



OPEN ACCESS

EDITED BY

Christos Mousas,
Purdue University, United States

REVIEWED BY

James Lewis,
Nottingham Trent University, United Kingdom
David Lindlbauer,
Carnegie Mellon University, United States

*CORRESPONDENCE

Ifigeneia Mavridou,
✉ i.mavridou@tilburguniversity.edu
Daniel N. Weiss,
✉ dweiss@bmhtllc.com

RECEIVED 16 May 2025

ACCEPTED 08 July 2025

PUBLISHED 31 July 2025

CITATION

Mavridou I, Seiss E, Ugazio G, Harpster M, Brown P, Cox S, Panchevski F, Erie C, Lopez D Jr., Copt R, Nduka C, Hughes J, Butera J III and Weiss DN (2025) "Did you hear that?": Software-based spatial audio enhancements increase self-reported and physiological indices on auditory presence and affect in virtual reality. *Front. Virtual Real.* 6:1629908. doi: 10.3389/frvir.2025.1629908

COPYRIGHT

© 2025 Mavridou, Seiss, Ugazio, Harpster, Brown, Cox, Panchevski, Erie, Lopez, Copt, Nduka, Hughes, Butera and Weiss. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

"Did you hear that?": Software-based spatial audio enhancements increase self-reported and physiological indices on auditory presence and affect in virtual reality

Ifigeneia Mavridou^{1*}, Ellen Seiss², Giuseppe Ugazio³,
Mark Harpster⁴, Phillip Brown¹, Sophia Cox⁵, Filip Panchevski⁵,
Christine Erie⁴, David Lopez Jr.⁴, Ryan Copt⁴, Charles Nduka⁵,
James Hughes⁴, Joseph Butera III⁴ and Daniel N. Weiss^{4*}

¹Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, Netherlands, ²Bournemouth University, Bournemouth, United Kingdom, ³University of Geneva, Geneva, Switzerland, ⁴Bongiovi Acoustic Labs, Port St Lucie, United States, ⁵Emteq Labs, Brighton, United Kingdom

Introduction: This study investigates the impact of a software-based audio enhancement tool Q6 in virtual reality (VR), examining the relationship between spatial audio, immersion, and affective responses using self-reports and physiological measures.

Methods: Sixty-eight participants experienced two VR scenarios, i.e., a commercial game (Job Simulator) and a non-commercial simulation (Escape VR), under both enhanced and normal audio conditions. In this paper we propose a dual-method assessment approach, combining self-reports with moment-bymoment physiological data analysis, emphasizing the value of continuous physiological tracking for detecting subtle changes in electrophysiology in VR simulated experiences.

Results: Results show that enhanced 'localised' audio sounds significantly improved perceived sound quality, immersion, sound localization, and emotional involvement. Notably, commercial VR content exhibited a stronger response to audio enhancements than non-commercial simulations, likely due to sound architecture. The commercial content featured meticulously crafted sound design, while the non-commercial simulation had numerous sounds less spatially structured, resulting in a less coherent auditory experience. Enhanced audio additionally intensified both positive and negative affective experiences during key audiovisual events.

Discussion: In this paper we propose a dual-method assessment approach, combining self-reports with moment-bymoment physiological data analysis, emphasizing the value of continuous physiological tracking for detecting subtle changes in electrophysiology in VR simulated experiences. Our findings support software-based audio enhancement as a cost-effective method to optimize auditory involvement in VR without additional hardware. This

research provides valuable insights for designers and researchers aiming to improve audiovisual experiences and highlights future directions for exploring adaptive audio technologies in immersive environments.

KEYWORDS

audio systems, virtual reality, sensor systems human-computer interaction, electrophysiology, signal analysis

1 Introduction

Virtual reality (VR) techniques have revolutionized the way users interact with simulated environments, providing high levels of immersion and the sensation of presence. Attaining high levels of immersion is crucial for delivering realistic and engaging VR experiences, with one of the key components being spatial audio (Warp et al., 2022; Hendrix and Barfield, 1996; Broderick et al., 2018; Kim et al., 2019; Kern and Ellermeier, 2020; Potter et al., 2022).

In VR immersive experiences, “localised” audio refers to the technique of positioning and rendering audio sources in specific locations within the virtual environment to create a sense of directionality and proximity for the listener (Naef et al., 2002). This provides spatial cues that mimic real-world auditory perception, thereby enhancing immersion. Spatial audio techniques have been applied across diverse domains, including entertainment, training, medical VR, and digital twins (Corrêa De Almeida et al., 2023). For example, in medical VR, increased immersion and presence have been linked to better therapeutic and training outcomes (Mahrer and Gold, 2009; Schiza et al., 2019; Brown et al., 2022; Standen et al., 2023; Cockerham et al., 2023).

While past research has focused predominantly on visual experiences, studies increasingly show that auditory properties - particularly spatial audio - significantly enhance presence and realism (Bormann, 2005; Broderick et al., 2018; Potter et al., 2022). Therefore, the design of appropriate solutions aiming towards increased audio fidelity, quality and spatialization can enhance the effectiveness of VR applications and improve patient outcomes.

Despite the growing interest in spatial audio and its potential benefits for VR applications, the effects of audio enhancement are not fully understood. Effective spatial audio design is complex, requiring knowledge of psychoacoustics, sound propagation, spatialization techniques (Hong et al., 2017). Furthermore, the auditory experience may be further enhanced when tailored to user preferences and context via a user-centered approach. This in turn may increase the subjective sense of presence without causing discomfort or disorientation. This is particularly important as discomfort and distortions, such as cybersickness, have been shown to negatively correlate with presence (Weech et al., 2019) thus potentially diminishing application effectiveness.

A potential solution that might greatly benefit VR immersive experiences is the creation of software-based tools to enhance audio, reducing reliance on expensive hardware and complex sound design. Such tools could help overcome hardware limitations and providing a consistent and immersive audio experience across various devices, environments, and content types (Kang et al., 2016). Moreover, software solutions could potentially offer additional customization

and adaptability options to suit different applications and user preferences, providing a flexible and cost-effective alternative.

Previous studies have shown that spatial audio can intensify emotional experiences in VR, particularly amplifying valence elicitation, and enhancing presence (Slater 2002; Rosén et al., 2019; McCall et al., 2022; Bosman et al., 2023; Huang et al., 2023). Techniques like binaural rendering and head-related transfer functions (HRTFs) can provide realistic directional cues, while audio processing (e.g., reverberation, propagation), and adaptive audio systems can contribute to increased fidelity and responsiveness (Antani and Manocha, 2013; Serafin et al., 2018).

Self-reports allow users to provide subjective feedback on their experience, including their perceived mental, cognitive, and emotional states (Slater, 1999; Bouchard et al., 2004; Tcha-Tokey et al., 2016). For this study, the two-dimensional model of valence and arousal of affect was chosen (Russel, 1980). Valence describes how negative or positive an experience is perceived, while arousal describes the physiological activation of an individual (i.e., calm to excited).

Additional physiological measures can continuously track changes of body functions over time whilst the participant is within an VR experience. For example, measures such as heart rate and skin conductance can detect arousal states (Wiederhold et al., 2001; van Baren and IJsselstein, 2004; Gao and Boehm-Davis, 2023). Facial electromyographic (EMG) sensors embedded within the VR headset can continuously measure spontaneous facial expression changes caused by emotional experiences within VR (Mavridou et al., 2018; 2021). These facial expressions mostly reflect pleasure (valence states). Generally, physiological measures do offer greater granularity in user experience responses to immersive content.

The present study investigates the impact of a novel software-based spatial audio enhancement tool, developed by Bongiovi Acoustics Labs (BAL), on both subjective and physiological experiences of VR users with standard commercial headset headphones. This tool uses BAL solutions algorithms to enhance audio based on perceived loudness and utilizes HRFT-based “V3D” technology for more natural sound localization. Human perception, attempting a more natural and immersive listening experience. The BAL solution also provides immersive localization of sounds using HRTF methods in their “V3D” algorithm. HRTF is a set of filters that adjust an input sound signal based on the anatomy of the listener’s head to mimic real world sound localization. By using custom HRTF parameters, developed in special acoustic labs, BAL has designed an HRTF solution specifically calibrated for VR experiences. The tool consists of two audio manipulation solutions: a Digital Power Station (DPS), aiming at enhancing the quality and clarity of the sound, and a Virtual 3D (V3D) model, aiming at creating a localised sound field.

