EUROREGIO BNAM2022
Joint Acoustics Conference
9th-11th of May, Aalborg, Denmark

Perceptual evaluation of auralizations from a wave-based method in a virtual environment

Tanmayee Pathre[1,*], **Maarten Hornikx**[1], **Armin Kohlrausch**[2]

[1]Department of the Built Environment, Eindhoven University of Technology, Eindhoven, Netherlands.

[2]Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, Netherlands.

[*]t.u.pathre@tue.nl

Abstract

In recent years, Virtual Reality (VR) has become a powerful tool in studies involving perceptual evaluation of auralization methods. The aim of this study was to investigate the extent to which binaural auralizations from a wave-based method are perceptually close to those from measurements for two different room acoustical scenarios. The signals used for auralization are simulated using the Discontinuous Galerkin method. The simulations encompass geometric details of the room, material properties of the objects in the room and receiver directivity including head orientation. The measured Binaural Room Impulse Responses (BRIRs) were limited to an upper frequency of 2.5 kHz to match the bandwidth of the simulations. The reverberation time results show a fairly good agreement between the measured and simulated data, thereby indicating that the wave-based method can be held as a reference for achieving perceptually realistic auralizations. To evaluate this, a listening experiment was carried out in an interactive VR environment employing a 3D model of the room integrating a dynamic convolution framework with headphone reproduction. The BRIRs were convolved in real-time with two types of source signal – percussion and female voice speech. Twelve normal hearing participants performed a rating task for judging differences between measured and simulated BRIRs for four perceptual attributes - reverberance, clarity, bassiness and externalization. The results indicate that the mean difference between measurement and simulation is statistically significant for reverberance, clarity and bassiness but not for externalization. Additionally, this study highlights a state-of-the-art experimental framework which can be used for perceptual evaluation studies independent of the computational method used to derive the auralizations.

Keywords: perceptual evaluation, wave-based method, auralization, virtual reality, perceptual attribute.

1 Introduction

Several computational acoustics methods are applied for room acoustic modelling and auralization purposes. Amongst them, the wave-based methods are known to accurately model low-frequencies and physical phenomena such as reflection, diffraction and scattering. Although computationally expensive at high frequencies, the time-domain wave-based methods have been successfully applied for room acoustic modelling [1], [2], [3]. The room acoustic simulation when rendered for auditory reproduction serves as an input for subjective evaluation studies. For an auralization to be realistic, the room material properties, directivity of source, geometrical details of the room and receiver directivity needs to be taken into account. Several studies on perceptual evaluation of room acoustic modelling methods and various auralization systems provide an account of the needed requirements aiming to achieve realistic auralizations [4], [5] [6]. Lately, Virtual Reality (VR) has widely been implemented in the domain of building design and building acoustics [7]. The incorporation of VR is growing in addressing questions such as accuracy of sound localization in reverberant environments [8] and sound effects of building characteristics on cognitive performance [9].

The current study implements VR in a listening test for subjective evaluation of the auralizations derived from a wave-based method. Our goal is to investigate the degree to which the auralizations come close with the measurements for the four perceptual attributes - reverberance, clarity, bassiness and externalization [10]. The choice of perceptual attributes was primarily based on the computational method used to derive the simulation and what answers it can provide to potentially improve the simulation.

This paper starts by providing an overview of the approach adopted for the measurements and wave-based simulations and the objective results [11]. Next, an elaborate account on the listening experiment design is presented. It informs the reader about the VR system integrated with a real-time convolution framework and design of the 3D user interactive environment. Finally, the results from the listening experiment are presented.

