DOI: 10.61782/fa.2023.0429

forum acusticum 2023

PRELIMINARY EVALUATION OF THE AURALIZATION OF A REAL INDOOR ENVIRONMENT FOR AUGMENTED REALITY RESEARCH

Alessandro Giuseppe Privitera[1]* **Marco Noro**[2] **Michele Geronazzo**[2]

[1] Dept. Mathematics, Computer Science and Physics, University of Udine, Italy
[2] Department of Management and Engineering, University of Padua, Italy

ABSTRACT

This paper describes the calibration procedure and the technical setup for a realistic real-time acoustic reconstruction of a real indoor environment, a corridor, in the context of Audio Augmented Reality (AAR). The acoustic phenomena inside such space are simulated using the scattering delay network (SDN) algorithm. Wall reflection coefficients have been estimated using room dimensions, wall materials, and RT60 decay measurements. Auralisation has been dynamically conveyed by using a personalized head-related transfer function (HRTF), modeled by combining (i) a spherical head model with ear displacement with (ii) the high-frequency magnitude of an HRTF selected from the CIPIC database by using two 2D images of the user's head. Moreover, the iPad's AR camera tracking system and AirPods pro accelerometers have tracked the listener's head and body position in real space. The proposed preliminary evaluation focuses on the impact of the different rendering factors in a simple AAR environment, suggesting that personalization, room calibration, and volume gain help render a more plausible AAR scene.

Keywords: *augmented reality, personalized HRTF, sonic interactions, room acoustics, auralization*

**Corresponding author*: privitera.alessandrogiuseppe@spes.uniud.it.

1. INTRODUCTION

Audio augmented reality (AAR) integrates real-world acoustic space with virtual sounds in a hybrid environment. Recent advances in sensors and immersive audio technologies have added further realism and richness to user experiences in various applications, from entertainment and gaming to education and training [1]. One of the most important aspects in realistic AAR is *auralisation*, the audio rendering process involving models that combines listener acoustics, i.e., head-related transfer function (HRTF), and environmental acoustics, namely room impulse response (RIR). Effective auralisation of sound sources is fundamental to providing a plausible listening experience, but the flexibility of the rendering process is crucial for a real-time environment [2]. On the other hand, the vast majority of commercial headphones models alter natural hearing [3] resulting in a mediated AAR scenario in which real audio sources are not authentic [4] can present differences with respect to the real environment, e.g. in volume and equalization. In this work, we develop the drag-and-drop MUSHRA [5] in a mobile application to evaluate the effect on an AAR scenario of acoustic personalization and calibration in terms of perceived *plausibility*. The objective of this pilot test is twofold: to preliminarily assess the impact of different rendering factors in an AAR setup and to verify the efficacy and usability of the procedure. As a theoretical background for the interpretation of the obtained preliminary results, this study draws upon Sonic Interaction in Virtual Environments (SIVE) [6], an emerging field that explores the interplay between humans and computers through immersive auditory feedback, focusing on sound as a primary carrier of information, meaning, and emotional qualities.

In SIVE, three main concepts are introduced: *Immersion*, the degree to which the virtual simulation engages the range of sensory channels, *Coherence*, the effectiveness of the sonic interaction design in terms of rendering and meaningful behaviors for each user [7], and *Entanglement*, the dynamic and mutual adaptations and active participation among the user and all the other key actors, including the user, technology, and content [8].

Sec. 2 summarizes the main models and technologies to implement a personalized AAR. In Sec. 3 and 4, different degrees of personalization and calibration are evaluated in terms of plausibility of the augmented experience.

2. AN AURALISATION FRAMEWORK FOR AR

The auralization framework, developed in Unity3D [1], models a campus corridor (width 19.70 m, depth 2.32 m, height 3.38 m) at the Department of Management and Engineering of the University of Padua, reproduced in a VE. The virtual acoustic space was obtained using Scattering Delay Network (SDN) reverb for RIRs [9] and personalized HRTF for dynamic auralisation of any space. Apple iPad and AirPods Pro sensor data and libraries provided by the vendor were used as for user tracking. Fig. 1 depicts the overall AR framework developed for this work.

