Audiovisual models for virtual reality: Underground station

Ľuboš Hládek, Bernhard U. Seeber

Audio Information Processing, Technical University of Munich, 80290 Munich, lubos.hladek@tum.de

Introduction

Virtual acoustic environments provide new opportunities to address current issues of hearing and acoustic research. In the area of hearing research, the often reported issue is a mismatch between good performance in standard audiological assessment of auditory function of people with hearing difficulties, and the difficulties the people experience in their daily life during communication (for more details see [1] and other publications in the special issue). One of the identified shortcomings is that testing is done with static sounds in environments are very simplistic, unrealistically sounding or missing reverberation, without visual representations or in very simplistic visual scenes. Another problem is that test procedures do not capture all functions of communication and mainly lack the interactivity of the people with the environment or with fellow people. Performing auditory experiments in realistic audio and visual environments in interactive virtual or augmented reality can address some of the difficulties regarding ecological plausibility [2]. However, the research community lacks tools and content. Therefore, we created a set of digital assets that can be used as content for a virtual reality representation of an underground station, a common communication environment, but we focused on the aspects of acoustics and reproducibility. The package is freely available for research [3] and it has been introduced in previous publications [4], [5]. The current publication provides a comprehensive overview of the package and additional information on acoustic measurements.

From the complementary perspective of acoustic research, interactive virtual reality needs new approaches for acoustic simulations, which are coupled with sound field reproduction techniques, for instance, real-time implementations of image source method or ray tracing methods and real-time implementations of diffuse reverberation including the effects of diffraction and scattering [6], [7]. However, such improvements require reference acoustic recordings for acoustic benchmarking. In the current work, the models of virtual reality are accompanied with such reference measurements that were performed in the context of the pre-defined acoustic scenes (for discussion please see [5]).

Underground station environment

Underground stations are common places of communication, such as the U-Bahn station on Theresienstraße in Munich. The station is a large space with high reverberation time and with static and dynamic noise sources: from trains, escalators, ventilation systems, PA systems, or people at the platform; all of which make the communication challenging.

Scenes

In this environment, two scenes were defined and they are shown in Fig. 1. In this context, the ‘scene’ refers to a defined set of sound source positions and receivers. In Scene 1, the listener is standing on the platform and the talker or multiple talkers speak to the listener from different directions – a common spatial configuration in audiological research. In Scene 2, a sound source can be approaching or receding from the listener – a common situation in which people experience different levels of reverberation.

