Student Conference Proceedings
Vol 1, No 1, 2025, Article ID 1997, 4 Pages

Student Conference Lübeck

Proceedings Article

Validation of an experimental setup for creating augmented acoustic environments

Luca Götzke [a,b,*] · Piotr Majdak [b] · Florian Denk [c] · Ľuboš Hládek [b]

[a] Student of Auditory Technology, Universität zu Lübeck, Lübeck, Germany
[b] Acoustics Research Institute of the Austrian Academy of Sciences, Vienna, Austria
[c] German Institute of Hearing Aids, Lübeck, Germany
[*] Corresponding author, email: luca.goetzke@student.uni-luebeck.de; florian.denk@uni-luebeck.de

Received 11 February 2025; Accepted 03 June 2025; Published online 23 July 2025

Abstract

This study aimed to validate calibration routines for an experimental setup that simulates and presents realistic acoustic environments via open headphones, with real-time motion tracking. A virtual acoustic environment was implemented, using the real-time Simulated Open Field Environment (rtSOFE) software package, which employs the image source method for real-time applications. This setup, including a set of calibration filters and the procedures to create them, was previously used to study face-to-face communication in nosy environments. The results in the present work support the use of this system and it's filter creation procedures in future auditory research, providing a reliable platform for real-time experimentation in virtual and augmented acoustic environments.

I. Introduction

This study aimed to validate the SPL calibration of an simulated acoustic environment with open headphones. The real-time Simulated Open Field Environment software package (rtSOFE) [1], capable of real-time binaural room impulse response (BRIR) calculations, was used to augment a two-person conversation including dynamic position changes with arbitrary reverberations and background noise. Head-related transfer functions (HRTFs) can be used for the receiver as well as a mouth-directivity model to simulate the source. This makes rtSOFE comparable to other real-time binaural spatialisation software such as the 3D Tune-In Toolkit [2], the Binaural Rendering Toolbox (BRT) [3], the Toolbox for Acoustic Scene Creation And Rendering (TASCAR) [4] and the room acoustics simulator RAZR [5].

This experimental setup, which was previously used to study face-to-face communication in noisy environments [6], relies on precise equalization filters to ensure accurate transmission of audio signals without the influence of hardware artifacts. This work details the methods utilized to validate the production of the correct SPL by the system, including headphone and headset microphone equalization, integration of motion tracking, and discusses the results of these validation steps.

II. Methods and materials

This experimental setup used rtSOFE, a software package designed to simulate acoustic environments using the image source method. On the one hand, the software package consists of the rtSOFE instance, which calculates the BRIRs for a given source and receiver position while using the provided room parameters to incorporate reverberations. It can also use a directivity profile for the source and and a HRTF for the receiver. Directivity is implemented by applying a multiplier, similar to an absorption coefficient, at the beginning of the image source method. On the other hand it has the convolver, which takes an input signal and calculates the convolution with