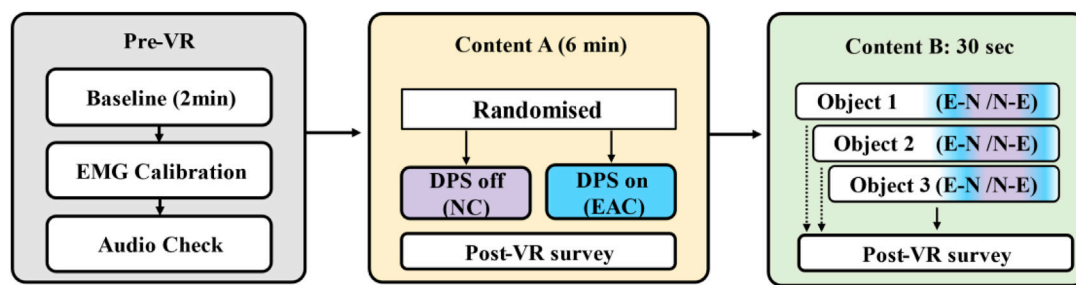


FIGURE 1

Study procedure overview. EAC and E stand for enhanced audio condition, NC and N stand for normal audio condition. The Pre-VR phase included headsets and sensor placement, calibration and baseline recording, followed by a quick audio check. Next, users experienced Content A (EscapeVR) experience with and without the audio enhancement (audio properties stayed the same across individuals). Upon completion of each condition, participants reported on all auditory presence items (from Witmer and Singer, 1998). The process was repeated again for content B (commercial content), which permitted controlled event-triggering. Four segments were recorded with both conditions, a 30-second background audio and three separate audio-emitting objects. After each repetition and condition, the user completed the post-VR survey. Once all events and contents were experienced, an optional brief semi-structured interview session took place.

We investigated the effects of the auditory enhancement on self-reported measures of the auditory experience, immersion, and the emotional experience. In addition we used the physiological measures of affect as described above. Two VR immersive experiences were used: a) a commercial game with 3D sound design that simulated an indoor office environment, and b) a simulated war zone with emotionally intense scenes. Both experiences were investigated under standard and enhanced audio conditions. Here, the effect of audio enhancement on the overall affective user experience was measured.

We analysed subjective reports and physiological measures from 58 participants in a within-subject design to investigate these four hypotheses:

Hypothesis 0 (H0): The enhanced audio mode condition increases audio quality, sound identification, sound involvement and sound localization compared to normal condition (manipulation check).

Hypothesis 1 (H1): The enhanced audio mode condition increases subjective levels of presence.

Hypothesis 2 (H2): The enhanced audio mode condition elicits the same range of auditory effects on the non-commercial Escape VR content (A) and the commercial Job Simulator VR content (B).

Hypothesis 3 (H3): Enhanced audio in VR can intensify affective valence responses in users - both for contextually positive and negative experiences. This was only tested for the Escape VR scenario (content A) because this is the affect inducing scenario.

Hypothesis 4 (H4): Enhanced audio settings in VR increase affect-based arousal responses in users. This was only tested for the Escape VR scenario (content A) because this is the affect inducing scenario.

For the analysis, a content-based together with an event-based analysis approach is proposed, taking advantage of the continuous signals available for the multimodal set-up tailored for this feasibility study, focusing on key moments within the experience where physiological and emotional reactions changes were expected (see all methods in Section 2).

2 Materials and methods

2.1 Participants

This study received ethical approval for the experimental protocol by the University of Geneva's ethical commission. For this study, 68 participants, mostly Bachelor students from the University 2 of Anonymous, were recruited. Their mean age was 28.65 years (± 10.54), and the self-identified gender split was 30 males and 38 females. From the total number of participants, ten were excluded due to noisy PPG signals. The remaining 58 datasets were used for the analysis of physiological data, whereas all participant data was utilized for the analysis of self-reported values. All participants received monetary compensation of 10 CHF.

2.2 Study design and procedure, and VR content

An overview of the study protocol and design is shown in Figure 1. The entire study protocol lasted approximately 1 hour. Upon reading the information sheet and signing the participant agreement form, participants were first asked to complete a short demographic questionnaire (see Section 2.6 Questionnaires), with age, experience with VR, fitness, and health-related questions to screen against the exclusion criteria. Post exclusion, participants were familiarised with the VR headset, and the sensor set-up. The OBS software was set up to record high-resolution video and audio from the VR content, the signal monitoring software, a front-facing webcam, and the system time using a clock windows widget (similar to the universal clock widget) as shown in Figure 2, which were used to monitor and time important moments in the user's experience.

After placing the sensor set-up, a two-minute-long physiological baseline (sitting still) was recorded for each participant. A short EMG calibration was then executed, in which participants were asked to perform three expressions with maximum intensity (happy - smiling, angry - frowning, and surprise - eyebrow raising). Individual calibration recordings with maximum EMG-registered

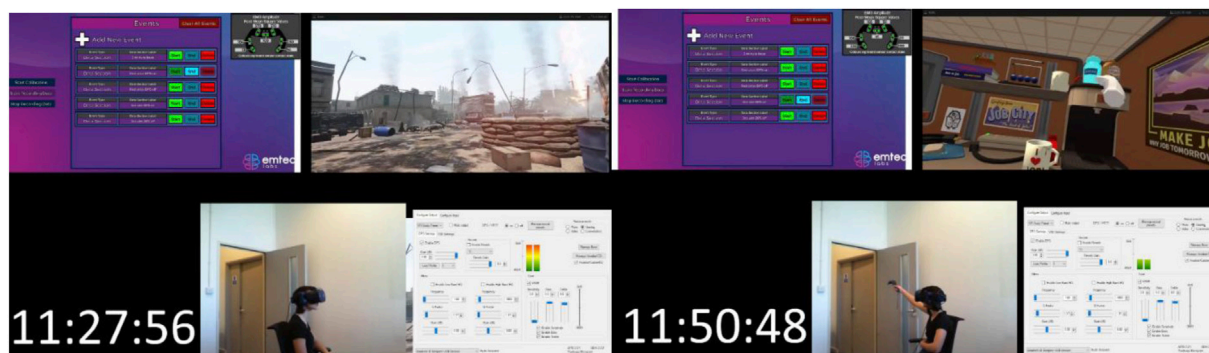


FIGURE 2

Examples of screen recording of the experiment for content A (left) and content B (right). In both, the top left window shows the event annotation app used. Top right window shows the view of the user in VR; left for the EscapeVR (Content A) and right for the Job Simulator (Content B). Bottom, left to right: current time, Camera view of the user, Audio enhancement DPS app.

muscle contractions were essential to permit the normalization of signals in the data processing and analysis stage.

After calibration, participants were taking part in two parts of the experiment, i.e., Part A and Part B. In both parts of the study, the participants experienced both audio conditions but they were not aware of the chosen audio condition at the time of presentation. Part A was presented before Part B to minimize the participant's bias towards focusing on the audio early on so as they could experience the events in Part A as they would in a normal VR experience. There was a short break between Part A and Part B.

Part A (Escape VR): The EscapeVR was provided by the International Committee of the Red Cross (ICRC). It featured a war zone fully immersive scenario, with 3D designed objects, characters, dialogs and interaction points. The user entered as a first-person view and quickly started to witness events taking place on the street ahead of them. In those, the user is trying to evade the conflict, by hiding and moving towards the red cross rescue team whilst trying to protect a young kid. The experience lasts on average 6 min, involving several events with dedicated background and event-specific localized sounds. Participants went through a six-minute VR experience, which simulated the attempted rescue of a child from a war zone. Each participant experienced the Escape V VR experience twice—once with enhanced audio (EAC) and once with normal audio (NC), with condition order counterbalanced across participants resulting in 31 participants starting with EAC condition. This counterbalancing was largely preserved in the final sample after exclusions (29 starting with EAC, and 32 starting with NC). The positive ending where the child and user were moved to safety was shown for all participants in both scenario repetitions.

Part B (Job Simulator VR, commercial, from Steam): The Job Simulator VR is an example of a fully immersive fictional VR simulation game, which is a type of game that attempts to model a real or fictional situation or process. The game is set in the year 2050, where robots have replaced all human jobs, and the humans can experience what it was like to work by using the Job Simulator. The game offers four different jobs to choose from. For this study the office worker was chosen. The game is set in an indoor office environment with ambient office sounds and little distractions. Once the user can complete a series of tasks in any order, using various objects (with spatially designed auditory cues and sounds)

that are available in the environment. All objects are interactive and inhibit natural movement behaviours due to the employment of physics. Participants were in this virtual environment for 30 s observing the various sounds, without interacting with anything in the room. The experimenter then asked the participant to remove the headphones and answer the PostVR Job Simulator questionnaire. Next, participants were instructed to interact with one of three virtual sound-emitting objects in the virtual space: the cafetiere, the telephone, or the copier. These objects were chosen because of the duration of the sounds, their interactive nature and the ability to trigger them multiple times for assessment. Participants were instructed to pay attention to the auditory changes made during the first 30 s (background sound) and when interacting with three objects, the “Copier,” the “Telephone,” and the “Cafetiere.” The experimenter switched between the audio conditions (presented in counterbalanced order) and asked the user to repeat the same interactions with the objects.

After the completion of each VR condition, the users were asked to answer the PostVR questionnaire. Here, the user was asked to rate the events in the PostVR questionnaire. The sound identification questions were not included for the three objects as all sound sources were identified for the purposes of the assessment.

