VPS VIBRATIONS IN PHYSICAL SYSTEMS

DOI: 10.21008/j.0860-6897.2022.1.12
Vibrations in Physical Systems, 2022, 33(1), 2022112

Virtual Reality Environments for Soundscape Research

Anne HEIMES[1], Ming YANG[2], Michael VORLÄNDER[1]

[1] IHTA - Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany
[2] HEAD acoustics GmbH, Herzogenrath, Germany

Corresponding author: Ming YANG, email: mingkateyang@163.com

Abstract Soundscape research demands a holistic approach for the analysis of environments, yet any research method, e.g. soundwalk or lab-based listening test, has its advantages and limitations. The virtual reality (VR) and augmented reality (AR) technology provides an alternative method for soundscape research, which may retain as much context as possible while enabling control conditions. The current phase of this research project focuses on the ecological validity of virtual reality environment for soundscape, by creating virtual environments for comparison of human experiences and responses in a real soundscape and three-dimensional virtual ones. The workflow of scene rendering on different levels of detail is described. The results are made publicly available.

Keywords: virtual reality, soundscape, soundscape recording, auralization.

1. Introduction

Virtual reality (VR) and augmented reality (AR) have become more popular and stable during the last years. The applications span a very wide range including human factors, cognitive science, training and rehabilitation, engineering and product design, cultural heritage, and entertainment [1]. VR and AR also become an important tool in research areas, such as psychology and neurobiology, due to their flexibility and predictability for creating environmental stimuli and conditions [2].

Soundscape is a complex construct; it involves the auditory perception of everything that makes up our physical environment, multisensory/multimodal information integration, other context factors – such as people's activities, mind states, expectations, memories, personalities, cultural backgrounds, and especially the complex interaction effects of all the different elements/factors audio-visual interaction [3]. Thus, soundscape research demands a holistic approach [4], rather than methods that consider only single or a few factors. Regarding soundscape research methods, the traditional survey/experiment methods have their limitations. VR and AR provide the possibilities of an alternative research method, which might allow the combination of the advantages of classic research methods including in-situ soundwalk and lab-based listening test. It enables the control over the environmental stimuli experienced by participants while preserving as much as possible the context.

There are a number of ways for creating virtual soundscapes. Sun et al. [5] used immersive audio-visual recordings collected at urban sites, which combined 360-degree video with spatial audio, for the presenting soundscapes. This method presents virtual scenes closest to the real environments. However, it does not enable the changes to the elements or conditions of the virtual visual and acoustic environments. Jiang et al. [6] demonstrated a fully synthetic environment, with both the simulated visual environment from 3D models of objects and simulated acoustic environment by spatial rendering of individual sound source recordings. Such simulated VR scenes provide the greatest flexibility to control or change any environmental stimuli/components. Using AR technology, Aalmoes et al. [7] superimposed virtual objects and cues (aircraft in their study) upon audio-visual recordings, by mixing a 3D aircraft model flying aircraft trajectories with 360 degrees video recordings made in the community and simulated aircraft sound with ambient noise recordings. It enables the addition of the objects/elements to the recorded scenes while retaining the high degree of realism.

Our research project will compare these different methods for creating virtual soundscapes (audio-visual recordings, simulated VR scenes, AR scenes based on recordings, and traditional audio-only recordings), by comparing human experiences and responses in these virtual soundscapes to those in real ones, to study the ecological validity of these virtual reality environments for soundscape.

In the first step of this research study, we thus designed a number of different virtual scenes according to a real environment, based on the above methods, to compare with the real situation. These include one original recorded audio-visual scene, three VR scenes – a VR scene including a rebuilt virtual scene in high visual quality, in middle visual quality and a VR scene with very simple geometry in one colour, one AR scene that is based on the recorded scene with artificial modifications, and finally one scene without any visual cues is presented. In this way a comparison of audio-visual scenes with audio-only ones is done to study the effects of visual cues on soundscape. Additionally, different levels of detail of virtual reality scenes are compared to investigate the human response to the differences. In total, there are six different virtual scenes. Our research study will examine the influence of the authenticity of virtual audio-visual scenes on human responses to the environment to search for the threshold of complexity for reproducibility.

