In *Euronoise*, 2107–14.

Lindebrink, J, and J Nätterlund. 2015. "An Engine for Real-Time Audiovisual Rendering in the Building Design Process." *Acoustics 2015 Hunter Valley*, 1–8.

Lokki, Tapio, and Lauri Savioja. 2005. "Evaluation of Auralization Results 1 Introduction 2 Evaluation of Auralization Qual- Ity 3 Performed Evaluations of Aural- Ization Quality." In *Forum Acousticum*.

Pelzer, Sönke, Lukas Aspöck, Dirk Schröder, and Michael Vorländer. 2014. "Integrating Real-Time Room Acoustics Simulation into a CAD Modeling Software to Enhance the Architectural Design Process." *Buildings*, 113–38. https://doi.org/10.3390/buildings4020113.

Rees, Ian., and Adrian. James. 2016. "COMMON PITFALLS IN COMPUTER MODELLING OF ROOM ACOUSTICS." In *Proceedings of the Institute of Acoustics Vol. 38. Pt.1. 2018*. Vol. 38.

Rungta, Atul, and Chapel Hill. 2017. "Dissertation Proposal : Simulation and Evaluation of Sound Propagation Effects for Virtual Environments," 1–12.

Rungta, Atul, Nicholas Rewkowski, Roberta Klatzky, Ming Lin, and Dinesh Manocha. 2017. "Effects of Virtual Acoustics on Dynamic Auditory Distance Perception," no. c.

Savioja, Lauri, and U. Peter Svensson. 2015. “Overview of Geometrical Room Acoustic Modeling Techniques.” *The Journal of the Acoustical Society of America* 138 (2): 708–30. https://doi.org/10.1121/1.4926438.

Savioja, Lauri, and Ning Xiang. 2019. “Introduction to the Special Issue on Room Acoustic Modeling and Auralization” 145 (1). https://doi.org/10.1121/1.5099017.

Southern, Alex, Jeremy Wells, and Damian Murphy. 2009. “Rendering Walk-through Auralisations Using Wave-Based Acoustical Models.” *European Signal Processing Conference*, no. Eusipco: 715–19.

Stewart, Rebecca, and Mark Sandler. 2010. “Database of Omnidirectional and B-Format Room Impulse Responses.” In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 165–68.

Thery, David, and Brian F G Katz. 2019. “Anechoic Audio and 3D-Video Content Database of Small Ensemble Performances for Virtual Concerts.” In *Proceedings of the 23rd International Congress on Acoustics*, 739–46.

“VR Ready Equipment.” 2019. RoadtoVR. 2019. https://www.roadtovr.com/how-to-tell-pc-virtual-reality-vr-oculus-rift-htc-vive-steam-vr-compatibility-tool/.

Stewart, Rebecca and Sandler, Mark. "Database of Omnidirectional and B-Format Impulse Responses", in *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2010), Dallas, Texas, March 2010*.