2.3 Hardware

In total, two VIVE Pro Eye VR headsets with integrated EmteqPRO masks were used (see Figure 3). The VIVE Pro Eye is a premium VR headset comprising built-in eye tracking, a 110-degree field of view, a combined resolution of $2,880 \times 1,600$ pixels, and a frame refresh rate of 90hz. The audio system uses built-in on-ear “Hi Res Audio Certified” headphones, which were used for the study. While the VIVE Pro Eye provides eye-tracking capabilities, this data stream was excluded from analysis. Instead, physiological and expression data were recorded using the emteqPRO system. The emteqPRO system incorporates 1 photoplethysmographic (PPG), 7 electromyographic (EMG), and an inertial measurement unit or IMU sensor and allows data to be continuously recorded at 1,000 Hz fixed rate.



FIGURE 3
Equipment used in the study, the VIVE PRO eye headset, the built-in headphones and the EmteqPRO mask.

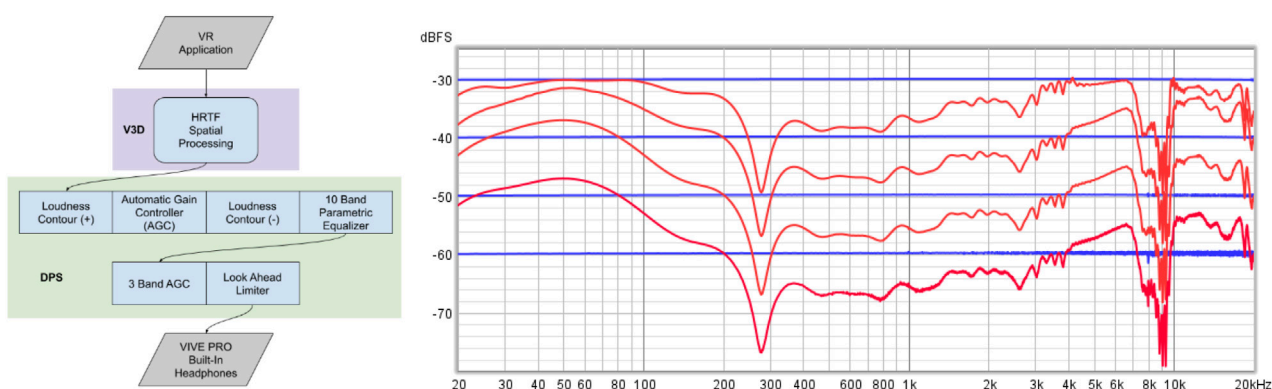


FIGURE 4
Left panel: Audio processing signal path block diagram. Right panel: The audio enhancement result used in this study (red traces) relative to the input level (blue traces) measured from the output of the headphone amplifier. Low frequencies are boosted in the Equalizer stage and attenuated by the low frequency band of the 3 band automatic gain controller (AGC) to deliver a bass enhancement effect. High frequency energy incurred during HRTF filtering and sound cue enhancement equalization is attenuated as part of the pre/post-compression blocks and the high frequency band of the 3 band AGC. Finally, audio signal clipping is prevented by a look-ahead limiter as observed in the highest input level trace (-30dBFS recorded into the Room EQ Wizard version 5.31.3 measurement system).

Photoplethysmogram (PPG) – PPG is monitoring physiological heart measures such as heart rate (HR in beats per minute) and heart rate variability (HRV). The PPG sensor on the emteqPRO is positioned over the middle of the forehead.

Electromyography (EMG) – EMG sensors over the zygomaticus, corrugator, frontalis, and orbicularis muscles are built into the emteqPRO system. Mean amplitudes of facial muscle movements give information about facial expressivity during VR experiences.

Inertial Measurement Unit (IMU) – The emteqPRO system's inertial measurement unit (IMU) is comprised of a magnetometer, a gyroscope, and an accelerometer. Each sensor produces data for the x, y, and z axes. These sensors enable tracking of head movements and motion information.

2.4 Software

The free Supervision software and OpenFace executable was used for the recording and monitoring of the electrophysiological signals. Valence and arousal detected scores were extracted from the recorded data using emteq cloud analytics (described in Gnacek

et al., 2021). In addition, a custom audio enhancement tool and GUI interface was designed which is described below.

The Audio Enhancement tool– To improve the auditory experience within the VR environment, we designed a custom Anonymous Digital Power Station (DPS) audio enhancement solution (version 2023), implemented as a Windows Audio Processing Object (APO). The system operates at 48,000 Hz with 32-bit floating-point precision and incorporates two low-latency modules: DPS (as an Effects APO) and V3D (as a Spatial APO).

Audio enhancement is achieved using a series of calibrated audio processing algorithms in a specific sequence (see Figure 4). First, the V3D module applies spatial processing using head-related transfer functions (HRTFs) derived from impulse responses recorded with a binaural dummy head and torso. Calibration involved optimizing reverberant characteristics, speaker-to-head distance, and post-convolution equalization. Audio channels are cross-mixed through HRTF filters into a virtual 360° speaker array, producing a stereo output spatialized for headphones. A panning algorithm, driven by a look-up table, determines the angular placement of input channels and supports arbitrary multi-channel configurations via channel mask parsing. Next, DPS algorithm dynamically adjust audio gain based on perceptual loudness, using pre-/post-

compression filtering and equalization to enhance quiet sounds without distorting dominant ones. This process also compensates for frequency response limitations of the headset and audio source, as well as changes incurred by HRTF filters. These advanced dynamic range controls also serve to adapt the audio output to the listener's equal-loudness contour (ISO 226:2023). The parameters (also referred to as “enhancement profiles”) of these algorithm blocks are tuned subjectively by experienced audio engineers to provide a desired listening result for the target application.

A customized enhancement profile—developed in collaboration with the Audio Neuroscience Organization (ANO)—was applied, optimized for HMD-based VR use. Real-time parameter tuning is possible via a companion GUI application during the testing phase. Floating point math is used throughout the audio processing modules. Both enhancements combined were tested in our study. In addition, a custom GUI control application was designed to allow real-time adjustment of the DPS properties (although for this study the VR settings preset provided by BAL was used across all participants).

2.5 Questionnaires

Two main questionnaires were employed for this study. The first, referred to as PreVR questionnaire was administered before the study, while the PostVR one was administered during the study, at the end of each content testing.

The PreVR questionnaire was administered prior to participants donning the VR headset. It encompassed inquiries regarding demographics (gender and age), VR and gaming experience, and health-related aspects. Experience with VR, computer games and experience of similar experiments were rated on a 1-9 scale, ranging from “not at all experienced” to “very experienced. Next, participants provided information on variables related to cardiovascular health, such as smoking, fitness and diseases as potentially confounding variables (Fisher et al., 2015; Van Elzakker et al., 2017; Kotlyar et al., 2018; Virani et al., 2020). Participants were screened for cardiovascular and respiratory conditions, previous disease diagnoses including psychological and neurological conditions, conditions affecting facial movements, hearing impairments, and the use of glasses for everyday activities and reading, specifying prescription strength, if applicable (see complete list in the [Supplementary Appendix SA](#)).

Post-VR Questionnaire: After each of the two VR experiences, participants rated their experiences on a 9-point semantic differential scale (1 = very negative, 9 = very positive, and 5 = neutral) on 10 single items. These ratings were collected immediately after each experience while participants were still wearing the VR headset. The first two items (valence and arousal) covered the affective impact of the experience and were only reported for the emotional stimuli part only. In item 3, the feeling of presence was reported, aimed at measuring overall the feeling of being there in the VR experience. Single item presence scores have been supported for VR studies as a reliable method whilst minimizing intrusion (Bouchard et al., 2004). Items 4 and 5 focused on experiences of discomfort and dizziness/nausea during the experience and were reported in the emotional stimuli only.

Next, items 6-10 (audio quality, auditory involvement, sound identification, and sound localization) were used to assess ability to evaluating audio clarity, and auditory presence factors related to involvement, identifying and locating sounds (extracted from items 20-22 from Presence Questionnaire by [Witmer and Singer, 1998](#)). The list of questionnaire items along with their anchors are provided in [Supplementary Appendix SA](#). With the exception of the audio quality question, these items were utilized in previous studies assessing the effect of spatial audio in immersive experiences ([Hendrix and Barfield, 1996](#); [Zhang et al., 2021](#)). At the end of the emotional Escape VR experience participants were also asked to describe particularly realistic or intense events within the VR experiences from memory. Once both audio conditions were experienced, participants were asked whether they identified differences between the conditions and its effect on their feelings, to gauge potential emotional changes between audio conditions.