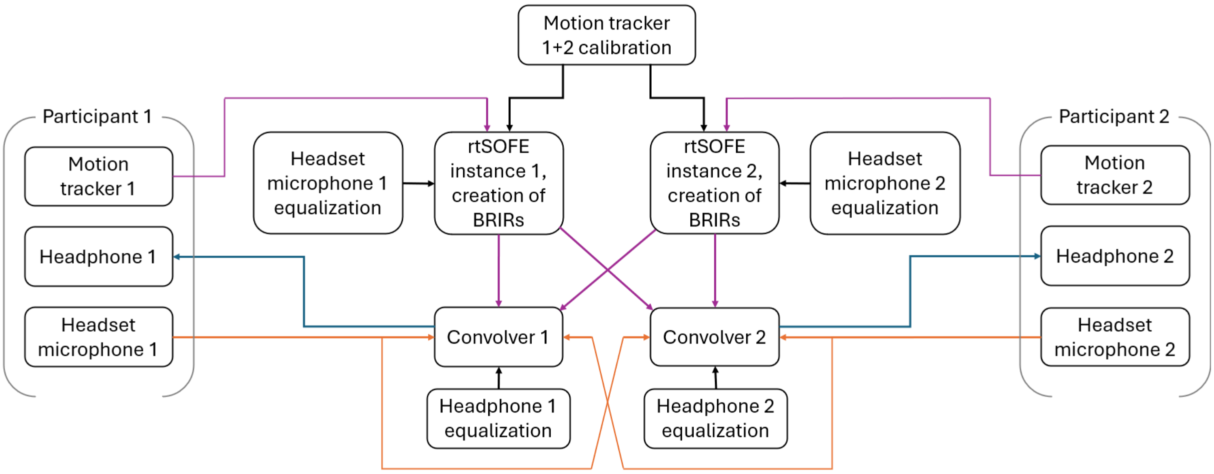

Figure 1: Interaction flowchart of the interaction of the different parts of the experimental setup. Participant 1 is represented by Motion tracker 1, Headphone 1 and Headset microphone 1 on the left and Participant 2 respectively on the right. The purple path shows the processing chain for the positional updates and the resulting BRIRs. The orange path shows the transmission from each of the headset microphones to both of the convolvers. The blue path shows the transmission from the convolvers to the headphones.

the BRIRs and headphone equalization filters as an output. The flowchart in Figure 1 shows the connection between the different parts of this setup. Participant 1 on the left is talking (source), and participant 2 on the right is listening (receiver). The position of the source (Motion tracker 1) and the position of the receiver (Motion tracker 2) are both sent to the two rtSOFE instances. Here the BRIRs from the source to the receiver are computed and sent to the convolvers. Convolver 1 sent its output to Headphone 1, enabling Participant 1 to hear the reverberation of it's own voice. Convolver 2 sent its output to Headphone 2, letting Participant 2 hear Participant 1 with reverberations. The positions of the participants get updated approx. 43 times per second, enabling real-time directional auditory cues for the participants. Both participants are functioning as source and receiver. Further rtSOFE instances can be used to simulated additional virtual conversations from recorded speech or for added background noise. Due to the use of open headphones (AKG K 1000, see Figure 2), direct sound transmission between participants is preserved.

To produce an accurate auditory impression for the participant, the influence of the open headphone and headset microphone (AKG MicroMic C 520 Vocal) capturing participants speech had to be minimized. This was done by creating equalization filters for each of the devices, to ensure a correct SPL presentation at the participants' ears. The filters for the headphones were applied in the convolvers. The headset microphone equalization filters were implemented by adding frequency-specific attenuation for all directions on top of the mouth directivity [7] utilized in the rtSOFE instances.

Lastly the motion-tracker (HTC Vive Tracker 2QAB100) calibration was done by defining a virtual room in Steam

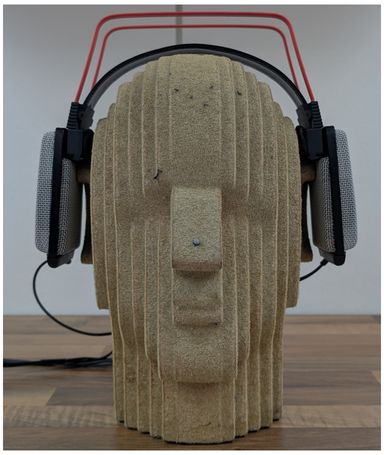

Figure 2: Open headphones (AKG K 1000) on a head stand. The open nature enables the simultaneous reception of the natural acoustic environment as well the headphone output.

VR and introducing an offset for the motion-tracker positions to change the reported position to the middle of the head instead of beeing on top of the headphones. Another part was transforming the coordinate system used by Steam VR to the coordinate system used by rtSOFE. The test setup (see Figure 3) for the validation of the equalization filters and the motion-tracker calibration simulated a receiver on the left and a source on the right. The test signals were played back by a loudspeaker (custom made, with VIFA 10 BGS as drivers [8]) and were picked up by the headset microphone or the reference microphone, which was located near the motion tracker on the left. The test setup was located in a sound booth (IAC Acoustics, based on 120-A-Series). Octave-band-filtered pink noise with center frequencies of 125 Hz to 8 kHz as well as a broadband pink noise and a pure tone 1 kHz tone were used as the test signals. All test signals were normalized to the same RMS level.