2 Background

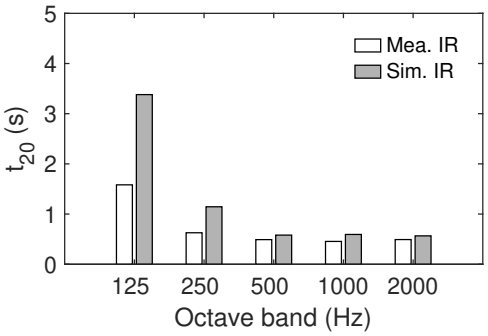
2.1. Measurements

A small room (87.63 m^3) in a laboratory was acoustically treated to achieve two different room acoustical scenarios (see Figure 2). The room received acoustical treatment with carpet and acoustic porous panels. This well-treated environment was referred to as Scenario 1. The second scenario received a fairly less amount of acoustical treatment and was referred to as Scenario 2. For each scenario, BRIR measurements were carried out for a source-receiver distance of 1.5m and 3.5m. The BRIR measurements were carried out using B&K TYPE 4125, an omnidirectional source with an exponential sine sweep of 175s and the B&K 4128 Head-and-Torso Simulator (HATS) as a receiver. The measurements were performed for 72 different head orientations in the horizontal plane with a resolution of 5°. The surface impedances of the acoustic panels and the carpet tiles and were measured both with a pressure-velocity (PU) probe and an impedance tube.

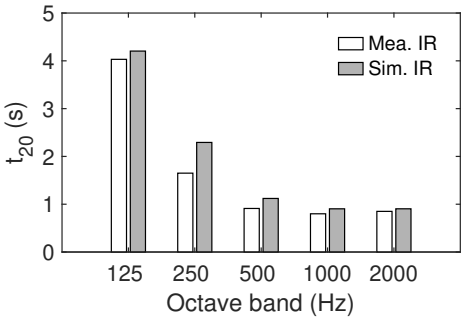
2.2. Wave-based simulations

The simulations use the frequency dependent time-domain boundary conditions using the time domain Discontinuous Galerkin (DG) method. They encompass geometric details of the room, surface impedances of the porous panel and carpet, and the reflection coefficient of the hard concrete walls. The sound source was simulated as an omnidirectional Gaussian pulse. In order to simulate the head-rotations and binaural aspects of human hearing, the simulations were rendered to be reproduced binaurally over headphones. A dual concentric sphere of receiver points was placed in the simulation domain. Spherical harmonics along with the Head Related Transfer Functions (HRTFs) from a database of measured BRIRs [12] were applied to simulate the same set of head-rotations as in the measurements [13]. For the purpose of the listening experiment, an improved set of simulations was re-computed compared to the ones mentioned in [11]. The new calculations were carried out for all four room acoustic scenarios with an upper frequency bound set to 2.5 kHz . Within this new set of simulations, the receiver grid for spherical harmonics reconstruction was corrected. The reflection coefficient of the hard concrete walls was adjusted from 0.991 to 0.996. An optimization algorithm was used for angle adjustment in the horizontal plane between source and receiver for a better match with the measurements. The reader can refer to [11] for further details on measurements and simulations and the methods adopted for their post-processing.

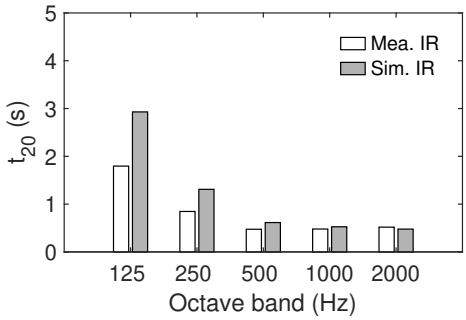
The reverberation time results are presented in Figure 1. It can be observed from Figure 1 that overall, for all four scenarios the simulations are in fairly good agreement with the measurements. A closer look at the results tells us that the simulations are in a much better agreement with the measurements for Scenario 2 as compared to Scenario 1 for both source-receiver distances 1.5m and 3.5m.