2.6 Data analysis

The analysis process followed the order by which the hypotheses were stated, beginning with testing normality for both self-reported and physiological data, followed by extraction of condition-specific averages and measurements. Mean self-rating scores (and standard deviations) across participants were computed the self-report ratings. Differences between conditions in self-reported ratings were statistically evaluated using repeated measures ANOVAs, paired t-tests or non-parametric equivalence. Post-hoc tests were conducted and Bonferroni corrected for multiple comparisons where necessary. As the physiological data did not meet normality assumptions, Wilcoxon signed-rank tests were used for the pairwise comparisons (valence, arousal, HRV). An event-based analysis concentrating on affect changes in the physiological measures (e.g., helicopter event in content A), allowed for the examination of fluctuations in valence, arousal, and HRV over time. Specifically, the Helicopter event was chosen for illustrative purposes, given its intensity in and also its temporal and spatial clarity, which enabled more precise alignment with physiological data. Event-based analyses were not conducted on the Job-Simulator VR events because physiological segments were too short and events did not induce affect changes. Statistical analysis was conducted with Python and SPSS.

3 Results

3.1 Audio enhancement effects on perceived audio quality and auditory presence

3.1.1 Job simulation VR (Part B)

The self-reported scores on audio quality were compared between the EAC and the NC conditions as a manipulation check (related to H0) Auditory involvement, and sound localization were compared between the EAC and NC conditions as a way to understand how audio enhancement affected these audio factors pertaining to presence (related to H1). For the self-ratings, four sound events were analyzed, i.e., background sounds (first 30 s) and three VR objects (“Copier,” “Telephone,” and “Cafetiere”).

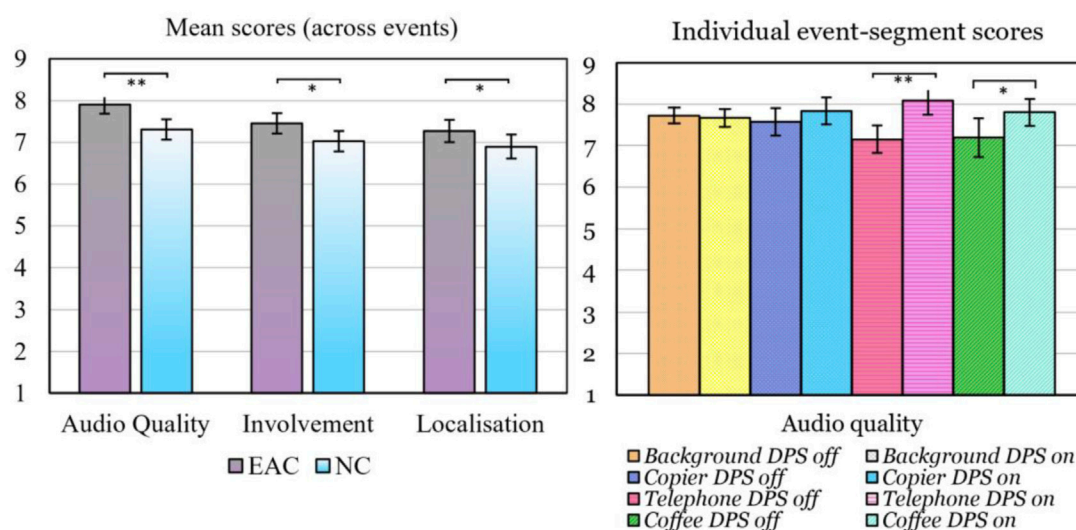


FIGURE 5

Average scores for audio quality, auditory involvement, and sound localization, for the normal (NC) and enhanced audio (EAC) conditions. SE error bars. Significant differences are indicated with (* for $p < 0.05$, ** for $p < 0.01$). This might indicate that audio quality was enhanced at the attended location only (as per task instructions) when tuning in could improve the perception of the specific task-relevant event. In case of the cafetiere, this audio quality change between audio conditions might not have occurred because the sound was already very loud, and a ceiling effect was already reached in the NC condition. Note, the task-irrelevant perception of sound in the periphery (background noises) was unaffected by the audio enhancement.

These were chosen because of the duration of the sounds, their interactive nature and the ability to trigger them multiple times during the simulation. Participants were instructed to pay attention to the auditory changes made during the first 30 s (background sound) and when interacting with three objects, the “Copier,” the “Telephone,” and the “Cafetiere.”

The ratings for audio quality, involvement and sound localization are presented in Figure 5. For audio quality ratings, a 2×4 ANOVA (factors: audio condition (EAC vs. NC) \times 4 events revealed a significant main effect of audio condition ($F(1, 63) = 24.49$, $p < 0.001$), meaning that the quality was rated higher for the EAC than for the NC condition (EAC Mean: 7.43 ± 0.35 vs. NC Mean: 7.11 ± 0.36). There was also a significant interaction between audio condition and event ($F(3,192) = 2.96$, $p = 0.034$), showing that significantly enhanced ratings for the EAC condition were found for the copier sound that was activated when interacted with ($p = 0.045$) and more static telephone sound ($p < 0.001$) but not for the background noise and cafetiere events (p 's > 0.05) (see Figure 5 right panel). Interestingly, the cafetiere audio was perceived as loud in all conditions.

Similar 2×4 ANOVAs were conducted for the two aspects of auditory presence, i.e., involvement and localization. For audio involvement, the main effect of audio condition was significant ($F(1, 63) = 6.79$, $p = 0.011$), again, showing that the EAC has a significant influence. For audio localization, the main effect of audio condition was significant too ($F(1, 63) = 5.95$, $p = 0.018$). This means that self-ratings for auditory presence were enhanced in the EAC compared to the NC condition, independent of the event type (see Figure 4) left panel.

Overall, ratings for auditory quality, involvement and sound localisation were enhanced for the EAC compared to the NC condition showing that the Job Simulator VR benefited from the enhanced audio settings. For the auditory quality ratings, these

enhancements depended on the event type. Significant changes were found for the copier and telephone events but not for the cafetiere and background events. This might indicate that audio quality was enhanced at the attended location only (as per task instructions) when tuning in could improve the perception of the specific task-relevant event. In case of the cafeteria, this audio quality change between audio conditions might not have occurred because the sound was already very loud and a ceiling effect was already reached in the NC condition. Note, the task-irrelevant perception of sound in the periphery (background noises) was unaffected by the audio enhancement.

3.1.2 Escape VR (Part A)

Participants reported ratings for audio quality, auditory involvement, sound localization and sound identification after experiencing the content with normal and enhanced audio settings. Mean ratings are displayed in Figure 6. Overall, audio quality ratings were rather low for both conditions for the overall experience, i.e., EAC ratings were 4.63 ± 0.42 and NC ratings were 5.08 ± 0.40 . In contrast, the auditory involvement, sound identification and sound localization ratings were higher. Ratings were compared between the EAC and the NC conditions by performing separate paired t-tests. Results showed that EAC significantly increased the feeling of involvement within the VR experience compared to the NC condition ($t(2) = 2.30$, $p = 0.025$) which is one of the three immersion-contributing auditory aspects. All other ratings were not significantly different between the EAC and NC conditions (p 's > 0.05).

The Escape VR scenario induced affective responses to the content. Hence, participants were asked to give additional ratings about valence, arousal, presence, discomfort and dizziness for the EAC and NC conditions. These are shown in Figure 7. Separate paired-t-tests for these ratings showed no difference between the two

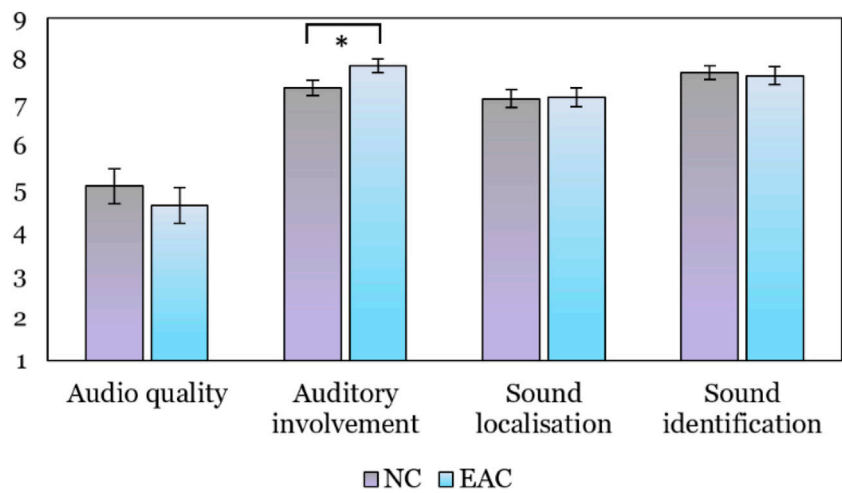


FIGURE 6
Subjective audio and sound ratings for the Escape VR experience for the EAC and NC conditions. Significant differences are indicated with (* for $p < 0.05$, ** for $p < 0.01$).