2.1 Room Acoustics

The SDN algorithm integrates delay networks and acoustic artificial reverberators, yielding a physically and perceptually accurate acoustic simulation [9, 10]. The SDN algorithm employs the image-source method [11] to render the direct sound path (source-to-listener) and 1^{st}-order reflections in both time and space. The following equation can express the computed RIR:

$$R(z) = \frac{1}{K}\mathbf{k}_M^T(z)\left[\overline{\mathbf{A}}^T\overline{\mathbf{H}}^{-}1(z) - \mathbf{PD}_f(z)\right]^1 \mathbf{k}_s(z) + \overline{g}z^{-D_{S,M}} \quad (1)$$

where $\mathbf{k}_M^T(z) = \mathbf{\Gamma}_M^T\mathbf{D}_M(z)\mathbf{G}_M\mathbf{W}$ is the combination of the block diagonal weight matrix for pressure extraction, the source directivity vector, the delay matrix, and the attenuation matrix; $\mathbf{k}_s(z) = \mathbf{U}\mathbf{G}_S(z)\mathbf{D}_S(z)\mathbf{\Gamma}_S$ is the block diagonal matrix that assigns source signals to input wave variables and listener. $\overline{\mathbf{H}}(z)$ is the wall absorption matrix, $\mathbf{D}(f)$ is the inter-node delay matrix, $\overline{A}$ is the scattering matrix, and $\mathbf{P}$ is a network topology-dependent permutation matrix. Moreover, HRTF filters can be easily incorporated to produce a fully customized binaural room impulse response (BRIR).

[1] Available at `https://unity.com/`

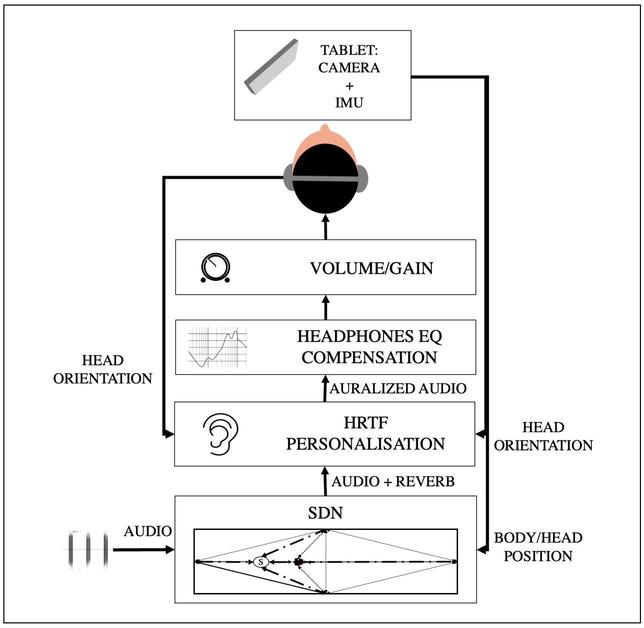

Figure 1. Schematic representation of the auralisation framework.

To correctly reproduce room acoustics, the RT30 decay time of the room was measured with a Sound Level Meter (Brüel & Kjær 7830) with the setup configuration displayed in Fig. 2. The RT30 of the virtual room was matched by visual inspection, varying two SDN parameters:

1. absorption material: corresponding to a 4^{th} order FIR filter, $H(z)$ designed using [12] and selected by considering wall materials of the real room,

2. reflection coefficients: a manual gain that can be used to adapt the theoretical filter to the real material, independently for each wall.

Table 1 displays the final values. The average RT30 value is 1.38 s in the real corridor 1.27 s in the VE. Fig. 3 shows the RT30 decay time of the real and virtual room for each frequency. Notwithstanding those real and virtual environments present similar average decay times, values per band present differences that might influence spatial perception. It is important to note that while many methods have been suggested in literature to compute real room

Table 1. Absorption materials and reflection coefficients in SDN simulation.

Wall	Material	Refl. Coeff.
Back,Front, Left	gypsum	0.9
Right	glass	0.9
Floor	concrete	0.98
Ceiling	concrete	0.99

Figure 2. The campus corridor and the RT30 measurement setup.

acoustics and absorption filter, the task is still considered complex to solve [13].

The SDN implementation used in this work is made in Unity using C. In game cycle (*void Update*), the algorithm detects the distance and direction from the virtual audio sources to every wall in the virtual room in order to correctly compute audio delays of reflected sounds and applying the correct absorption filter. In the built-in audio thread accessed by *void OnAudioFilterRead*, dynamic HRTF filters are applied to each reflection based on user position and orientation.