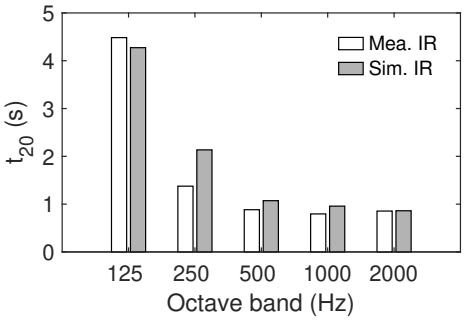
(a) Scenario 1 1.5m



(b) Scenario 2 1.5m



(c) Scenario 1 3.5m



(d) Scenario 1 3.5m

Figure 1: Reverberation Times T_{20} of IRs for the four room acoustic scenarios

3 Listening Test Experimental Design

Virtual Reality Dynamic Convolution Framework

A virtual reality integrating a dynamic convolution framework forms the cornerstone of the listening experiment setup. Unity Game Engine [14] and Max/MSP (Cycling'74) [15] are the two major building blocks of the system design. A new object, called dynamic convolution object was developed with the convolution approach presented in [16]. The object was programmed for Max/MSP to perform a real-time convolution on the CUDA. The source code for the dynamic convolution object was written in C and compiled using Visual Studio 2017. The object was implemented as a dynamic library incorporated in Max/MSP for general usage and specifically connected to the Unity Game Engine to feed the spatial parameter (head movements from the user). The dynamic convolution object takes pre-computed BRIRs for left and right channels separately to be convolved with source signals.

Unity Game Engine provides a platform for functioning of VR. When the user/s when mounts the VR Head Mounted Display (HMD), their head-movements/rotations are represented as angles within Unity Game Engine [14]. The angles (real valued numbers) are converted into bytes and sent over UDP (User Datagram Protocol) to Max/MSP. In Max/MSP, these bytes are translated into real numbers which correspond to the angles or head-orientation of the user. The Max patch is equipped with functions to evaluate the angles and match it with the associated BRIR filter number. This filter convolved with the source signal in real-time on CUDA. A change in head-orientation of the user, would update the BRIR filter number allowing convolution with a different filter. The convolved audio signal is then reproduced over the headphones to be heard by the user.

3.1. 3D Interactive Environment Design

Four 3D models of the room, each corresponding to a room acoustical scenarios were created in SketchUp [17] as shown in Figure 2.

(a) Scenario 1 1.5m

(b) Scenario 2 1.5m

(c) Scenario 1 3.5m

(d) Scenario 2 3.5m

Figure 2: 3D room models of four room acoustic scenarios created in SketchUp.

The assignment of materials and textures to various room objects was also performed in SketchUp. The 3D models of the room were imported to Unity (Version 2019.3.15f) in .fbx format. Visual rendering and offline lighting was done in Unity to achieve realistic and immersive scenarios. The 3D models of all the four room acoustic scenarios were rotated by 270°in the Unity Game Engine as SketchUp and Unity follow different co-ordinate systems. The material and textures of all the 3D room model components such as tables, chairs, wall panels, carpet were further upgraded using built-in high definition post-processing features in Unity to obtain visually high quality materials. An example of the 3D room model in the Unity Game Engine is provided in Figure 3. The experimental interface was designed using the User Interface (UI) components from Unity as shown in Figure 4.

The UI components included buttons for stimuli presentation, recording the response of the participant, loading source signals and saving the response of the participants. Four sliders each corresponding to the four perceptual attributes were implemented as a bipolar continuous rating scale with anchors -50 to +50 at either ends. However, only the scale end labels were visible to the participants as shown in Figure 4. An additional slider, with a unipolar continuous scale ranging between 0 to 100 was implemented for the overall difference question. The behaviour of UI components was controlled by means of C# scripts compiled in Visual Studio Environment.

Figure 3: 3D model of Scenario 1 1.5m as seen in Unity Game Engine. The camera in the image (indicated by ”L” in red) is the position of the listener inside the virtual room when VR HMD is mounted.