In this paper, focusing on the first stage (the creation of virtual scenarios) of the research project, we will present the methods for creating the different virtual environments, in terms of both virtual visual and acoustic environments, for the proposed soundscape research.

2. Virtual scenarios

2.1. Scenario site

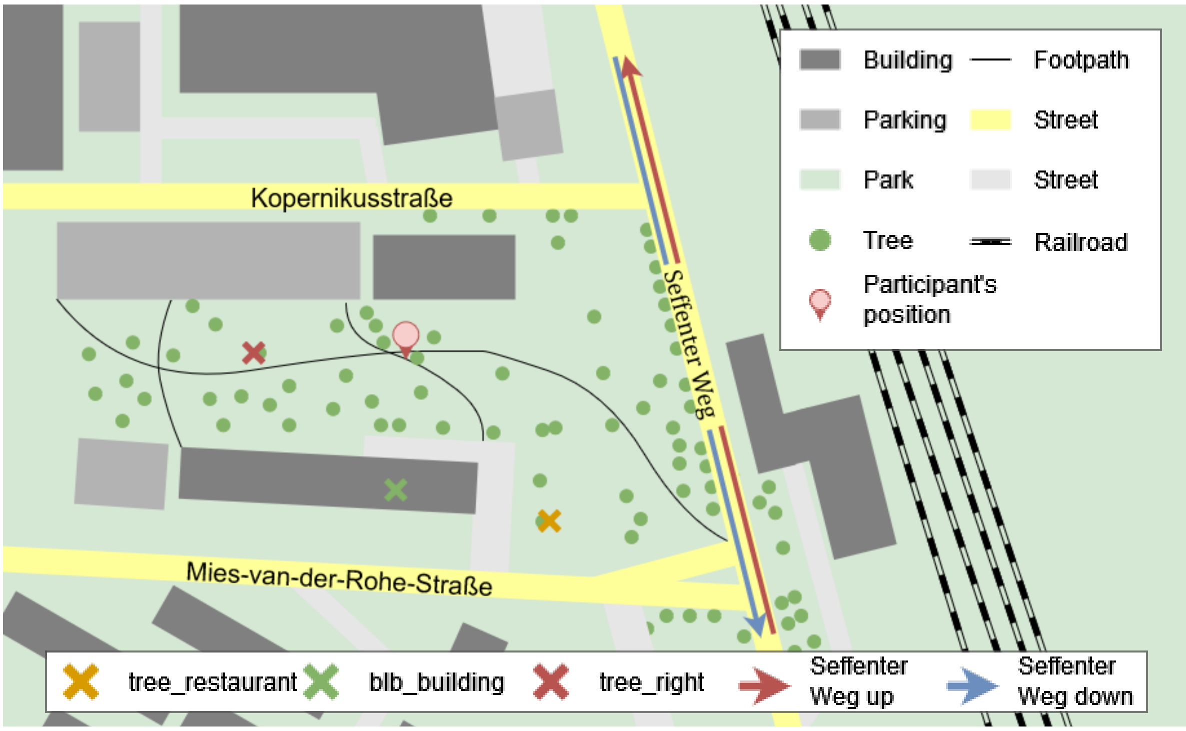

Figure 1. Map of the scene where the experiment is conducted. The orange dot is marking the place where the participant is sitting. The arrow is marking the default view direction of the participant.

To test the stimulation of virtual environments with a variety of frequently heard sound sources in soundscapes of daily life, we chose the small park behind the Institute for Hearing Technology and Acoustics (IHTA) of RWTH Aachen University (Kopernikusstraße 5, 52074 Aachen, Germany) as the experiment site, considering the large availability of such sound sources. The park is surrounded by streets, especially the Seffenter Weg and the Mies-van-der-Rohe-Straße have medium traffic including frequent bus traffic. Additionally, there is a railway track nearby, where regular passenger and cargo trains are passing by. In the south-west is a street for walking and behind it is a garden area. Around Kopernikusstraße are more university buildings and a street with rare car traffic. In the park, the different sound sources and activities mainly include birds, wind through the leaves, and people in the park, for example they are walking through the gravel footpath.