2.2 User Acoustics

Using individual HRTFs, instead of the generic dummy head HRTFs, combined with reverberation and head tracking can greatly reduce localization error and improve externalization [10]. However, measuring individual HRTFs is time-consuming and requires specialized facil-

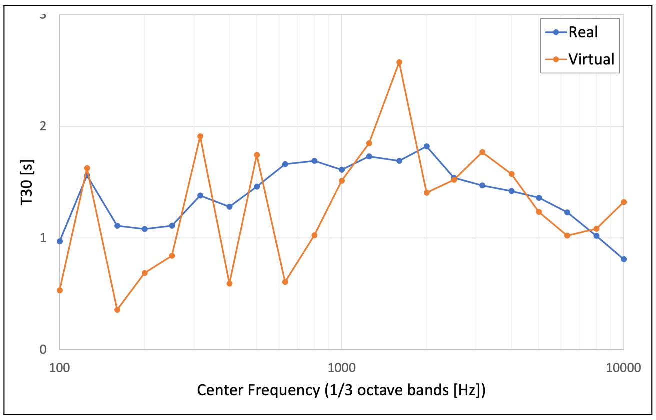

Figure 3. RT30 per 1/3 octave bands for real vs. virtual room.

ities. Our study employs mixed structural modeling proposed and evaluated by Geronazzo et al. [10] to balance optimization and efficiency in HRTF representation. This model requires a minimal set of anthropometric data, including the ear contour ($C1$) that outlines the helix, head width ($X1$), and head depth ($X3$). Using $X1$ and $X3$, low-frequency content (up to 1 kHz) can be modeled using a spherical head model with ear displacement [14]. Assuming diametrically opposed ears, the spherical Interaural Time Difference (ITD), one of the primary cues available to the auditory system for determining the spatial location of sound sources [15], in the horizontal plane is estimated as:

$$ITD(\gamma, a) = \begin{cases} \frac{a}{c}\left(\sin\gamma + \gamma\right) & \text{if } 0 \leq \mid \gamma \mid < \frac{\pi}{2} \\ \frac{a}{c}\left(\pi - \gamma + \sin\gamma\right) & \text{if } \frac{\pi}{2} \leq \mid \gamma \mid < \pi \end{cases}, \tag{2}$$

where c is the speed of sound, the head radius a is given by $a = \left(0.41\frac{X1}{2} + 0.22\frac{X3}{2} + 3.7\right)$, and γ is the angle between the source vector $\vec{S}$ and the ear vector $\vec{e}$ with origin in the center of the sphere that is displaced relative to the interaural axis to $x_0 = [0, e_b, e_d]$ with e_d/e_b the downwards/frontwards ear shift, by using:

$$\gamma = cos^{-1}\left[\frac{(\vec{S} - \vec{x_0}) \cdot (\vec{e} - \vec{x_0})}{\vec{S} - \vec{x_0}\vec{e} - \vec{x_0}}\right]. \tag{3}$$

On the other hand, high-frequency spectral content is selected from the best-fit HRTF included in CIPIC [16] dataset according to anthropometric characteristics. For each participant, n = 15 ear contours $C1$ and k = 20 ear canal positions were manually annotated by using a Matlab script [2] and three photos of the subject (front and side views and ear closeup) with a measuring tape as a reference. Based on these annotations, we estimated pinna notch frequencies f_0 for each elevation angle ϕ [17]. Such notch frequencies f_0 are compared with the corresponding frequencies F_0 extracted from the measured HRTFs from CIPIC employing a mismatch metric:

$$m_{(k,n)} = \frac{1}{N_\phi} \sum_{\phi} \frac{|f_0^{(k,n)}(\phi) - F_0(\phi)|}{F_0(\phi)}, \tag{4}$$

given N_ϕ the number of elevation angles.The best matching HRTF from the CIPIC database in terms of Eq. 4 was selected. Low and high-frequency contents are combined with a linear crossfade at 1Khz.

3. AUDIO QUALITY EVALUATION

The pilot test consists of evaluating different degrees of personalization and calibration of the acoustical space. It involved 16 participants (all male, age 20-35), self-reporting normal-hearing capabilities.