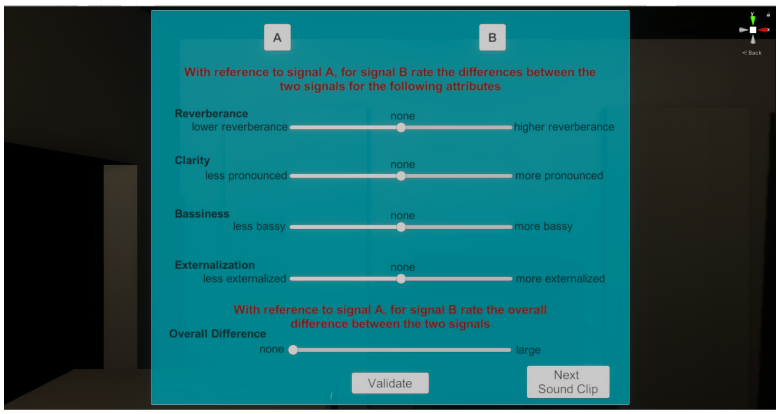


Figure 4: Experimental Interface as seen by the participants while wearing VR HMD

The C# scripts handled the following processes for Max/MSP and Unity respectively:

- Sending commands to Max/MSP for loading new sets of measured and simulated BRIRs, cleaning the GPU off the previously loaded BRIR set, randomising the order of source signals and loading a source signal from the list of source signals, playing the convolved measured and the convolved simulated signal.
- Loading a random 3D model of the room such that there was no mismatch between the BRIRs loaded into the memory and the corresponding 3D room model.

4 Materials and Method

4.1. Participants

Twelve participants (5 Female, 7 Male) took part in this study. The age of the participants ranged between 20-32 years. All participants reported that they were normal hearing listeners.

4.2. Source Signals and Audio Apparatus

Two types of source signals were used - percussion instruments and speech sentences by a female talker. Anechoic recordings of two different percussion instruments, Turkish drums, Darbuka and Bendir played with the same rhythm "dum-tek-tek" [18] and four anechoic recordings of speech sentences by a female talker [19] were used in the experiment. The duration of the signals was 3-6s. The source signals were re-sampled at 44.1 kHz. The audio playback was handled by Max/MSP. The output signal was routed through the Audient iD14 audio interface to Sennheiser HD 800 S headphones.

4.3. Procedure

The experiment was divided into four steps. Figure 5 shows the steps of listening experiment procedure.

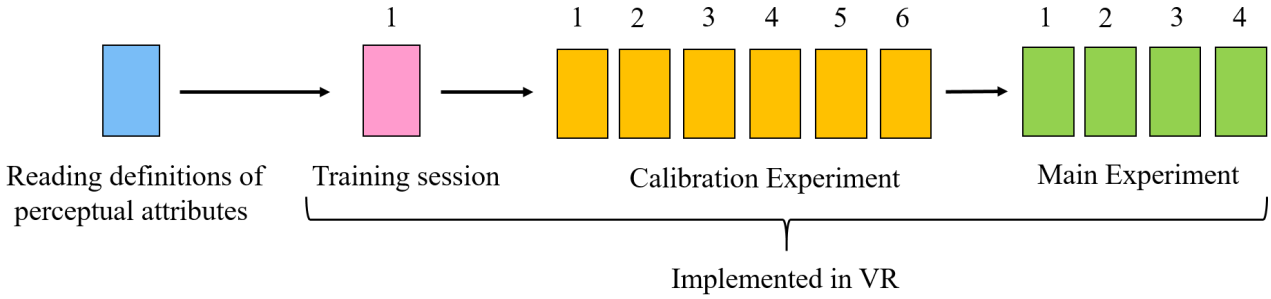


Figure 5: Graphical representation of the listening experiment procedure. The listening experiment consists of four steps. The numbers on top of the coloured blocks indicate the number of sessions in each experiment. The first step involves participants reading the definitions of the four perceptual attributes. The second step was one training session in VR. The third step was a calibration experiment composed of 6 sessions and the last experiment was the main experiment having 4 sessions.