In this park the listening experiment takes place and accordingly the virtual scenes are created. The orange circle marks the position where the participant sits on a park bench during the experiment, whereas the red arrow marks the view position.

2.2. Original recorded scenario

The scene was recorded in November 2020 with an Insta360 Pro 2, which supports 3D 360° special video recording. The video was recorded with a resolution of 3200 x 2400 for each eye of the six eyes. With the help of the Insta360 Sticher the single videos can be rendered to full 3D 360° Videos. The highest possible resolution is 12k, whereas we choose a resolution of 4k to maintain a stable playback and a good resolution. After that, the videos are post-processed in Adobe After Effects to blur the faces of people presented for de-identification. The audio was captured by a Zoom H3-VR 1st order Ambisonics microphone. Figure 2 shows the recording setting.

Figure 2. Recording Setting in the park, with an Insta360 Pro 2 and a Zoom H3-VR 1st order Ambisonics microphone.

From the measurement, two different settings were selected. First, a scene as shown in Figure 3. Notice that this picture is an equi-rectangular 360° picture. In this scene, background sources as wind in leafs, background noise from the traffic nearby, bird sounds and a person crossing the way are audible and visible. The second recording was almost the same but emptier without a person crossing the way. The first recording is presented as its original in the listening experiment later, whereas the second recording is used for further processing to create the AR scenario as described in Section 2.5. The first recording is also used as a reference for visual and audible events for the virtual scenes. The final video has a length of 5.5 minutes.

Figure 3. 360° Picture of the recorded scene.

2.3. VR scenario at high resolution

We create the VR scenarios according to the field recording. The left building in Figure 4 is taken from [9], the other buildings were created in SketchUp and imported to Unity 3D. The environment is rebuilt in a high resolution to be as close as possible to the real scene. We use Unity 3D to render one scene, as it is shown in Figure 4. Furthermore, we use Unreal to render a scene with higher resolution, as it is shown in Figure 5. All virtual buildings and objects have the same dimension as the real ones.

Before the auralization, all the sound sources in the field recording are annotated by the authors, through investigating the video as well as the audio. A list was created and used as a script for the virtual scenes. The detailed list of all sound sources at which time can be found in Table 1. The Locations are marked on the map in Figure 1.

Table 1. List of all sound sources and events in the recorded Scene.

Start time	End time	Source type	Location	Visibility
00:00	00:15	car	Seffenter Weg up	no
00:00	02:55	bird	tree_restaurant	no
00:00	05:30	air conditioner	blb_building	yes
00:36	00:45	car	Seffenter Weg down	white
01:25	01:38	car	Seffenter Weg up	black
01:27	01:40	car	Seffenter Weg up	silver
01:56	02:25	bird	tree_right	no
02:10	02:23	car	Seffenter Weg down	black
02:12	02:25	car	Seffenter Weg up	white
02:31	02:55	car	Seffenter Weg down	black
02:46	03:10	bird	tree_right	no
03:06	03:15	car	Seffenter Weg up	white
03:15	03:25	motorbike	Seffenter Weg down	motorbike
03:26	03:35	car	Seffenter Weg down	black
03:36	04:25	bird	tree_right	no
04:16	04:25	car	Seffenter Weg up	black
04:35	04:46	car	Seffenter Weg down	black
04:37	04:48	car	Seffenter Weg down	black
04:39	04:50	car	Seffenter Weg down	white