II.I. Headphone equalization

The first measurement validated the equalization filters for the headphones, which are applied by the convolver. The process began by calibrating an artificial ear (Brüel&Kjær Type 4153) connected to a sound analyzer (Brüel&Kjær Type 2260), using an acoustical calibrator (Brüel&Kjær Type 4231). The impulse response from one of the headphones was measured with the Impulse Response Measurer app in MATLAB [9] using a maximum length sequence, which was then used to generate equalization filters. The calibration was validated by playing a 1 kHz pure tone at −17 dB FS through the headphones, ensuring a playback level of 80 dB SPL at the artificial ear. Subsequently, the broadband noise and octave-band-filtered pink noise were played back. The difference between the measured SPL of both noises and the 1 kHz pure tone reference was calculated to confirm the accuracy of the equalization.

II.II. Headset microphone equalization

The second measurement assessed the level of the transmitted sound by the headset microphones. A reference microphone (PreSonus PRM1) was placed 1 m in front of the loudspeaker and was calibrated using an acoustic calibrator (Cirrus CR:511E). One of the headset microphones was placed in front of the loudspeaker. To create a reference SPL without the equalization filter for the headset microphone, the test signals were played back via the loudspeaker. They were picked up by the reference microphone and sent to the headphone on the artificial ear outside the sound booth. Here the output SPL of each test signal was measured. To remove the influence of the loudspeaker, the digital level of each test signal was adjusted to ensure the same SPL for all test signals on the artificial ear.
The transfer function of the headset microphone was derived using spectral division by comparing its frequency response with that of the reference microphone. Equalization filters for the headset microphone were generated and applied to the mouth directivity. An rtSOFE instance simulated a room with no reflections and no HRTF to remove any attenuation on the receiver side. The resulting BRIR was then convolved with the adjusted test signals to include the headset microphone equalization. As a last step the adjusted and headset microphone equalized test signals were played back from the loudspeaker via the headset microphone to the headphone on the artificial ear outside the sound booth and the output SPL was measured. The deviation between the SPL of the test signals before and after the adjustment and equalization process was then calculated.

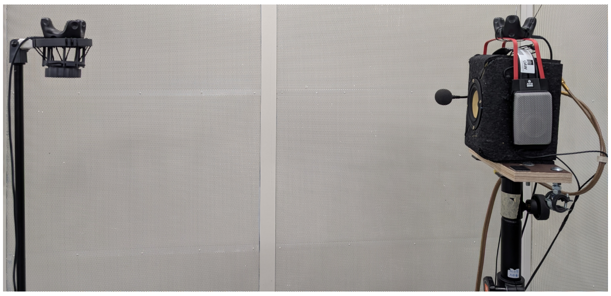

Figure 3: Test setup for motion-tracker validation inside a sound-treated, double-walled sound booth. A motion tracker on the left, representing the listener. A loudspeaker with headset microphone and motion tracker mounted on top of the headphone on the right, representing the speaker.

II.III. Motion-tracker validation

The test setup as seen in Figure 3 was used to measure the motion-tracking integration. The signal processing chain followed the steps shown in Figure 1. The loudspeaker played back test signals, which were captured by the headset microphone and processed through an rtSOFE instance and a convolver. The resulting output was played back via a headphone on the artificial ear outside the sound booth. The rtSOFE instance simulated a room with no reflection and the receiver used no HRTF to remove any frequency weighting for the incoming sound. Only the base mouth directivity without microphone equalization was used for the source. To evaluate the correct integration of the motion tracker, the test signal SPL measurement via the artificial ear was first done for zero degrees azimuth, i.e. source and receiver were facing each other directly. The motion tracker on the left was stationary and the combination of motion tracker, loudspeaker and microphone on the right was rotated for each measurement. The output SPL at the artificial ear for the test signals were measured for three different rotations of 45°, 90° and 135°. The difference to the zero degree measurement were calculated for each rotation and test signal. Lastly the results were compared to the attenuation values of the underlying mouth directivity and the deviation was calculated.

III. Results and discussion

The experimental validation demonstrated the functionality of the creation and application of equalization filters and the motion-tracker calibration. The headphone equalization filter led to a maximum deviation of 1.2 dB from the desired SPL for the broadband pink noise (Table 1, row 1). The headset microphone's equalization

10.18416/SCP.2025.1997

Table 1: Results of the validation measurements. The first row shows the relative SPL change from the reference pure tone with the headphone equalization. The second row shows the relative SPL change from the desired SPL before and after the adjustment and headset microphone equalization were applied. The content of the third to fifth rows shows the relative SPL changes between the attenuation at each motion-tracker rotation and the intended attenuation from the directivity profile used.