The listening experiment was carried out in a double-walled sound insulated listening booth. The participants were first informed about the general nature of the experiment involving sound quality evaluation in Virtual Reality for four perceptual attributes - reverberance, clarity, bassiness and externalization. Before they began the experiment, they read the definitions of the four attributes. The definitions of the attributes were modified compared to mentioned in [10] to be understood by naive listeners. In case the definition of the attributes was not clear to the participants, the experimenter elaborately explained the meaning of the attributes through examples. After the participants read the definition of the attributes, a short training session was administered before the commencement of the experiment. The purpose of the training session was to familiarise the participants with the task involved in the experiment and to get them accustomed to the VR environment. Oculus Rift VR was used during this study. They first mounted the Head Mounted Display (HMD) and then the headphones. Throughout the experiment, the primary button of the Oculus Rift right hand touch controller was used by the participants to provide their response.

After the training session, a calibration experiment was conducted prior to the main experiment. The motivation for the design of calibration experiment was to understand how the participants map the meaning of perceptual attributes onto the rating scale used in the experiment. The calibration experiment was an audio only session implemented in VR. The 3D models of the rooms were eliminated. However, for the sake of brevity the data from calibration experiment has not been presented and discussed in this paper.

The experiment consisted of four sessions. Within each session, participants were presented with 6 source signals. The order of the source signals and experiment sessions was randomised throughout the experiment. Each session in the experiment represented one audio-visual condition corresponding to one of the four room acoustic scenarios as shown in the Table 1.

Table 1: Table showing the audio-visual sessions implemented in the listening experiment

Sessions	Signal A (Convolved measured signal)	Signal B (Convolved simulated signal)
1	Scenario 1 – 1.5m	Scenario 1 – 1.5m
2	Scenario 1 – 3.5m	Scenario 1 – 3.5m
3	Scenario 2 – 1.5m	Scenario 2 – 1.5m
4	Scenario 2 – 3.5m	Scenario 2 – 3.5m

The participants began the experiment by pressing 'Start Experiment' button on the interface. A 3D model of one of the four room acoustic scenarios was generated in the VR headset such that the participant was placed in the virtual room. Simultaneously, the measured and simulated BRIRs corresponding to the room acoustic scenario were loaded into the computer's memory. The process of loading BRIRs took around 110 seconds. During this time, the participants were prompted to wait by a display message on the interface until the files were loaded in the memory. They were asked to have a look at the virtual room in which they were placed or to take a break by taking off the VR HMD. Once the files were loaded, a 'start' button appeared on the interface. The participants started the session by pressing the button.

The method adopted in the listening experiment was an A/B testing method in which Signal A was the reference signal, calculated by convolving the source signal with the measured BRIRs and signal B was calculated by convolving the source signal with the simulated BRIRs as mentioned in Table 1. The participants were not informed about the nature of the signals A and B. They were instructed to provide their response for signal B with signal A as the reference. Their task was to rate the differences between the two signals A and B for the four perceptual attributes and an additional question regarding the overall difference between the two signals.