3.1 D&D MUSHRA

To collect user preference, we implemented the evaluation method suggested by [5, 18] called *drag-and-drop MUSHRA* (Multiple Stimuli with Hidden Reference and Anchor, D&D MUSHRA), a modified MUSHRA test [3] aimed to simplify the rating interface, especially using touch screens. MUSHRA is a standard methodology for audio stimulus comparison tests in terms of perceived quality, usually employed in evaluating lossy audio compression algorithms. The suggested modification introduces a unified interface (Fig. 4) for playback and rating using drag-and-drop actions, resulting in a reduced time to complete the ranking task [18]. The stimulus playback is triggered by the user pressing the corresponding button with a generic icon and no stimulus numbering. The rating is achieved by moving the button on a xy-grid, where the horizontal axis modifies the rating scale. The vertical axis can help users visually organize their ratings without affecting stimulus scoring. In the upper right corner of the GUI, a light blue button is available to play reference stimulus on demand.

[2] https://github.com/msmhrtf/personalization2020_app

[3] ITU-R recommendation BS.1534-3 available at `https://www.itu.int/rec/R-REC-BS.1534-3-201510-I`

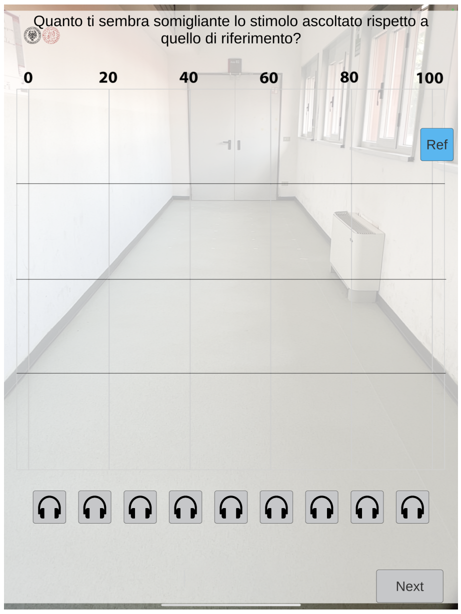

Figure 4. The D&D MUSHRA User Interface.

The D&D MUSHRA could be easily adapted for testing audio quality taxonomies such as Spatial Audio Quality Inventory (SAQI) [19]. Ratings obtained from the D&D MUSHRA interface were comparable to those obtained with the classic interface in terms of reliability and discrimination ability. To the authors' knowledge, this study implements the D&D MUSHRA interface in an experimental setup for the first time.

3.2 Setup

The setup consisted of a loudspeaker Genelec 8030C (Woofer 5", Tweeter 3/4", frequency response: 47 Hz - 25 kHz, positioned at 1.7 m from the floor) connected to an Edirol UA-4FX audio interface. In Unity, a virtual corridor was created with the same dimension as the real experimental room, spatially aligned with the real room during the experiment using the AR framework. The virtual user avatar was positioned at the same position as the participant, at 9.85 m from the front of the room, with the windows on the right. The virtual audio source was positioned behind the participant in the same position as the real loudspeaker, leaving the VE hidden on the iPad screen. As depicted in Fig. 1, an Apple iPad Pro (12.9-inch 5^{th} gen., iPadOS 15.4) and Apple AirPods Pro (1^{st} gen.) tracked the participant's body. iPad front and rear cameras tracked head and body position in the physical

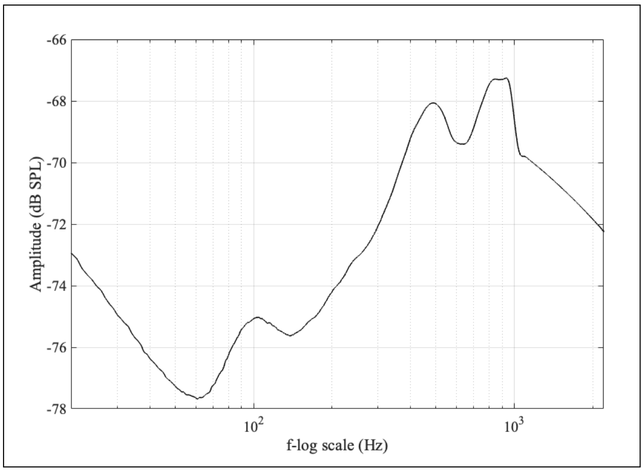

Figure 5. Headphone compensation.

space, while the AirPods Pro IMU sensor tracked head orientation only. In order to create a more ecological audio environment, though mediated by the use of earphones, AirPods *Audio Transparency Mode* was used.