Signal / SPL [dB]	Pure tone 1 kHz	Pink noise broadband	Center frequency for octave-band-filtered pink noise [Hz] 125	250	500	1k	2k	4k	8k
Headphone	ref. value	-1.2	0.8	1.1	-0.6	0.3	-0.3	-0.2	-0.7
Headset microphone	0	-1	1.1	1	-0.9	-2.1	-2	-1.7	-1.9
Motion tracker at 45°	-1.1	-0.6	1.3	1.6	1.2	-1.9	-2.4	-1.7	-1.3
Motion tracker at 90°	0.8	1.9	0.8	0.3	0.6	0.9	1.5	2.9	2.1
Motion tracker at 135°	1.4	2.1	-0.8	-1	0.6	1.5	1.5	2.3	3.1

filter resulted in a maximum deviation of 2.1 dB from the desired SPL at 1 kHz (Table 1, row 2). The integration of motion-tracking and rtSOFE demonstrated accurate real-time binaural room impulse response calculations, as shown by a maximum deviations of 3.1 dB relative to the mouth directivity used in rtSOFE (Table 1, rows 3-5). While the results of the headphone equalization show only small deviations of approx. 1 dB to the desired SPL, the equalization of the microphone show slightly larger deviations of approx. 2 dB. The motio-tracking measurement show a larger deviation of approx. 3 dB. A potential explanation for the significant deviations observed in the equalization of the headset microphone and the motion tracking may be attributed to the discrepancy between the spectrum of the test signals utilized and the reference values derived from the mouth directivity, which were obtained through single-frequency recordings. Additionally the presence of roll and pitch angle deviations from zero degrees could result in deviations from the directivity pattern. A last factor would be fluctuations in the level meter readings. Motion-tracking update latency has not been tested.

IV. Conclusion

This study validates the equalization filters and motion-tracker calibration used in our setup, designed to simulate realistic acoustic environments with dynamic position changes in real-time. The validation measurements showed an accurate SPL reproduction with the applied equalization filters and motion-tracker calibration. The results support the use of this system in future auditory research, providing a reliable platform for creating virtual acoustic environments.

Acknowledgments

The work has been carried out at the Acoustics Research Institute of the Austrian Academy of Sciences, Vienna, Austria and co-supervised by the German Institute of Hearing Aids, Lübeck, Germany. Ľuboš Hládek is a recipient of a Seal of Excellence Postdoctoral Fellowship of the Austrian Academy of Sciences. Research funding: The author state no funding involved.

Author's statement

Conflict of interest: Authors state no conflict of interest.

References

[1] B. U. Seeber and T. Wang, Real-time Simulated Open Field Environment (rtSOFE) software package, version 1.1, Zenodo, 2021. doi:10.5281/zenodo.5648305.

[2] M. Cuevas-Rodríguez et al. 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation. *PLOS ONE*, 14(3), 2019, doi:10.1371/journal.pone.0211899.

[3] D. González-Toledo et al., The binaural rendering toolbox. A virtual laboratory for reproducible research in psychoacoustics, in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, 331–337, Turin, Italy: European Acoustics Association, 2024. doi:10.61782/fa.2023.1042.

[4] G. Grimm et al., Toolbox for acoustic scene creation and rendering (TASCAR): Render methods and research applications, in *Proceedings of the Linux Audio Conference*, Mainz, 2015.

[5] T. Wendt et al. A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. *Journal of the Audio Engineering Society*, 62(11):748–766, 2014.

[6] Ľ. Hládek et al., Communication Conditions in Virtual Acoustic Scenes in an Underground Station, in *2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, 1–8, Bologna, Italy: Institute of Electrical and Electronics Engineers, 2021. doi:10.1109/I3DA48870.2021.9610843.

[7] J. L. Flanagan. Analog measurements of sound radiation from the mouth. *The Journal of the Acoustical Society of America*, 32(12):1613–1620, 1960, doi:10.1121/1.1907972.

[8] P. Majdak et al. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics*, 72(2):454–469, 2010, doi:10.3758/APP.72.2.454.

[9] T. M. Inc., Audio toolbox version: 23.2 (R2023b), Natick, Massachusetts, United States, 2022. URL: https://www.mathworks.com.