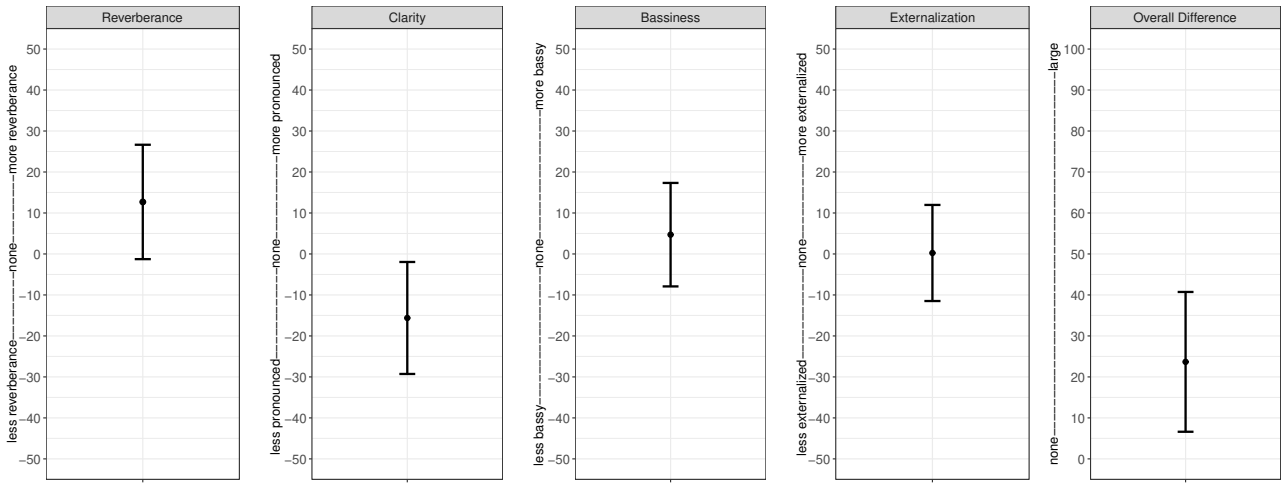
The participants listened to signals A and B as many times as they would like. They provided their response by dragging the slider for each of the attributes. The participants pressed the Validate button on the interface to save their response. After which they would press the Next Sound signal button on the interface. This button would load a different source signal to be convolved with the BRIRs.

After the completion of each session, the participants pressed the Next session button. By pressing next session, the participants were placed in another virtual room (3D model) and the BRIR set corresponding to the visual condition was loaded into the system. This process happened until the participants completed all four sessions. All the participants performed the rating task for all source signals and for all four sessions in the experiment only once. In total there were, 4 room acoustic conditions x 6 source signals = 24 observations per participant. The approximate duration of the experiment was 30 minutes.

5 Results

For each of the four perceptual attributes, a direct difference between the two signals, A (measured) and B (simulated) was obtained from the rating task. Figure 6 shows the judged difference between the measurement and simulation for each of the four perceptual attributes. Each sub-figure provides the mean response combined over all four scenarios for each perceptual attribute. The Y-axis of each sub-figure shows the scale labels for the four attributes as implemented in the listening experiment. A positive value on the Y-axis represents a higher rating for the simulation than for the measurement. It can be observed from Figure 6a that the perceived reverberance was judged higher for the simulation than for the measurement. From Figure 6b, it can be seen that the clarity in the stimuli was perceived to be less clear or less pronounced for the simulation than for the measurement. The perceived bassiness was judged higher for the simulation than for the measurement as shown in Figure 6c. For the externalization attribute based on the VR experience, the simulation was rated to be slightly more externalized than the measurement as seen in Figure 6d. Lastly, from Figure 6e the overall

difference between simulation and measurement was not perceived as large but a perceptual difference between the two stimuli was still observed.



(a) (b) (c) (d) (e)

Figure 6: Mean of the difference rating between measurements and simulations for each of the four perceptual attributes ($Mean \pm s.d.$).(a) Reverberance (13.35 ± 13.96), (b) Clarity (-15.62 ± 13.66), (c) Bassiness (4.71 ± 12.62) (d) Externalization (0.24 ± 11.72) and (e) Overall Difference (23.76 ± 17.05). Error bars indicate standard deviation.

Statistical analysis was performed using R Statistical Software. One sample Wilcoxon signed rank test, a non-parametric test, was performed as the responses for each of the attributes was found to be not normally distributed. The analysis revealed that the mean difference between measurement and simulation was found statistically significant for reverberance ($V = 27695$, $p < 0.001$), clarity ($V = 974$, $p < 0.001$), bassiness ($V = 2468$, $p < 0.001$) but not for externalization ($V = 10456$, $p = 0.7216$). The overall difference between the simulation and measurement was found to be statistically significant ($V = 40755$, $p < 0.001$).