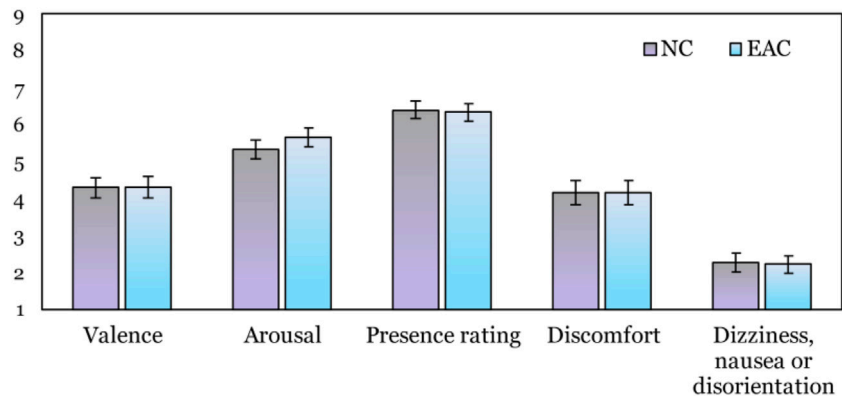


FIGURE 7
Subjective affective, presence and comfort ratings for the Escape VR experience for NC and EAC conditions.



FIGURE 8
The event called “Helicopter” featured a variety of auditory stimuli from intense fire exchange between a tank and a helicopter whilst the user stands in close proximity to both.

conditions (all t 's ≥ 0.4 , all p 's > 0.2), meaning that the EAC condition neither reduced or increased feelings of valence, arousal, or presence. EAC did also not change levels of discomfort or dizziness/nausea.

3.1.3 Event-based analysis–Introducing the “helicopter event”

The next step was to investigate effects of the EAC on specific events in the EscapeVR scenario. For this, participants were asked to

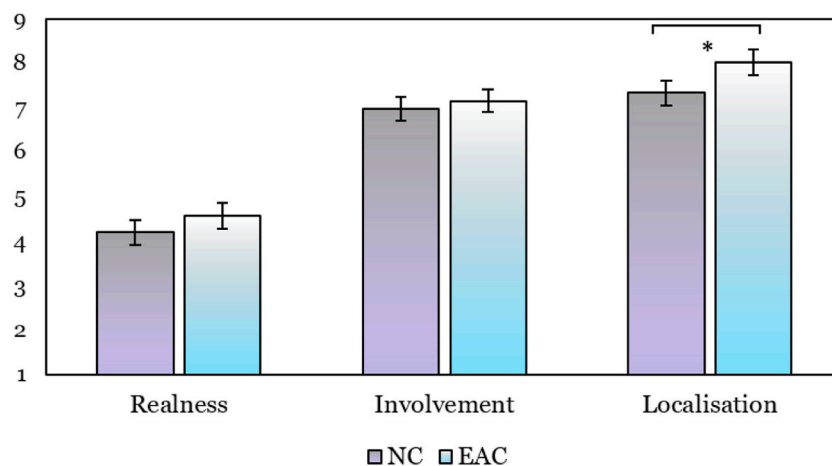


FIGURE 9
Post-VR single-item ratings for the helicopter event (content A), separately displayed for the NC and EAC conditions.

identify a key event in the experience from memory, which was the appearance of a helicopter followed by a bombing (short title: *helicopter event*), shown in Figure 8.

After identifying the helicopter event, participants were asked to rate several characteristics of the event, e.g., its level of realness, feeling of involvement (auditory immersive properties of the sound), and sound localization. These ratings are displayed in Figure 9. Realness of Helicopter event was generally low (EACmean = 4.63 ± 2.45 ; NCmean = 3.92 ± 2.11), and it did not significantly differ between the EAC and NC conditions ($t(47) = -1.39$, $p = 0.085$). The feeling of involvement was rated higher for both conditions, but there was no significant difference between the EAC and NC conditions (EACmean = 7.21 ± 1.67 ; NCmean = 7.012 ± 2.0 ; $t(55) = -0.531$, $p = 0.299$). More importantly, sound localization has significantly improved in the EAC compared to the NC condition (EACmean = 9.09 ± 1.10 ; NCmean = 7.45 ± 2.07), $t(55) = -2.017$, $p = 0.024$, showing that localising the relevant helicopter event and related sounds was enhanced in the EAC condition.

Overall, ratings for auditory quality, involvement and sound localisation indicated condition effects visible in closer inspection of the data using event-based analysis and not in experience-based one. The findings in this experience agree with the ones from the Job Simulator experience, on one of the auditory presence scores showcasing potential efficacy of methods across different types of content (related to H2).

3.2 Audio enhancement effects in emotional VR experiences: Physiological evidence

3.2.1 Valence detection with facial EMG signals

Valence could also be detected based on the facial EMG measurements using the existing insight models provided by emteq labs based on (Boxtel, 2010; Mavridou et al., 2021; Gnacek Michal et al., 2022). The valence scores for this measure ranged from 1 to 9. (1 = negative, 5 = neutral, 9 = positive). The Job Simulator and Escape VR condition was further analysed for the entire experience. Individual events during the Job Simulator were not analysed due to

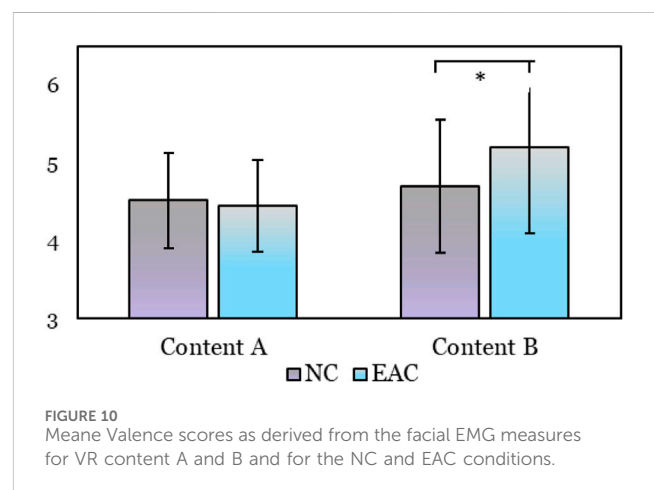


FIGURE 10
Mean Valence scores as derived from the facial EMG measures for VR content A and B and for the NC and EAC conditions.

their short duration and expected neutrality in emotional elicitation as a result of the user's interaction them (i.e., copier, telephone and cafetiere). In addition, the helicopter event segment within the Escape VR experience was analysed in more detail. For the Job Simulator condition we expected neutral to positive values, whereas we predicted neutral to negative values for the EAC condition. We also predicted that the valence scores become more affective (positive or negative) in the EAC condition (H3). The mean valence scores for both VR conditions are shown in Figure 10.

3.2.2 Job Simulator VR (Part B)

For content B, the valence scores were significantly more positive in the EAC condition (EACmean = 5.2 ± 1.1) compared to the NC condition (NCmean = 4.7 ± 0.86 ; Wilcoxon $t = 569.0$, $p = 0.0265$), showing that the enhanced audio settings had a clear effect on the valence scores derived from the facial EMG measures.

3.2.3 Escape VR (Part A)

The average valence scores for the entire experience was within the neutral to negative range for both conditions (EAC Mean $4.45 \pm$

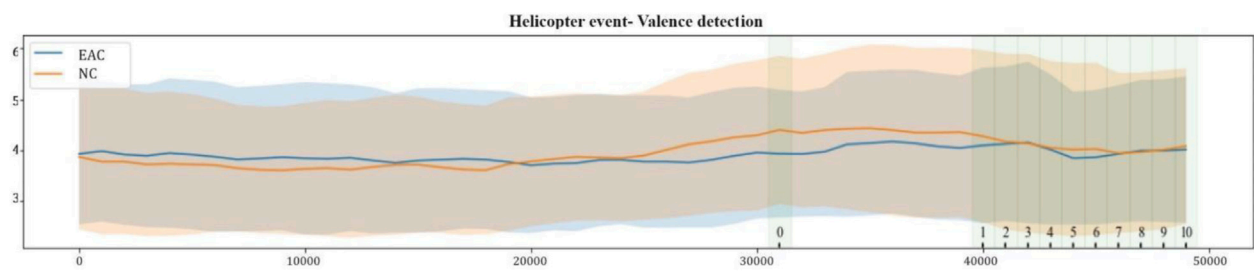


FIGURE 11
Mean valence detection scores for the helicopter event (Part A), separately for the NC and EAC conditions (lines with SD). The overall duration selected for the time window was 50 s, including 15 s before the onset of the event. Significant differences are shown on highlighted numbered time-windows (all t 's > 95.5, $p_0 = 0.44$ and all p 's(1-10) ≤ 0.001 , Bonferroni corr. Individual tests in [Supplementary Table SA2](#), [Supplementary Appendix SB](#)).

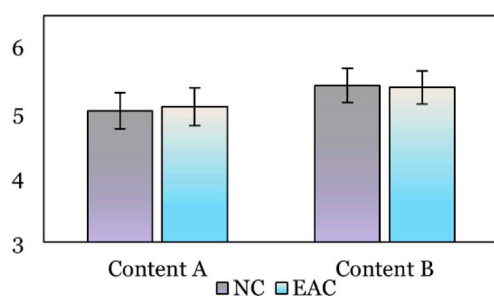


FIGURE 12
Average arousal for content A and Content B for NC and EAC conditions.