Acoustic properties

Reverberation time T_{30} of the station was measured with an

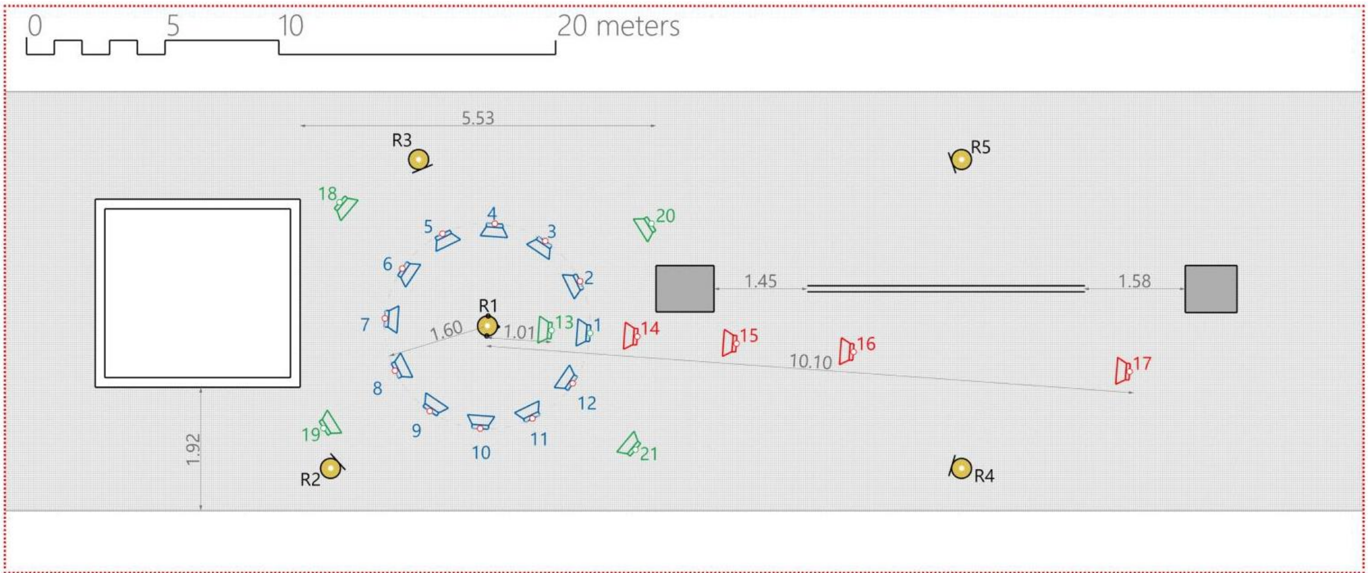

Figure 1: Two acoustic scenes were defined around listener position R1 at the station platform. Scene 1 (blue and green symbols), Scene 2 (red and green symbols). Additional receiver positions were defined at R2-R5. Picture replotted from the model documentation [3].

omni-directional sound source and omni-directional microphones 1.73 s, 2.44 s, 2.05 s, 1.7 s, 1.47 s, 1.11 s, 0.65 s (in octave frequencies from 125 Hz to 8 kHz) following recommendations of ISO 3382-2.

Acoustic model

A precise geometric acoustic model was created and it includes all acoustically relevant planes at the platform and in the area around escalators on the first underground floor of the station. The precision was achieved by calibrating the model to the point-cloud data obtained from a laser scan of the whole environment. The model has three variants: full model (183 planes) and two sub-models with 56 and 28 planes to represent the area around the defined scenes for more effective rendering for real-time reproduction. The model was acoustically verified against the measured baseline impulse responses and perceptually validated in a speech perception experiment [4]. For this reproduction in the real-time Simulated Open Field Environment (rtSOFE) [8], early reflections were modelled using image source method and late reverberation was created from multi-channel impulse response recording of the reverberant tail at a discrete position.

Visual Model

The representation was created using the images and point-cloud measurements and stored in an open-source format. The visual and acoustic models were implemented in virtual reality for reproduction over loudspeakers and projection screens in SOFE and for reproduction over headphones and HMD [9].

Environmental sounds recordings

A set of high-quality close-microphone recordings and a microphone-array recording were performed. The recordings include typical sounds of the running and idling escalators, background noise of the ventilation system, sounds of the escalator, and a multi-channel recording of the train passing, stopping, and leaving the station.

Impulse response recordings

Acoustic impulse response recordings were performed at the scene positions with three types of receivers: omni-directional microphone, artificial head (HMS-II.3-33, Head Acoustics, Germany) with behind-the-ear (BTE) capsules and multi-microphone array (EM 32, Eigenmike, USA) and two types of sound sources: a studio monitor loudspeaker (BM6A MK-II, Dynaudio, Denmark) and an omni-directional sound source. The excitation signal was a frequency-equalized one-minute-long sine-sweep such that the excitation level was set beforehand in the anechoic chamber to limit distortion of the loudspeaker and maximize signal-to-noise ratio. The recordings with the loudspeaker and the artificial head cover all defined positions in Scenes 1 and 2 (source at positions 1-17 microphone at R1), the recordings with the omni-directional microphone and the multi-channel microphone array cover a subset of positions (microphone was at R1). For the complete list, see the documentation [3].

Horizontal directivity of HMS II.3-33 with BTE capsules

The directivity of the artificial head (HMS II.3-33, Head Acoustics, Germany) with anthropometric pinnae (P.57 standard, type 3.4) with additionally mounted BTE capsules was characterized by measuring HRTFs at 0.5° steps at elevation 0° and distance 2.1 m. The excitation signal was an MLS sequence presented from the monitor loudspeaker at 80 dB SPL and equalized with a FIR filter. The head was carefully positioned on the center of rotation of the motorized turntable (ET250-3D, Outline, Italy) which controlled the rotations. The reference recording was obtained using a measurement omni-directional microphone (MM210, Microtech Gefell, Germany) placed at the center of the head. The measurements were performed in the anechoic chamber at the institute with an low cut-off frequency of 100 Hz. The 11-ms long HRTFs were obtained by computing FFT of the windowed (3 ms cosine ramps) head-related impulse responses.

Fig. 2 shows data of left-ear HRTFs measured in the ears and at the two microphones of the BTE shells at five exemplary angles.