6 Discussion

The VR integrated dynamic convolution framework explained in section 3 provides the listener with a 3D simulated audio-visual environment. The dynamic convolution framework is independent of the number of BRIR filters allowing to feed BRIRs of a desired resolution. For this study, we used 72 BRIR filter pairs (360°/5°), 5°resolution for measurements and simulations. A finer resolution between the BRIR filter pairs might prove useful in perceptual sensitivity studies.

For attribute reverberance, the perceptual judgement results shown in Figure 6a are in fairly well agreement with the objective results for T_{20} Figure 1.

For the attributes clarity, the results shown in Figure 6b are in line with the inverse relationship between reverberation time and clarity [20]. This opposite trend can also be observed in Figure 6a and Figure 6b showing that the reverberant aspect of the simulation resulted in lack of its perceived clarity.

The bassiness attribute was judged higher for the simulation than for the measurement (Figure 6c). The T_{20} results from Figure 1 for octave bands 125 Hz and 250 Hz are in well agreement with perceptual results for all room acoustic conditions except for 125 Hz band of Scenario 2 3.5m (see Figure 1d where there is more energy in the 125 Hz band for the measured IR). The inherent bassy trait of the percussion instruments, Bendir and Darbuka, is possibly responsible for more perceived bassiness in the simulation than the measurement.

Externalization rating were not found to be statistically significant from zero. This finding tells us that for one of

the perceptual attributes the simulation was perceptually perceived close to the measurement (see Figure 6d). A review study on sound externalization provides evidence that the phenomenon is influenced by various cues such as binaural cues (Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs)), reverberation in the stimuli, visual cues and head-movements [21]. The measured and simulated stimuli used in this study both encompass identical source-directivity information and were reverberant in nature. They were dichotically presented to the participants in a VR environment capturing natural head movements of the listener. The fact that in both the measured and the simulated stimuli the cues required for sound externalization were preserved, may explain that the simulations were judged to be perceptually close to the measurements for the externalization attribute.

Lastly, the overall difference was found statistically significant (see Figure 6e). The overall difference refers to any perceptually noticeable differences including the four perceptual attributes that could be involved in rating the difference between the simulated vs. the measured signals.

7 Conclusions and Further Work

In this study, a listening experiment was conducted in VR for subjective evaluation of the auralizations from a wave-based method. The findings from the listening experiment show that the wave-based simulations were found perceptually indistinguishable from the measurements for the externalization attribute. The simulations and measurements were found to be perceptually distinguishable from the measurements for the attributes reverberance, clarity and bassiness, the results from the listening experiment are in agreement with the objective results of reverberation time T_{20}. Future work focuses on examining results from the calibration experiment to understand how listeners used the scale implemented in the listening experiment. The experimental design presented in this study will be extended for perceptual evaluation of auralizations using computationally efficient room acoustic modelling methods. Lastly, investigations on material properties of objects used to create the two room acoustic scenarios are still ongoing to obtain a better match between measurements and simulations.

Acknowledgements

We would like to thank fellow researchers of the Building Acoustics Group Wouter Wittebol and Baltazar Briere de la Hosseraye for providing simulation data and objective results. Alessia Milo for trouble-shooting complexities during the design of dynamic convolution framework.

References

[1] Brian Hamilton and Stefan Bilbao. FDTD methods for 3-d room acoustics simulation with high-order accuracy in space and time. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25 (11):2112–2124, 2017. doi: 10.1109/TASLP.2017.2744799.

[2] Jelle Van Mourik and Damian Murphy. Explicit higher-order fdtd schemes for 3d room acoustic simulation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12):2003–2011, 2014.

