CONCERT HALLS IN VIRTUAL REALITY – FROM AUDIO-VISUAL PERCEPTION TO A COMMON VOCABULARY

J Heck RWTH Aachen University, Institute for Hearing Technology and Acoustics, Aachen, Germany
J Llorca-Bofí RWTH Aachen University, Institute for Hearing Technology and Acoustics, Aachen, Germany
Universitat Politècnica de Catalunya, Centre for Qualitative Urban Ideation and Simulation (QURBIS), , Spain
M Vorländer RWTH Aachen University, Institute for Hearing Technology and Acoustics, Aachen, Germany

1 INTRODUCTION

Within the research project *Person-focused Analysis of Architectural Design*, it is a goal to elaborate a method which categorizes architectural design based on vocabulary describing audio-visual perception of built environments. The first step is to understand which dimensions are of importance to assess and verbalize the perception. In other words, for architects, acousticians and designers of modern concert halls remains one key question when proposing a new design: how will people perceive the space? To develop an understanding that closes the gap between experts who design or optimize the space and those who use and visit it, we propose to collect perceptual and descriptive vocabulary.

This poster depicts the laboratory setup as well as the experimental design to develop individual audio-visual descriptors for concert halls. For this purpose, different concert halls are studied at several seating positions, following the procedure proposed by Lokki et al. 2011 [1]. The sound stimuli – symphonic music pieces – are presented through headphones that are dynamically tracked to account for head rotations. The visual stimuli are 360° representations of the halls which are presented, e.g., through a head-mounted display. Graphical user interfaces in the virtual scenes allow participants to select attributes and to perform Mushra-like ratings comparing the different seating positions and concert halls.

2 AURALIZATION

The framework used for auralization is Virtual Acoustics (VA) [2]. It offers interfaces to multiple inputs such as the playback of pre-rendered sound files as well as various source models. Then, it provides different rendering options and sound propagation models, especially the room acoustics simulation software RAVEN [3] which allows for real-time renderings. Finally, it feeds loudspeaker arrays in the desired configuration or binaural output.

In the presented setup, we use dynamically adapted binaural playback via headphones which accounts for head rotations. A set of generic head-related and headphone-related transfer functions are used to equalize the audio stream. Recorded impulse responses are used in higher-order Ambisonics format (currently up to 4th order) to be convolved with anechoic recordings of orchestra instruments playing symphonic music pieces.

3 VISUALIZATION

3.1 Capture

Each room has been captured from each sitting position under study. The capture was done with a Kandao Obsidian S 360 3D camera. The camera position approximates the view position of a masculine adult sitting on the concert hall chairs. The camera height corresponds to the lens center. This is 80 cm above the seat level, and a total hight of 80+x cm above the floor level, depending on the chair’s design. The total camera height varies inside each room and across rooms and ranges between 105 and 120cm.

Figure 1: 360° capture of a concert hall (Cologne philharmonic hall).

3.2 Color calibration and postprocessing

The camera settings were close to a reference setting as follows: ISO = 200; Shutter time = ½ sec; Aperture = F/2.4; Temperature = 3800. This setup keeps the image levels range very close to the dynamic range of the camera for all seating positions. However, each photo was taken with slight variations to the reference settings to fit the actual dynamic range. It was decided to keep ISO and Aperture values fixed and only vary shutter time because this parameter has an impact on the brightness of the photo, but do not affect directly the colors [5, 6]. In the event of fluorescent or LED light flicker at a rate of 50 Hertz a shutter speed of less than 1/50 of a second would be subject to sudden change in color temperature. None of our Shutter time values are below that value. In order to calibrate the differences between photographs, all the photographs of the same room are matched with a reference picture of that room. The matched colors postprocess included change of luminance, change of color range and neutralization of color cast.

The output photographs are monoscopic and have an image size of 8760x2880 pixels, a vertical and horizontal resolution of 96 dpi and a bitrate of 24 in a jpg format. These pictures have been post processed increasing the bitrate to 32, increasing their size to 15360x8640 pixels, and saved as .hdr format.

3.3 Visual Reproduction in Unreal Engine

The .hdr pictures are used as HDRI backdrop images in Unreal Engine [7]. These use a High Dynamic Range (HDR) image as a backdrop spherically projected to a background mesh. The projection mesh is a sky dome with the actor position at the center of the dome sphere and the floor adjusted to the actor height, according to the tracking data. Each photograph is loaded in a different level, with the corresponding backdrop. All backdrops have the same center point. The reproduction of the photographs is achieved by hiding or unhiding the correspondent levels. Finally, a persistent level includes the first actor at the center position.

Several GUIs have been implemented to facilitate the simultaneous comparison between the stimuli. The GUIs provide buttons to select which stimulus perceive. The order of apparition of the stimuli is randomized for each participant. Finally, the GUIs provide sliders to introduce ratings to specific stimuli.