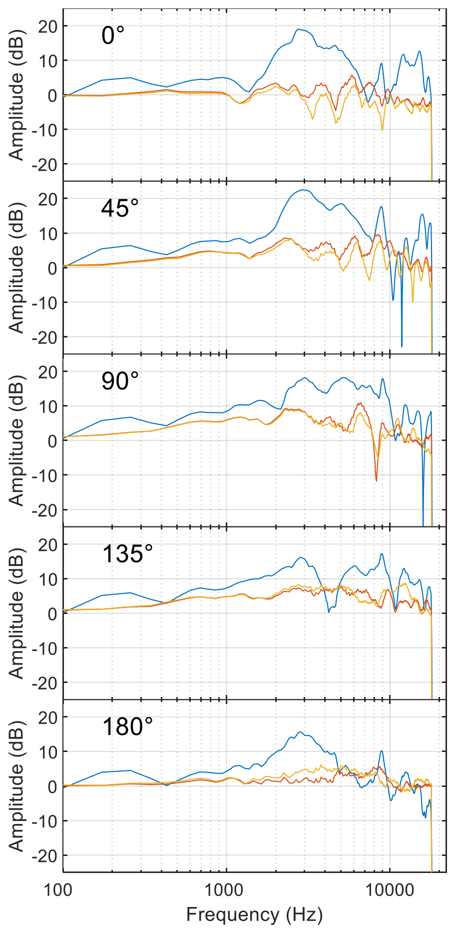

Figure 2: Left-ear HRTFs of HMS II.3-33, type 3.4 pinnae (blue) equipped with behind-the-ear capsule with 2 microphones (front – red, rear - yellow) and at 5 example angles measured at 0° elevation and 2.1 m distance.

Source directivity – Dynaudio BM6A MK-II

The directivity of the sound source (BM6A MK-II, Dynaudio, Denmark) used during the impulse response recordings was characterized by measuring the impulse response in horizontal (5° steps) and vertical directions (10° steps). The blue dots in the bottom panel of Fig. 3 shows all measurement points, red and black circles show points that were analyzed in the top and middle panels of Fig. 3. The measurements were performed in the anechoic chamber by rotating the loudspeaker around its axis of rotation and such that the measurement microphone (MM210, Microtech Gefell, Germany) was placed at 1.5 m distance. The motorized turntable controlled the horizontal rotation, the vertical rotation was fixed by a clamp that was manually adjusted. The vertical position was controlled using a digital tilt sensor (GLM 80, Bosch, Germany). The directivity was normalized to the on-axis direction and computed in 1/3 octave bands, and these data are provided in SOFA format.

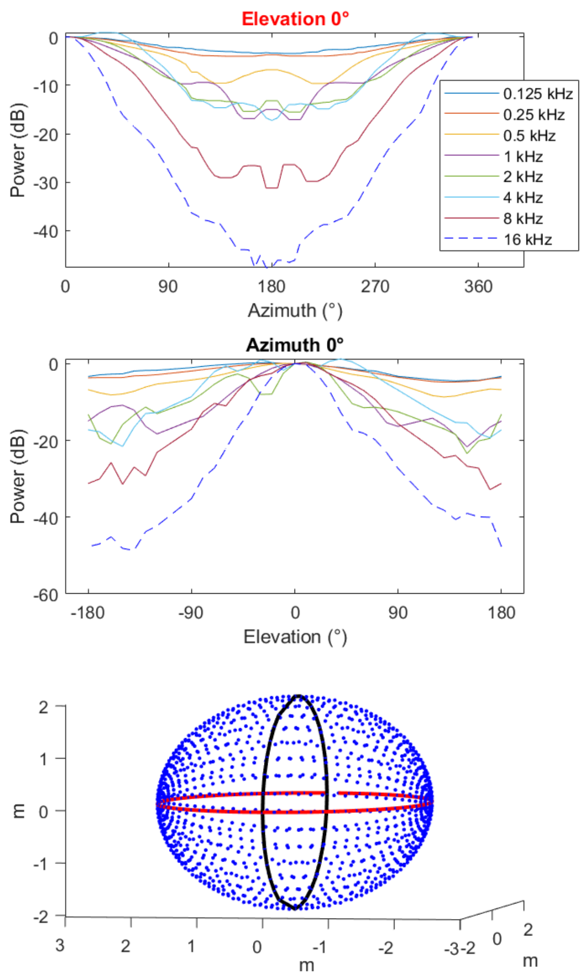

Figure 3: Normalized directivity of the sound source (Dynaudio BM6A MK-II) used during the measurements computed in 1/3 octave bands (octave bands shown). The bottom panel shows all measurement points (1368) and the subsets depicted in the panels above.

Implementation in the real-time Simulated Open Field Environment

The audio-visual model was implemented in the rtSOFE [8] [10]. Figure 4 shows a picture from the demonstration (available online as a video [11]) in which the complete simulation of the environment was run in real-time including a virtual loudspeaker, sounds of the escalators, voice of the speaker, and people on the platform. Altogether five independent sound sources were simulated and reproduced over 61 loudspeakers. The visual projection and the control of the actual spatial configuration of the sounds was managed using Unreal Engine and custom-built plugins [9].