[3] Huiqing Wang, Indra Sihar, Raúl Pagán Muñoz, and Maarten Hornikx. Room acoustics modelling in the time-domain with the nodal discontinuous galerkin method. *The Journal of the Acoustical Society of America*, 145(4):2650–2663, 2019.

[4] Matthias Blau, Armin Budnik, Mina Fallahi, Henning Steffens, Stephan D Ewert, and Steven Van de Par. Toward realistic binaural auralizations–perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario. *Acta Acustica*, 5:8, 2021.

[5] Fabian Brinkmann, Lukas Aspöck, David Ackermann, Steffen Lepa, Michael Vorländer, and Stefan Weinzierl. A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*, 145(4):2746–2760, 2019.

[6] Barteld NJ Postma and Brian FG Katz. Perceptive and objective evaluation of calibrated room acoustic simulation auralizations. *The Journal of the Acoustical Society of America*, 140(6):4326–4337, 2016.

[7] Finnur Kári Pind Jörgensson, Cheol-Ho Jeong, Hermes Sampedro Llopis, Kacper Kosikowski, and Jakob Strømann-Andersen. Acoustic virtual reality – methods and challenges. In *Proceedings of BNAM 2018*, 2018. Baltic-Nordic Acoustics Meeting 2018, BNAM 2018 ; Conference date: 15-04-2018 Through 18-04-2018.

[8] Hermes Sampedro Llopis, Finnur Pind, and Cheol-Ho Jeong. Development of an auditory virtual reality system based on pre-computed b-format impulse responses for building design evaluation. *Building and Environment*, 169:106553, 2020.

[9] Imran Muhammad, Michael Vorländer, and Sabine J Schlittmeier. Audio-video virtual reality environments in building acoustics: An exemplary study reproducing performance results and subjective ratings of a laboratory listening experiment. *The Journal of the Acoustical Society of America*, 146(3):EL310–EL316, 2019.

[10] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman, and Stefan Weinzierl. A spatial audio quality inventory (saqi). *Acta Acustica united with Acustica*, 100(5):984–994, 2014.

[11] Fotis Georgiou, Baltazar Briere de la Hosseraye, Maarten Hornikx, and Philip W. Robinson. Design and simulation of a benchmark room for room acoustic auralizations. Proceedings of the International Congress on Acoustics, pages 723–730, September 2019. doi: 10.18154/RWTH-CONV-239684.

[12] Ramona Bomhardt, Matias de la Fuente Klein, and Janina Fels. A high-resolution head-related transfer function and three-dimensional ear model database. In *Proceedings of Meetings on Acoustics 172ASA*, volume 29, page 050002. Acoustical Society of America, 2016.

[13] Maarten Hornikx. Reconstruction of binaural room impulse responses using spherical harmonics. In *23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019 (ICA2019)*, 2019.

[14] Unity game engine. https://unity3d.com/unity/whats-new/2019.3.15, 2019.

[15] Max/msp/jitter. https://cycling74.com/, 2019.

[16] Øyvind Brandtsegg, Sigurd Saue, and Victor Lazzarini. Live convolution with time-varying filters. *Applied Sciences*, 8(1):103, 2018.

[17] Sketchup pro 2020. https://www.sketchup.com/node/4446, 2020.

[18] Sound examples of percussion detection. http://users.spa.aalto.fi/ajylha/percussion/.

[19] P Demonte. Harvard speech corpus—audio recording 2019. *University of Salford Collection*, 2019.

[20] Gilbert A Soulodre and John S Bradley. Subjective evaluation of new room acoustic measures. *The Journal of the Acoustical Society of America*, 98(1):294–301, 1995.

[21] Virginia Best, Robert Baumgartner, Mathieu Lavandier, Piotr Majdak, and Norbert Kopčo. Sound externalization: A review of recent research. *Trends in Hearing*, 24:2331216520948390, 2020.