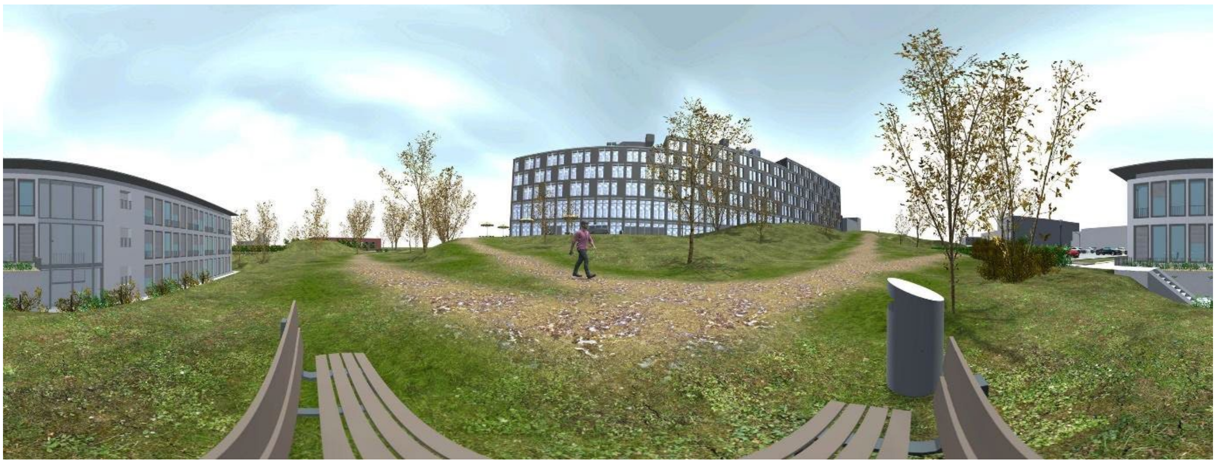

Figure 4. 360° Picture of the VR scene at medium resolution in Unity 3D.

The auralization is made by using "Virtual Acoustics" (VA) from Institute for Hearing Technology and Acoustics (IHTA), RWTH Aachen University [8], in which an outdoor geometrical acoustics model is used [10]. This model includes diffraction and reflection up to a certain order as well as doppler shifts and moving sound sources. Different options are available, the first option is the real-time approach, where all sound paths are pre-calculated before and the final auralization takes place in real-time. In our case, all positions and orientations of the sources are known before, for the listener the position is also fixed. This gives us the option to simulate everything beforehand and get an Ambisonics file, where all sounds are included. We choose the second option to reduce the computational time during the experiment and to increase the order of the reflection and diffraction for the final auralization.

For the auralization just dry recordings are used, such as birds, steps on gravel and wind noise.

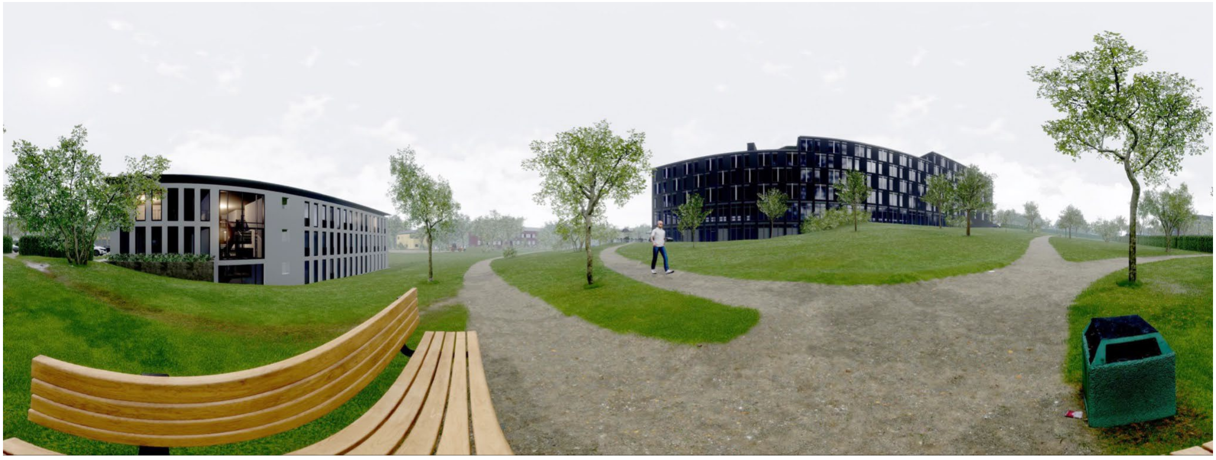

Figure 5. 360° Picture of the VR scene at high resolution in Unreal Engine.