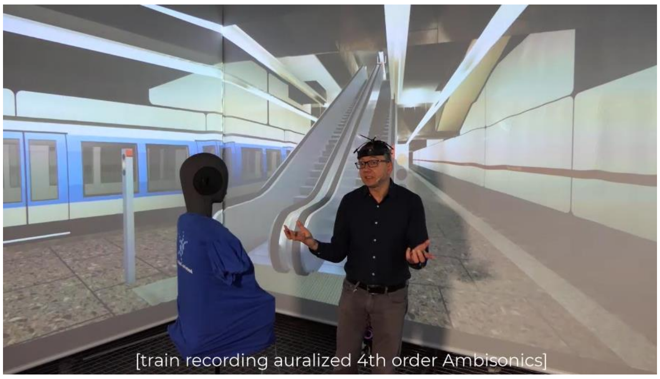

Figure 4: Picture from the demonstration of the underground station environment rendering in the real-time Simulated Open Field Environment inside our anechoic chamber. The picture shows author BS. The video recording can be downloaded from [11].

Conclusion

Here we present the audiovisual model 'Underground station environment' and its implementation in virtual reality using rtSOFE. During the demonstration, people can experience the environment of the underground station in the laboratory with several moving sound sources, realistic real-time simulation of reverberation, and with comprehensive visual simulation on four big projection screens or with a head mounted display. Therefore, interactive studies of communication in a life-like underground station situation or experiments on new approaches for real-time acoustic simulations are possible in our or other laboratories.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330, Project C5. rtSOFE development is supported by the Bernstein Center for Computational Neuroscience, BMBF 01 GQ 1004B.

References

[1] G. Keidser et al., "The Quest for Ecological Validity in Hearing Science: What It Is, Why It Matters, and How to Advance It," Ear Hear., vol. 41, no. Supplement 1, pp. 5S-19S, Nov. 2020, doi: 10.1097/AUD.0000000000000944.

[2] V. Hohmann, R. Paluch, M. Krueger, M. Meis, and G. Grimm, “The Virtual Reality Lab: Realization and Application of Virtual Sound Environments,” Ear Hear., vol. 41, no. Supplement 1, pp. 31S-38S, Nov. 2020, doi: 10.1097/AUD.0000000000000945.

[3] L. Hladek and B. U. Seeber, “Underground station environment.” 2022. doi: 10.5281/zenodo.5532643

[4] L. Hladek, S. D. Ewert, and B. U. Seeber, “Communication Conditions in Virtual Acoustic Scenes in an Underground Station,” in 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), 2021, pp. 1–8, doi: 10.1109/I3DA48870.2021.9610843

[5] S. van de Par et al., “Auditory-visual scenes for hearing research,” Nov. 2021 URL: `http://arxiv.org/abs/2111.01237`

[6] T. Wang, M. Lambacher, and B. U. Seeber, “Extension of the real-time Simulated Open Field Environment for fast binaural and non-isotropic reverberation rendering,” in Fortschritte der Akustik - DAGA ’22, 2022, p. same issue.

[7] S. Fichna, B. U. Seeber, C. Laudeau Bobadilla, and S. D. Ewert, “Evaluation of complex acoustic scenes for hearing research and audiology,” in Fortschritte der Akustik - DAGA ’22, 2022, p. same issue.

[8] B. U. Seeber and S. W. Clapp, “Interactive simulation and free-field auralization of acoustic space with the rtSOFE,” J. Acoust. Soc. Am., vol. 141, no. 5, pp. 3974–3974, May 2017, doi: 10.1121/1.4989063.

[9] F. Enghofer, Ľ. Hládek, and B. U. Seeber, “An ‘Unreal’ Framework for Creating and Controlling Audio-Visual Scenes for the rtSOFE,” in Fortschritte der Akustik - DAGA ’21, 2021, 2021, pp. 1217–1220.

[10] B. U. Seeber, S. Kerber, and E. R. Hafter, “A system to simulate and reproduce audio–visual environments for spatial hearing research,” Hearing Research, vol. 260, no. 1–2, pp. 1–10, Feb. 2010, doi: 10.1016/j.heares.2009.11.004.

[11] “Audiovisual scenes demonstrated in the real-time Simulated Open Field Environment (rtSOFE),” 2021. URL: `https://www.youtube.com/watch?v=yN7vD1khTOI`. [Accessed: 09-Mar-2022]