0.59; NC Mean: 4.52 ± 0.61), and they were not significantly different between conditions after conducting a paired-samples Wilcoxon test (Wilcoxon $t = 766.0$, $p = 0.630$). This is in line with the findings from the self-reported valence ratings.

A more in depth analysis of the change of valence over time was conducted on these continuous data because emotional events during the experience were expected to induce short-lived valence changes. For this, valence scores were extracted per second (from 30-second rolling) and were then averaged every 1 s across participants. These mean valence scores across participants are displayed separately for each of the two conditions in [Figure 11](#).

As previously, the *helicopter event* was selected as a key event in the experience due to its intensity and for its discreet audio triggered appearance. The event onset was synchronized across all participants' data using screen recorded timestamps, and a time-window of overall 50 s duration was selected for analysis. Individual data segments of 1 s duration during the 'helicopter' segment were compared with individual Wilcoxon tests between the EAC and NC conditions. Significant differences were found for several time windows (see green highlighted sections in [Figure 10](#) for significant differences (all $p < 0.04$). In those time windows, the EAC condition had more negative valence scores compared to the NC condition during the "helicopter" event. Valence scores return faster to neutral while the EAC levels remain more negative for longer (lower scores). Visual differences in the averages across participants per segment may appear subtle due to between-

subject variability, but statistical significance reflects consistent within-subject effects. Additional, subject-based graphs of physiological measures are shown in [Supplementary Appendix SC](#), and all tests are listed in [Supplementary Appendix SB](#).

3.2.4 Arousal detection from physiological sensors

Arousal was extracted from the data using the insight models (see Methods Software) was ranging from 1 to 9 (1 = calm; 5 - moderate arousal, 9 = extreme arousal), which was developed to use a range of features from all sensors of the system. Arousal is dependent and related to valence response intensity. A heightened arousal reaction as a result of sense stimulation was anticipated in the EAC (related to H4) in response to the spatial audio manipulation and localization enhancement, and the subsequent effects on felt immersion during the experience. For both VR conditions, the mean arousal values across the entire experience are displayed in [Figure 12](#). The arousal levels were moderate for all conditions (considered as around the score of "5").

3.2.5 Job Simulation VR (Part B)

A medium arousal level was expected as a result of its gaming contextual nature. This was indeed the case. Note, the audio condition did not have a significant effect on the arousal levels (EACmean = 5.39 ± 1.90 , NCmean = 5.42 ± 0.96 ($t = 565.0$, $p = 0.637$).

3.2.6 Escape VR (Part A)

For the entire content of the Escape VR experience, arousal levels were rather neutral which is not surprising because increased arousal was only expected predominantly in certain stressful events, like intense bombing events. The audio condition had no significant effect on these ratings (EACmean = 5.09 ± 0.39 ; NCmean = 5.03 ± 0.43 ; Wilcoxon $t = 531.0$, $p = 0.917$).

Delving deeper within the *helicopter event*, arousal segments were extracted as previously described for the valences cores. Wilcoxon tests indicated significant differences towards the end of the event ([Figure 13](#), highlighted sections: all $p < 0.05$), showcasing higher arousal levels for the EAC condition.

3.2.7 Heart rate variability (HRV) measures

Heart variability (HRV) metrics were extracted and analysed from the individual time-windows within the experience. HRV

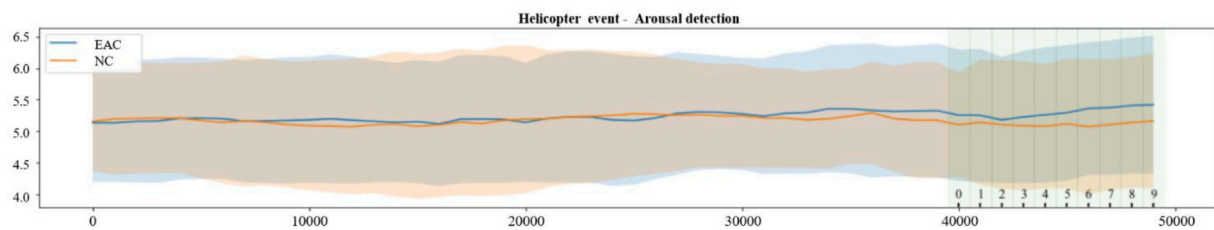


FIGURE 13
Mean arousal scores over time for the Helicopter event (Part A), separately for the NC and EAC conditions. All significant time-windows are highlighted and numbered from 0 to 9 (all t 's ≥ 24 , all p 's < 0.001 ; Individual tests in [Supplementary Table SA3](#)).

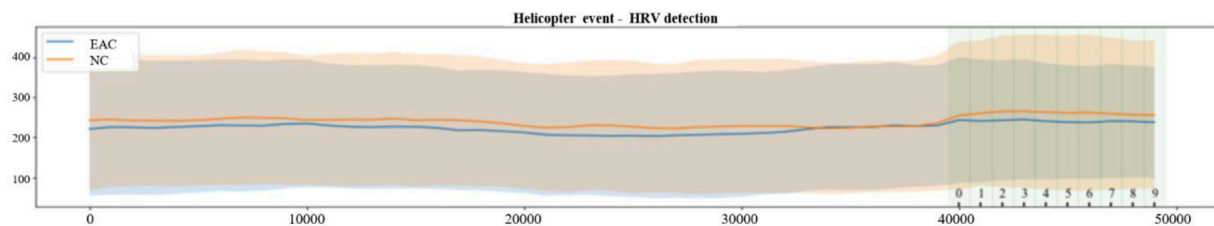


FIGURE 14
HRV RMSSD scores for the helicopter event (Part A), separately for the NC and EAC conditions. All significant time-windows are highlighted and numbered from 0 to 9 (all t 's ≥ 421 , all p 's ≤ 0.002 ; Individual tests in [Supplementary Table SA3](#), [Supplementary Appendix SAB](#)).

consists of changes in the time intervals between consecutive heartbeats called inter-beat intervals (IBIs). The fluctuations of a healthy, relaxed heart are intricate and regularly changing, which allow the cardiovascular system to rapidly adjust to homeostasis (Shaffer and Ginsberg, 2017). Even in healthy people, stress and elevated negative emotions could temporarily reduce the HRV (Kaczmarek et al., 2019) caused by enhanced autonomic nervous system activity (ref added). Therefore, in response to the EAC during content A it was hypothesized that as an extension of increased negative arousal, an overall reduction of HRV compared to NC conditions is present.

3.2.8 Job Simulation VR (Part B)

A similar average value was expected for HRV (RMSSD) in both conditions, per the effect of the audio enhancement on background sounds. In fact the audio condition did not have a significant effect on the HRV levels (EACmean = 142.4 ± 61.1 , NCmean = 140 ± 59.1 ($t = 727.0$, $p = 0.89$)).

3.2.9 Escape VR (Part A)

For the entire content of the Escape VR experience, the extracted HRV RMSSD values were not consistently low, which is not surprising because increased effects were expected predominantly in certain stressful events, e.g., helicopter bombing. The audio condition had no significant effect on these ratings (EACmean = 129.6 ± 46.8 ; NCmean = 139.2 ± 52 ; Wilcoxon $t = 447$, $p = 0.10$).

When investigating the HRV changes over time for the *helicopter event*, HRV was significantly reduced in the EAC condition compared to the NC condition for the last 10 s of the event (all $p < 0.05$) which was predicted. This is shown in [Figure 14](#).

Overall lower RMSSD equals higher autonomic nervous system activation. From these results it may be inferred that participants potentially felt higher levels of stress during the EAC condition compared to the normal one, especially for highlighted segments.

3.3 Self-observed auditory differences

Semi-structured interviews following the VR experiment, showcased subjective perceptions of the auditory experience, with links to affective events. In the Escape VR condition (Part A), from 68 total participants, 37 reported noticeable differences between the conditions. When asked to elaborate on those differences, participants reported changes in audio fidelity and intensity ($N = 26$). Of those ($N = 17$) reported that in the EAC condition objects and events seemed louder using terms such as “*more violent*”, “*realistic*”, or “*intense*”. For example, participant 8 reported that the missile event felt more violent in the EAC condition. Twenty participants (out of 31 who self-reported on sense of immersion) reported feeling more immersed in the virtual environment when the EAC condition was applied. Immersion was often linked to heightened emotional responses and a deeper connection to the virtual contents. For instance, participant 12 noted, “*the intense explosion in front of you, felt very realistic*” and Participant 35 mentioned that the scene “*when I was hiding behind the car with the refugees, felt realistic*”. In addition, participants ($N = 18$) also noted that sounds in the EAC condition felt “*easier to follow and localise*”, supporting improved directional ability and sound localization, allowing them to perceive the direction of sounds more accurately contributing to a heightened sense of immersion. For example, participant 18 stated that “*the helicopter was more noticeable in the*

first repetition [EAC condition]”. In addition, when asked about noticeable changes between the conditions, 12 participants described the change in feelings in the EAC condition. For example, Participant 46 stated that “he felt more anxious on the second experience [EAC condition] because he expected the events”. Fifteen participants provided additional comments, including observations of new details, changes in colour perception, and a sense of being placed differently in the virtual environment. For example, participants 5 and 22 noted that “the colors felt different/more realistic”, and participants 12 and 50 mentioned that they were “more focused on the [EAC] experience.”