2.4. VR scenario with simplified visual cues

In this scene, the visual cues are reduced compared to the VR scenario in section 2.3. The audio is modelled in the same way. The landscape is simple, and the buildings are monochrome blocks, in this way the visual cues are reduced to a minimum. An image of this scene is shown in Figure 6.

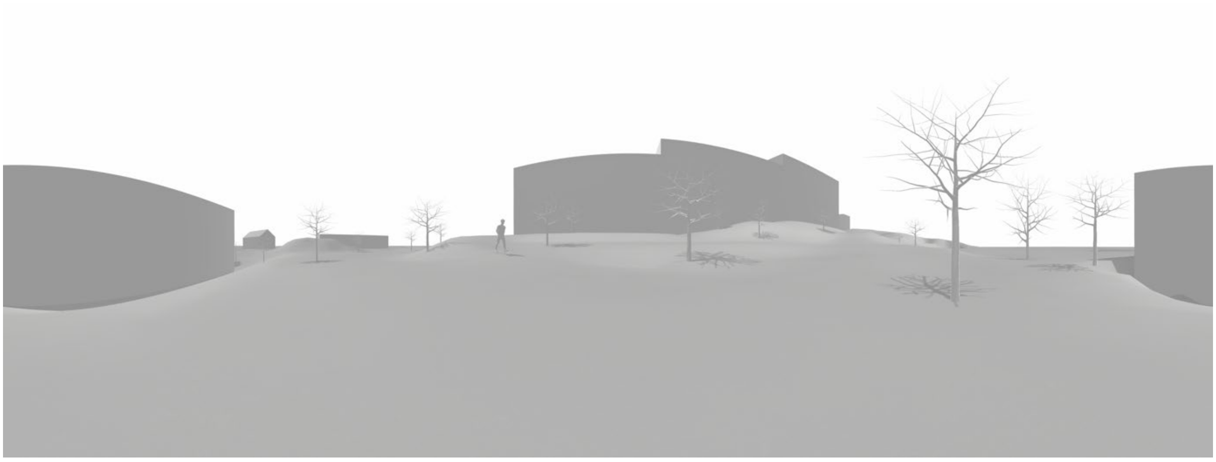

Figure 6. 360° Picture of the VR scene at low resolution.

2.5. AR scenario

The AR scenario is a mixture of the recorded scene including the video and the Ambisonics recording from Section 2.2 and an additional source, in our case only one walking man, added visually and acoustically. . The walking person passed by after 30 seconds in the whole duration of 5.5 minutes. In this way, in each of the virtual environment scenes (three VR and one AR scenes), we include one man avatar at the same occurrence time as that in the video recording, to make the conditions consistent.

For audio, the real Ambisonics recording from this scene is used and the auralised sound of the walking person is added to the recording. The auralisation was done as described in Section 2.3. The visual cues of a walking person were created and rendered in Unity 3D, and added to the video recording, with the help of Adobe After effects. An image of this scene is shown in Figure 7.

Figure 7. 360° Picture of the AR scene.

2.6. Audio scenario with no visual cues

Finally, a scene without visual cues is presented. The scene is simply black, and the sound is simulated as described in section 2.3.

2.7. Demonstration and Data presentation

For demonstration purposes, all scenes are rendered as a 360° 3D Video and 360° Video (for better usage). All created videos are available for open access under the link https://git.rwth-aachen.de/ita/ihta-soundscape-environment.

3. Current progress and further work

Within-subjects experiment design will be used to compare the participants' experiences in the real and virtual environments, via a set of listening experiments and soundwalk at a fixed position. The five virtual scenes will be compared in terms of participants' experiences and responses in them, and also to the real soundscape in-situ.