3.3 Stimuli and Conditions

During the pilot experiment, participants could listen to a test sound consisting of a train of three 40 ms Gaussian noise bursts with 30 ms of silence between each burst. A reference stimulus with the same nominal position as the virtual sounds could be played back (volume matched at 70 db(A)) on demand through the loudspeaker in the room. Test audio was spatialized using nine different arrangements of the following components:

- **HRTF**: Include customized HRTF, a generic KEMAR head, or no HRTF.
- **Artificial reverberation**: A second room (20 m x 20 m x 4 m) was modeled in some experimental conditions in addition to the previously described experimental corridor. We applied one of these two reverb conditions or no reverb to the conditions.
- **Headphone equalization compensation**: audio impulses can be calibrated to reduce the coloration caused by headphones [4]. The equalization curve, inverse of the earpods transfer function, applied is shown in Fig. 5.

[4] Impulse file available at `https://github.com/jaakkopasanen/AutoEq`

- **Stimulus volume**: To achieve a comparable Sound Pressure Level (SPL) with the reference, the audio stimulus SPL was calibrated before the experiment. A custom setup was developed using a virtual source positioned within the VE, played through a fake silicone ear wearing AirPods, to simulate the placement of headphones over a human ear. Using a sound level meter, the gain in the simulation was adjusted to calibrate the volume according to the reference loudspeaker. To test the effect of volume on perceived stimuli, two different SPL were used: 70 db(A) corresponding to the real reference stimulus SPL and 75 db(A), i.e., a +5 db(A) gain. Volumes were measured at 1.21 m(4 ft) from the source.

The final experimental configurations are described in Table 2. Considered conditions include a fully customized setup (A), and a hidden reference with no spatialization (HR). To keep the experiment time short and not tiring participants, the other condition was chosen to have a representative set of all available arrangements, keeping the test set small:

- B: to evaluate the isolated effect of reverb, no HRTF spatialization has been applied along with a single channel (mono) corridor reverb and calibrated volume; no headphone calibration was employed;
- C: to evaluate the isolated effect of personalized HRTF, no reverb spatialization has been applied in a fully calibrated environment;
- D: fully customized environment with unmatched volume, to test the effect of volume on plausibility;
- E: complete auralisation with no customization and calibration with unmatched volume, to compare a fully customized environment with a generic auralisation;
- F: to evaluate the effect of reverb calibration, complete customization with no reverb calibration has been applied;
- G: to evaluate the effect of HRTF personalization, complete calibration has been applied with no HRTF personalization;
- H: HRTF personalization and reverb calibration; unmatched volume and no headphone calibration, to evaluate the effect of personalization without calibration

Table 2. Configurations available in the pilot test. HR is a hidden reference condition

	HRTF	Reverb	Vol. - db(A)	Hp Eq
HR	*none*	*none*	+70	*none*
A	personal	calibrated	+70	comp.
B	*none*	calibrated	+70	*none*
C	personal	*none*	+70	comp.
D	personal	calibrated	+75	comp.
E	kemar	large	+75	*none*
F	personal	large	+70	comp.
G	kemar	calibrated	+70	comp.
H	personal	calibrated	+75	*none*

3.4 Procedure

The participant entered the room, sat down, and completed the Informed Consent. The experimenter took and analyzed three photos of the subject (front and side of the head, ear closeup) to create the custom HRTF. [5] The HRTF files were transferred to the iPad app at the end of the procedure. The pilot experiment was explained to the participant, who wore AirPods and read the instruction on the iPad screen, which was put on a stand. To avoid front-back confusion issues [20, 21], the participant has been asked to navigate the room without turning toward the virtual speaker. The instructions also explained that if the iPad loose user tracking, the screen became black and the interface unusable. Once the tracking was restored, the interface returned to usable.

Using an iPad, all experimental conditions and the reference sound were administrated by the developed D&D MUSHRA GUI previously described which can be selected by touching the corresponding button. During the evaluation, the participant had to consider the following question:

”How similar is the virtual stimulus to the reference?”.

After a short tutorial about the usage of the GUI, the experiment started. To avoid accustoming the subject to

[5] For privacy issues: after the experimental session, all subject photos and anthropometric data were deleted.