Overall, in the EAC, sounds were accentuated and brought perceptibly closer to the user, resulting in an overall amplification of the auditory experience. This heightened audio presence potentially influenced the overall sense of auditory presence within the virtual environment. Particularly noteworthy were the substantial enhancements observed in events featuring specific ‘focused’ audio tracks. Activation of these tracks not only rendered them more intense, distinct and easier to localize but also fostered a deeper sense of involvement. This, in turn, facilitated the elicitation of emotional responses in response to intense events compared to standard VR audio configuration.

4 Discussion

The integration of auditory elements in VR has garnered substantial interest across diverse disciplines such as psychology, neuroscience, and human-computer interaction. VR is often conceptualized as a multisensory encounter and it emphasizes the pivotal role of audiovisual input (Finnegan et al., 2022), fostering three-dimensional immersion through auditory and visual engagement. Notably, within VR, external sensory information is confined to visual and auditory inputs, accentuating the crucial role of auditory stimuli in engendering a sense of presence and immersion (Weber et al., 2021). Spatial audio, in conjunction with three-dimensional graphics, field of view, and body tracking, has demonstrated efficacy in heightening user immersion. Immersion is notably advanced by headphone-enabled spatial audio in contemporary VR applications (Kapralos et al., 2008). Methodologies such as head tracking and HRTF facilitate sound localization, leveraging the omnidirectional pickup capability of the ears for a more natural listening experience (Begault et al., 1998).

In the present study, we investigated the impact of a software-based audio enhancement on various perceptual and physiological outcomes in VR. Participants experienced two types of VR content under both enhanced (EAC) and normal audio (NC) conditions. Our primary hypotheses concerned the effects of audio enhancement on perceived audio quality, sound localization, self-reported immersion, and emotional arousal/valence. Below, we discuss the findings in relation to each hypothesis, while considering methodological factors and alternative interpretations.

4.1 Enhancement improves perceived audio fidelity (H0)

This hypothesis served as a manipulation check to confirm that the enhanced audio mode had the intended perceptual effect. It was partially supported. In the Job Simulator scenario, enhanced audio

significantly improved participants’ ratings of audio quality, localization, and sound identification, indicating successful perceptual enhancement. However, in the Escape VR scenario, these benefits were either muted or absent.

A likely explanation lies in the baseline quality and structure of the audio content. Job Simulator contained clear, object-linked, and spatially varied sound sources, offering more “hooks” for enhancement to work with. Escape VR, by contrast, featured overlapping ambient sounds of similar volume and less spatial variety, potentially limiting the enhancement’s impact. Another contributing factor may be attentional framing. In Part A of the study, participants were not told to focus on audio, which could have reduced sensitivity to subtler changes. By Part B—where participants were asked to rate audio fidelity—perceptual differences became more pronounced. This reinforces the idea that perceptual sensitivity to enhancement is not just a matter of signal quality, but also of where the user’s attention is directed.

4.2 Enhancement increases subjective presence (H1)

This hypothesis predicted that enhanced audio increases the sense of presence which was the case for both VR scenarios. Participants reported significantly higher levels of immersion and presence under enhanced audio conditions. These findings suggest that software-based auditory enhancements—without requiring changes to hardware—can meaningfully enhance the users sense of being “present” in a virtual environment by enriching the auditory landscape, fostering an immersive multisensory experience.

Notably, the impact of auditory enhancement in VR experiences might be on both immersion and involvement, a distinction made by Witmer and Singer (1998). Immersion refers to the subjective experience of being enveloped in an interactive environment, whereas involvement is a psychological state resulting from directing attention to stimuli (Agrawal et al., 2020). As such, the effects of the auditory enhancement could potentially heighten both immersion and involvement and, hence, heighten attentional engagement. As with H0, attentional focus may have influenced outcomes. When participants were explicitly asked to evaluate audio in Part B, perceptual differences became more salient. This observation aligns with theoretical distinctions between immersion—largely driven by system fidelity—and involvement, which is shaped by user attention. Enhanced audio may support both dimensions of presence.

It is important to also note, that the presence measures employed in this study were not as comprehensive as in other studies. A more nuanced understanding of the implications on presence could be achieved by using standardized, multidimensional presence questionnaires in future research.

4.3 Audio effects are consistent across content types (H2)

This hypothesis was not supported. The pattern of results varied considerably between the two content types. Results using the more gamified, war scene content of the Escape VR scenario demonstrated

significant improvements in perceived involvement and immersion but no changes in audio quality and localisation. In contrast, audio quality, sound localisation and involvement were enhanced for the events in the Job Simulator VR condition. This commercial content featured spatially-emitted sounds for individual objects that users could interact with, as well as background audio. In contrast, the non-commercial simulation had numerous sounds triggered by specific story events that did not appear to be spatially designed. These sounds often blended at the same volume, regardless of user interactions or location, resulting in a less coherent auditory experience. This divergence underscores that audio enhancement benefits may be content-dependent and most pronounced when the underlying sound design is spatially structured. Enhancement techniques amplify what is already present. They do not compensate for weak or non-directional audio design.

Note, however, the ratings were for the overall six-minute application in the Escape VR condition where no specific events were selected for specific ratings. This might have blurred the picture. In contrast, the self-rating in the Job Simulator VR scenario were more specific and focused, task-relevant events. For these specific events, auditory quality, localisation and involvement were enhanced. This means auditory effects seem to be especially pronounced for items in the focus of attention highlighting the importance of attentional focus in interpreting the effects of audio enhancements. In Part A of the study, participants experienced the VR content without being directed to attend specifically to audio fidelity—an intentional design choice to preserve ecological validity and reduce demand characteristics. In contrast, Part B included explicit instructions to focus on the audio experience. The greater improvements in perceived audio quality and immersion observed in Part B may therefore reflect not only the objective effects of the DPS system, but also the modulatory role of directed attention. This distinction underscores a key methodological implication: the way in which participant instructions frame their perceptual focus can significantly influence reported outcomes in VR research. This interaction between attentional set and perceptual salience warrants further investigation. Future research should explore whether similar perceptual gains are evident without explicit instructions or whether attentional engagement is necessary for audio enhancements to be fully appreciated.

4.4 Enhancement modulates emotional valence (H3)

To test this hypothesis, only the Escape VR war-like environment was chosen to induce potential affective responses in participants which is an emotionally intense immersive experience. The hypothesis received partial support. Subjective self-reported emotional valence did not differ significantly between audio conditions. In addition, self-ratings for arousal and valence for the entire experience did not indicate the expected emotional intensities when data were collected retrospectively after the experience.

In comparison, physiological data indicated heightened affective responses - notably during the high-intensity helicopter scene in Escape VR under enhanced audio. This shows that physiological measures offer a complementary layer of information, capable of

detecting moment-to-moment changes in affective and experiential states. This was evident in the event-based analysis of the helicopter event, where spatial audio enhancement elicited distinct physiological responses associated with affect processing. While this nuanced approach proves to be a potentially more effective method for comprehending the impact of spatial audio enhancement in immersive experiences, caution in generalisation is warranted - single-event analyses may not fully represent the overall experience. For broader experiential insights, self-reports remain a valuable and appropriate tool.

This analysis highlights the additional value of continuous signal analysis and a limitation of retrospective self-report measures. Momentary affective reactions may be missed, if not captured through event-based or real-time metrics. At the same time, this finding support the use of facial electro-physiological data (fEMG) to detect more subtle emotional changes elicited by audio cues.

4.5 Enhancement increases arousal (H4)

Again, this hypothesis was only partially supported. While participants did not report significantly greater subjective retrospective arousal ratings, physiological measures suggested elevated arousal during emotionally salient moments in the enhanced audio condition.

As with H3, these results suggest the promise of continuous physiological monitoring in immersive research and highlight the importance of analysing specific affective events rather than relying on global averages. We acknowledge that the selection of this single event as the primary analytic focus could raise the concern of representativeness. The helicopter sequence featured spatially dynamic and intense audio stimuli, which is ideal for showcasing the benefits of enhancement. It is possible that in other parts of the Escape VR scenario, where audio cues were less distinct, may not had produced similar effects. Saying that, this event was chosen for illustrative purposes, given its intensity in and also its temporal and spatial clarity, which enabled more precise alignment with physiological data. Future studies could examine a broader array of events and avoid relying on a single event as a proxy for the entire experience.