The participants' experiences will be collected by self-report subjective evaluations, via questionnaire. The questionnaire includes a set of items/questions on perceived emotional/affective responses of the environment according to soundscape standard ISO/TS 12913-3 [11] and the previous research studies in soundscape, environmental psychology, and emotion, such as 'annoying', 'chaotic', 'eventful', 'monotonous', 'uneventful', 'vibrant', 'pleasant', 'relaxing', 'irritating', 'calming', 'frightening', 'tiring', 'boring', 'stressful', 'joyful', and 'exciting', using a 5-point categorical scale ('not at all', 'slightly', 'moderately', 'very', and 'extremely', coded from 0 to 4).

We include only participants who are very familiar with the chosen site, for the reason that the order of the real and virtual scenarios in the comparison experiments would have great impacts on the result if the participants are not familiar with the environment. Participants who are not familiar would have different feelings/evaluations of the environment when experiencing it for the first time and the times after. It thus restricts our recruiting of participants, from only the students and staff of IHTA, RWTH Aachen University. Twenty-one participants (16 males and 5 females, aged from 24 to 70 (mean 33.7)) took part in the experiment. Written informed consent was obtained from every participant.

3.1. Soundwalk

To compare with the experiences in virtual environments, the participants experienced the real outdoor environment of the site, the park outside the IHTA building. Each participant took part in this on-site test individually. Guided by the experiment conductor, each participant came to the site, sat at the fixed position in the park, experienced the environment/soundscape for 2 minutes, and then gave their emotional evaluations of the environment by answering the questionnaire on a tablet. The on-site test of each participant took about 15 minutes. The whole on-site test of all participants was completed in one week, so the objective visual/acoustic environment and weather conditions were similar for all the participants.

The test results are shown in Table 2, including the frequency table as well as mean and standard deviation (std) values of each of the rating items. The results showed that the real environment was rated as relaxing (mean = 2.810, std = 0.873), calming (mean = 2.619, std = 0.865), uneventful (mean = 2.333, std = 0.966), and pleasant (mean = 2.095, std = 1.091), of which mean rating values larger than 2 are presented here. Also, the standard deviations (std) of the ratings of the items of all the participants were not high (0.9 to 1.1), which suggested that the participants had quite accordant ratings on the environment. These results serve as baseline and documentation of the real reference of the VR and AR scenes.

Table 2. Rating results of the soundwalk.

	Frequency						
	not at all	slightly	moderately	very	extremely	mean	std
annoying	8	12	1	0	0	0.667	0.577
relaxing	0	1	7	8	5	2.810	0.873
eventful	3	11	6	1	0	1.238	0.768
pleasant	3	1	9	7	1	2.095	1.091
vibrant	6	6	6	3	0	1.286	1.056
uneventful	1	3	6	10	1	2.333	0.966
monotonous	4	7	6	4	0	1.476	1.030
chaotic	17	3	0	0	1	0.333	0.913
irritating	17	3	1	0	0	0.238	0.539
depressing	15	6	0	0	0	0.286	0.463
calming	0	3	4	12	2	2.619	0.865
frightening	20	0	0	1	0	0.143	0.655
tiring	14	6	1	0	0	0.381	0.590
boring	11	3	7	0	0	0.810	0.928
stressful	18	3	0	0	0	0.143	0.359
joyful	1	4	13	3	0	1.857	0.727
exciting	12	7	1	1	0	0.571	0.811
lively	1	9	6	5	0	1.714	0.902
distressing	14	4	2	1	0	0.524	0.873
interesting	1	6	10	4	0	1.810	0.814
unpleasant	19	1	1	0	0	0.143	0.478

3.2. Laboratory listening experiment

A set of laboratory-based listening experiments will be carried out to collect participants' experience in the virtual environments, as soon as listening tests in presence in the laboratory are allowed in the context of pandemic hygiene rules.

The six different scenarios will be reproduced in Unity 3D with a head mounted display (HMD Vive Eye Pro), with the sound being reproduced by VA [8] via equalized headphones. The six scenarios will have a randomized order in the experiment.