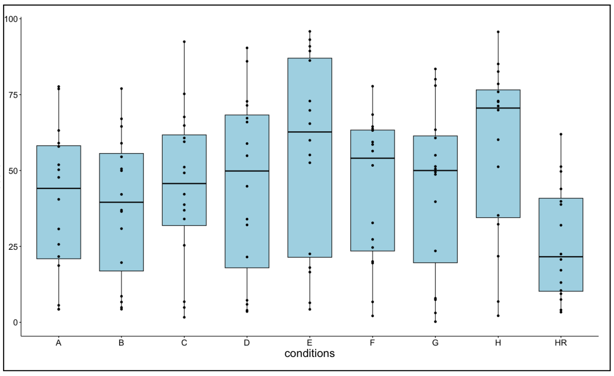

Figure 6. Boxplot of user ratings for each condition.

the stimulus. in the tutorial phase non-spatialized speech sound, played back through the real speaker, had been used. Even if it could be paused and restarted at any time, all participants carried out the experiment without interruption. Overall, the experiment lasted about 15 minutes, including of 5 minutes of configuration. At the end of the experimental task, each participant determined his/her individual score in tracking conditions.

4. RESULTS AND DISCUSSION

The pilot test aims mainly to evaluate the test procedure and the D&D MUSHRA interface. Collected rankings are shown in Fig.6. Repeated Measures ANOVA shows a significant effect of conditions, obtaining $F(4.06, 60.96) = 2.603$ ($p < 0.05$) and differences HR-C ($p < 0.01$), HR-E ($p < 0.01$), and HR-F ($p < 0.05$). These preliminary results exhibit a tendency for the H condition: HRTF and reverb personalization help in rendering a more plausible AAR. Extensive tests are needed to confirm this trend. Subjects typically rated the condition with a higher Volume (+75 db(A), including D, E, and H) with higher scores. This result can be attributed to the environmental noise of the corridor, which was frequently reported by subjects following the experimental session, and which acoustics were not controlled in order to obtain an ecological AAR. Presumably, increased volume improved hearing clarity, influencing participants' perceptions. Moreover, as mentioned in sec. 2, SDN parameters used in the virtual room simulation could influence the perceived plausibility, and can be improved by using other environmental measures such as RT60 decay. Another important bias factor can be given by the use of *Audio Transparency Mode* of Airpods, that can have altered the environmental room audio perception.

Users found it easy to operate on the D&D MUSHRA GUI and to navigate between conditions and the reference, and no subject reported problems. On the other hand, according to the experimenter, participants' tendency to compare experimental conditions to the real reference significantly impacts the evaluation results, making the requested comparison difficult. Often users did not compare stimuli with the reference, preferring a direct comparison between the conditions. For future tests, the D&D MUSHRA interface should be modified to implicitly require the participants to use the real reference more frequently.

User tracking was another important factor that diminished the significance of the test. In [22], the authors illustrate that subjects tend to adopt a movement that is consistent with the tracking method, resulting in the adaptation of the user to the system changing their behavior within the VE. Consequently, tracking methods could significantly affect the results. It is important to note that using consumer electronics devices has several advantages in design and implementation cost, experience reproducibility, and portability but often contains proprietary technologies that cannot be fully controlled and require a certain trade-off between accuracy and responsiveness. Considering Airpods, the manufacturer provides no specifications about IMU sensors available. Informal post-experimental user feedback reports significant tracking errors, probably due to no compensation or smoothing algorithms of orientation data that could lead to drift error over time and high tracking latency. During this pilot test, no information about the lost tracking rate has been recorded. About this topic, further accuracy measurement studies are advisable and our object of future studies.

5. CONCLUSION

In this pilot study, we implemented the D&D MUSHRA interface and applied it to an AAR evaluation experiment. The adopted scenario was developed in a non-laboratory environment using mobile AAR, which can be adapted to various contexts. Personalization and calibration of the VE help render a more plausible AAR, even though the number of subjects should be increased in future tests to confirm such a tendency. On the other hand, the specific contribution of each part of the system should be carefully evaluated in order to simplify the design of the VE. Moreover, the use of a noisy environment led participants to prefer conditions with higher volumes. According to the participants, the GUI was easy to use. However, the pilot test revealed some issues that need to be fixed for subsequent testing: (i) how the participants interact with the GUI should be adjusted to force them to use the real reference more frequently; (ii) additionally, user tracking should be carefully improved to obtain a more precise user position and orientation in the VE, while maintaining a trade-off between portability and accuracy. Even if these important factors significantly reduced the interpretability of the final outcome of the ranking test, this preliminary assessment allowed us to identify valuable insights for an ecological evaluation of mobile AAR scenarios.