4 EXPERIMENTAL SETUP

The virtual laboratory consists of a silent room with low illumination. The participant is sitting in a concert chair. The hardware used is as follows: VR glasses vive pro, one hand controller, a pair of Headphones Sennheiser HD650, and a Windows pc i7 (Figure 2, left). Virtual Acoustics and Unreal Engine renderers are synchronized sharing the location and orientation of the rendering point. This is achieved by the VA Unreal plugin developed at RWTH Aachen University [8].

The experiment procedure is as follows: informing the participant to the content of the experiment, filling out a demographics questionnaire, setting up the VR glasses and headphones, getting used to the virtual rooms, elicitation and revision of the individual attributes, rating the rooms with the elicited attributes using sliders (See Figure 2, right).

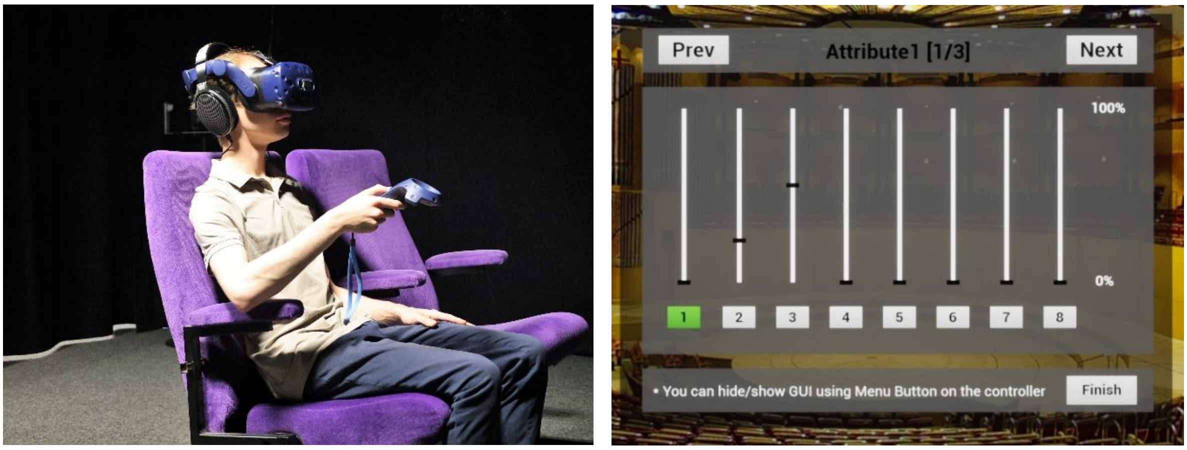

Figure 2: Experimental setup. Left side: Participant exploring a conccert hall with head-mounted display and headphones. Right side: The controller is used to interact with the graphical user interface.

5 OUTLOOK

The resulting set of vocabulary as well as the ratings based on the individual attributes do not only deliver a personal preference rating for one or another hall. But it promises to provide a bigger picture of which dimensions are important when people are exploring a concert hall not only auditorily but also visually. These results can be compared with previous studies [1, 4]. It is planned to collect case-specific vocabulary for several scenarios. Additional correlation between the subjects' attributes and objective acoustic and visual parameters of the rooms are planned. Furthermore, the application of this technology might be useful to evaluate other type of spaces, such as classrooms, music venues or even urban scenes.

6 REFERENCES

1. Lokki, T., Pätynen, J., Kuusinen, A., Vertanen, H. and Tervo, S. (2011). Concert hall acoustics assessment with individually elicited attributes. The Journal of the Acoustical Society of America 130, 835-849. https://doi.org/10.1121/1.3607422.
2. Virtual Acoustics (VA) – A real-time auralization framework (2023). http://www.virtualacoustics.org
3. D. Schröder and M. Vorländer Raven: A real-time framework for the auralization of interactive virtual environments. *Forum Acusticum*, pages 1541-1546, Aalbord Denmark, 2011 https://www2.ak.tu-berlin.de/~akgroup/ak_pub/seacen/2011/Schroeder_2011b_P2_RAVEN_A_Real_Time_Framework.pdf
4. Heck. J., Llorca-Bofí J., Aakriti, J., Vorländer, M. (2023). Text Mining of Audio-visual Descriptors for Concert Halls, 49th Annual German Conference on Acoustics, DAGA, Hamburg
5. Doeffinger, Derek (2009). Creative Shutter Speed: Master Your Camera's Most Powerful Control. Wiley. ISBN 978-0-470-45362-9
6. Filippo Naccari, Arcangelo Bruna, Alessandro Capra, Alfio Castorina, and Silvia Cariolo "Automatic image classification by color analysis", Proc. SPIE 5678, Digital Photography, (23 February 2005); https://doi.org/10.1117/12.587723
7. Unreal Engine. https://docs.unrealengine.com/4.26/en-US/BuildingWorlds/LightingAndShadows/HDRIBackdrop/
8. Unreal VA plugin: https://git-ce.rwth-aachen.de/vr-vis/VR-Group/unreal-development/plugins/unreal-va-plugin