The same as that on-site, each participant will come to the lab individually and sit in a chair in the lab (virtually at the fixed position in the park). They will experience the environment/soundscape for 2 minutes and then answer the questionnaire to give their emotional evaluations of the environment. Since the participants will use the HMD, the question/answer will be embedded in the virtual environment similar to a page on a tablet or a paper sheet. Each of the scenarios have a length of 5.5 minutes; the first 2 minutes of each of the scenarios are mandatory for the participants to listen. After the first 2 minutes the participants will start the evaluation progress (in the virtual environment), while the rest of each of the scenarios is continually presented the same as that in the real on-site situation.

4. Summary

In this research study, we designed five virtual audio-visual scenes according to a real environment, with a range of degrees of authenticity, using the different VR/AR methods. The workflow and its documentation is described and published as a set of video demonstrations (https://git.rwth-aachen.de/ita/ihta-soundscape-environment). In future our research study will compare the different methods and also to the real environment, on participants' subjective evaluations of soundscapes. This paper presents the methodology and current progress of the research study.

The results will help future research studies to choose the suitable method for creating virtual environments, based on the ecological validity and specific requirements of research design. The final results will contribute to the development of new methodology of soundscape research and to the potential applications in soundscape design and soundscape management.

Acknowledgments

This research study is funded by the HEAD-Genuit-Foundation under the project numbers P-19/03-W and P-20/02-W. The authors would like to thank Prof. Klaus Genuit, Prof. Brigitte Schulte-Fortkamp, Dr. Lukas Aspöck, and Dr. Andreas Herweg for the valuable discussions. The authors would like to thank Imran Muhammad, M.Sc. for the detailed IHTA Building in the virtual model and Yinyun Chen for her contribution in finalizing the virtual scene.

Additional information

The authors declare no competing financial interests.

References

1. K.S. Hale, K.M. Stanney K. M. (eds.); Handbook of virtual environments: Design, implementation, and applications; CRC Press, 2014.
2. P. Cipresso, I. Gigliol, M. Ray, G. Riva; The past, present, and future of virtual and augmented reality research: A Network and Cluster Analysis of the Literature; Front Psychol., 2018, 9, 2086. DOI:10.3389/fpsyg.2018.02086.
3. B. Schulte-Fortkamp, B.M. Brooks, W.R. Bray; Soundscape: An approach to rely on human perception and expertise in the post-modern community noise era; Acoustics Today, 2007, 3(1), 7–15.
4. International Organization for Standardization, ISO/TS 12913-2:2018 Acoustics-Soundscape-Part 2: Data collection and reporting requirements, 2018.
5. K. Sun, D. Botteldooren, B. De Coensel; Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation; Proc. 47th Int. Congr. Expo. Noise Control Eng., 2018, 1–10.
6. L. Jiang, M. Masullo, L. Maffei, F. Meng, M. Vorländer; A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment; Landsc Urban Plan, 2018, 170, 276–82. DOI:10.1016/j.landurbplan.2017.09.007.
7. R. Aalmoes, M. Boer, H. VeerbeekL; Virtual reality aircraft noise simulation for community engagement. INTER-NOISE: 47th International Congress and Exposition on Noise Control Engineering, 2018.
8. J. Stienen, F. Wefers, M. Vorländer; The open-source Virtual Acoustics (VA) real-time auralization framework; 23nd Int. Congr. Acoust. ICA 2019 Aachen Ger., 2019.
9. M. Imran, A. Heimes, M. Vorländer; Real-time building acoustics noise auralization and evaluation of human cognitive performance in virtual reality; Tagungsband - DAGA 2019 45 Jahrestag. Für Akust. 18-21 März 2019 Rostock, Berlin: Deutsche Gesellschaft für Akustik e.V. (DEGA), 2019, 551–4.
10. A. Erraji, J. Stienen, M. Vorländer; The image edge model; Acta Acust., 2021, 5, 17. DOI:10.1051/aacus/2021010.
11. International Organization for Standardization, ISO/TS 12913-3:2019 Acoustics-Soundscape-Part 3: Data Analysis, 2019.