6. REFERENCES

[1] J. Yang, A. Barde, and M. Billinghurst, "Audio Augmented Reality: A Systematic Review of Technologies, Applications, and Future Research Directions," *Journal of the Audio Engineering Society*, vol. 70, pp. 788–809, Nov. 2022.

[2] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl, "Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions," *IEEE Computer Graphics and Applications*, vol. 38, pp. 31–43, Mar. 2018.

[3] r. schlieper, s. preihs, and j. peissig, "an open dataset of measured hrtfs perturbed by headphones," *journal of the audio engineering society*, may 2022.

[4] S. A. Wirler, N. Meyer-Kahlen, and S. J. Schlecht, "Towards transfer-plausibility for evaluating mixed reality audio in complex scenes," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*, Aug 2020.

[5] C. Völker, T. Bisitz, R. Huber, B. Kollmeier, and S. M. A. Ernst, "Modifications of the MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) for use in audiology," *International Journal of Audiology*, vol. 57, pp. S92–S104, Sept. 2016.

[6] M. Geronazzo and S. Serafin, "Sonic Interactions in Virtual Environments: The Egocentric Audio Perspective of the Digital Twin," in *Sonic Interactions in Virtual Environments*, pp. 3–45, Springer International Publishing, Oct. 2022.

[7] R. Skarbez, M. Smith, and M. C. Whitton, "Revisiting Milgram and Kishino's Reality-Virtuality Continuum," *Frontiers in Virtual Reality*, vol. 2, p. 27, 2021.

[8] M. Geronazzo, "Egocentric Audio in the Digital Twin of Virtual Environments," in *2022 IEEE 2nd International Conference on Intelligent Reality (ICIR)*, pp. 7–10, Dec. 2022.

[9] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "Scattering Delay Network: An Interactive Reverberator for Computer Games," Audio Engineering Society, Feb. 2011.

[10] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A Minimal Personalization of Dynamic Binaural Synthesis with Mixed Structural Modeling and Scattering Delay Networks," in *Proc. IEEE Int. Conf. on Acoust. Speech Signal Process. (ICASSP 2020)*, (Barcelona, Spain), pp. 411–415, May 2020.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, Apr. 1979.

[12] M. Vorlander, *Auralization*. RWTHedition, Berlin, Germany: Springer, 2008 ed., Oct. 2007.

[13] C. Foy, A. Deleforge, and D. D. Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning," *The Journal of the Acoustical Society of America*, vol. 150, pp. 1286–1299, Aug. 2021.

[14] H. Bahu and D. Romblom, "Optimization and Prediction of the Spherical and Ellipsoidal ITD Model Parameters Using Offset Ears," in *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*, Audio Engineering Society, 2018.

[15] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A Bayesian Model for Human Directional Localization of Broadband Static Sound Sources," *Acta Acustica*, vol. 7, p. 12, 2023. Publisher: EDP Sciences.

[16] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The CIPIC HRTF Database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pp. 99–102, 2001.

[17] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini, "Applying a Single-notch Metric to image-guided Head-Related Transfer Function Selection for Improved Vertical Localization," *Journal of the Audio Engineering Society*, vol. 67, pp. 414–428, june 2019.

[18] F. Brinkmann and S. Weinzierl, "Audio Quality Assessment for Virtual Reality," in *Sonic Interactions in Virtual Environments*, pp. 145–178, Springer International Publishing, Oct. 2022.

[19] A. Lindau, "Spatial Audio Quality Inventory (SAQI). Test Manual.," tech. rep., 2014.

[20] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using Nonindividualized Head-Related Transfer Functions," *The Journal of the Acoustical Society of America*, vol. 94, pp. 111–123, July 1993.

[21] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do we need Individual Recordings?," *Journal of the Audio Engineering Society*, vol. 44, pp. 451–469, june 1996.

[22] A. G. Privitera, F. Fontana, and M. Geronazzo, "On the Effect of User Tracking on Perceived Source Positions in Mobile Audio Augmented Reality," in *CHItaly '23 - under review*, (Torino, Italy), Association for Computing Machinery, 2023.