Overall, the findings on valence and arousal (H3 & H4), advocate for a comprehensive study design for future investigations—one that integrates and analyses both subjective self-reports and physiological data to capture the full spectrum of user experience in immersive experiences. This nuanced methodology ensures a more accurate representation of the intricate and evolving nature of user experiences within immersive environments. We conclude that investigating the impact of audio on emotional and cognitive responses—both at an event-based and overall experience levels - could provide valuable insights into how tailored audio experiences contribute to the effectiveness and realism of VR applications across various domains, including gaming, education, therapy, and virtual tourism. It is likely that both the enhanced spatial separation and improved transient gain of the BAL enhancement tool contributed jointly to the perceived improvements, though this design did not permit disambiguation of individual enhancement types. Future

studies should isolate these factors to determine their relative contributions.

In conclusion, the study demonstrates a feasibility testing approach, highlighting the potential benefits and impact of enhanced audio tools not only within the realm of VR, but also in broader audiovisual contexts. This holds particular relevance for designers, indicating that high-quality auditory experiences can be achieved without the need for expensive hardware. These findings contribute to a growing body of research on the role of auditory involvement in immersive environments and the effectiveness of software-based sound enhancement. By emphasizing a dual methodology integrating self-reports and physiological data, we advocate for a more nuanced understanding of user experiences within immersive environments. The applicability of our findings extends beyond VR, suggesting broader implications for audiovisual content in diverse settings. As designers embark on creating new VR experiences, our research underscores the potential to optimize audio effects using techniques like the Enhanced Audio Condition (EAC), further enhancing the impact of immersive content.

5 Limitation and future directions

5.1 Participant pool, subjective traits and generalisability

The demographic of the participants were largely consisted of university students, which may limit the generalizability of the findings. In addition, some participants appeared to detect the differences easier than others during the study. Although all subjects reported good hearing ability, we expect that perhaps the ability to detect the enhanced audio output in both experiences could be linked to finer-grained subject differences in people. In future work, we could explore additional tests to evaluate hearing competency and frequency discrimination in spatial audio within VR. In addition, while enhanced audio significantly increased immersion ratings, the use of a limited number of immersion items—rather than a comprehensive questionnaire—warrants a cautious interpretation. The use of a more comprehensive presence instrument should be considered in future studies where experimental conditions allow for greater participant tolerance and reduced system complexity.

5.2 Reported issues in sound identification and localisation

In the Escape VR condition, sound identification and localisation were more difficult in the normal audio condition (NC) due the sounds. By comparison the Job simulator was less noisy, and easier for the participant to detect and recognise the source of the sounds. In the future, a custom application with controlled audio source triggers at prespecified locations could be designed in order to study the effects of enhanced audio identification and localisation in VR.

5.3 Event based analysis approach

These findings underline the importance of selecting multiple, diverse events for physiological analysis, as reliance on a single event—however salient—may limit their generalizability. Even so, the event analysis was more informative than the entire VR experience analysis, especially for experiences that are dynamic and interactive. In this study, we identified three distinct event segments within Content B and one in Content A (the Helicopter scene) to illustrate how specific moments can elicit strong emotional and physiological responses. These were selected based on their clarity, temporal alignment, and emotional salience. However, we acknowledge that other events, with differing contextual characteristics, may elicit different patterns of emotional and physiological responses.

Furthermore, in highly interactive VR environments, the timing and nature of event triggers can vary between participants, introducing variability that complicates direct comparisons. This variability should be considered when interpreting the results, and future studies should aim to include a broader range of event types and implement synchronization strategies to improve comparability across participants. For this automatic event annotation systems, such as the EmteqVR SDK, are crucial. They reduce human error and data processing time, for both more standardised passive and for interactive experiences. Importantly, predefined event annotations are very valuable particularly to designate the events and objects of interest for tracking, esp. when affective changes for these events are predicted based on the study design and background literature.

An alternative way is to inform the analysis of physiological data by the qualitative analysis of the participant responses to events. In our study the main event was selected based on the relevance to the scope of this work and by its affect-triggering impact on the participants, as reported by multiple participants during pilot sessions and the main study. This is a valid approach but, we acknowledge that other relevant shorter events with different emotional and auditory properties may have gone unnoticed.

5.4 Interpreting visualizations of dynamic data

In this study, we chose to present across-subject mean trajectories to visualize differences between conditions, as this is a common and interpretable approach in continuous affective and psychophysiological data analysis. However, this method inherently smooths individual variability and may mask consistent within-subject effects that are still statistically detectable. Particularly in dynamic, real-time experiences, such as those modelled here, individual responses may follow similar trends without aligning precisely in timing or magnitude. While statistical analyses can account for these within-subject differences, we acknowledge the group-level visualizations may not fully capture the richness of individual-level dynamics. Future work could benefit from complementary visualisations, such as exemplar single-subject plots or difference trajectories, to provide a more nuanced picture of moment-to-moment variability.

5.5 Audio driver compatibility

Future studies would benefit from testing with other VR headsets. The DPS enhancer worked with the HTC VIVE Pro Eye link box and audio driver.

Future work could explore audio enhancement in tailored VR experiences to optimize user engagement and immersion, incorporating personalised soundscapes, adaptive audio technologies, and spatial audio processing techniques. Integrating user preferences and physiological responses could further refine these enhancements, potentially improving overall user satisfaction and immersion levels in virtual environments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University of Gevena ethical commission (CUREG-2022-03-33). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

IM: Writing – review and editing, Writing – original draft, Investigation, Formal Analysis, Visualization, Conceptualization, Project administration, Methodology, Supervision. ES: Supervision, Conceptualization, Writing – review and editing, Methodology. GU: Resources, Validation, Supervision, Methodology, Writing – review and editing. MH: Investigation, Software, Methodology, Conceptualization, Writing – review and editing, Resources. PB: Writing – review and editing. SC: Writing – review and editing, Investigation, Data curation. FP: Writing – review and editing, Formal Analysis, Visualization. CE: Investigation, Software, Writing – review and editing. DL: Software, Investigation, Writing – review and editing. RC: Investigation, Writing – review and editing, Software. CN: Conceptualization, Project administration, Writing – review and editing, Funding acquisition. JH: Writing – review and editing, Funding acquisition.

References

- Agrawal, S., Simon, A., Bech, S., Bærentsen, K., and Forchhammer, S. (2020). Defining immersion: literature review and implications for research on audiovisual experiences. *J. Audio Eng. Soc.* 68, 404–417. doi:10.17743/JAES.2020.0039
- Antani, L., and Manocha, D. (2013). Aural proxies and directionally-varying reverberation for interactive sound propagation in virtual environments. *IEEE Trans. Vis. Comput. Graph.* 19, 567–575. doi:10.1109/TVCG.2013.27
- Begault, D. R., Ellis, S. R., and Wenzel, E. M. (1998). Headphone and head-mounted visual displays for virtual environments.
- Bormann, K. (2005). Presence and the utility of audio spatialization. *Presence Teleoperators Virtual Environ.* 14, 278–297. doi:10.1162/105474605323384645
- Bosman, I. de V., Buruk, O., 'Oz, Jørgensen, K., and Hamari, J. (2023). The effect of audio on the experience in virtual reality: a scoping review. *Behav. Inf. Technol.* 43, 165–199. doi:10.1080/0144929X.2022.2158371
- Bouchard, S., Robillard, G., St-Jacques, J., Dumoulin, S., Patry, M. J., and Renaud, P. (2004). "Reliability and validity of a single-item measure of presence in VR," in *Proceedings - 3rd IEEE international workshop on haptic, audio and visual environments and their applications - HAVE 2004*, 59–61. doi:10.1109/have.2004.1391882
- Broderick, J., Duggan, J., and Redfern, S. (2018). "The importance of spatial audio in modern games and virtual environments," in *2018 IEEE games, entertainment, media conference, GEM 2018*. doi:10.1109/GEM.2018.8516445

Project administration. JB: Writing – review and editing, Funding acquisition, Software, Project administration, Resources. DW: Supervision, Project administration, Conceptualization, Investigation, Writing – review and editing, Resources, Funding acquisition, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported in part by Bongiovi Acoustic Labs, which also funded the development of the audio enhancement tool used in this research. Several co-authors are affiliated with this organization and contributed to the design and implementation of the software. The funding body had no influence on the study design, data analysis, or interpretation of results.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2025.1629908/full#supplementary-material>

- Brown, P., Powell, W., Dansey, N., Al-Abbadey, M., Stevens, B., and Powell, V. (2022). Virtual reality as a pain distraction modality for experimentally induced pain in a chronic pain population: an exploratory study. *Cyberpsychol. Behav. Soc. Netw.* 25, 66–71. doi:10.1089/cyber.2020.0823
- Cockerham, M., Eakins, J., Hsu, E., Tran, A., Adjei, N. B., and Guerrero, K. (2023). Virtual reality in post-operative robotic colorectal procedures: an innovation study. *J. Nurs. Educ. Pract.* 13, 1. doi:10.5430/JNEP.V13N6P1
- Corrêa De Almeida, G., Costa de Souza, V., Da Silveira Júnior, L. G., and Veronez, M. R. (2023). “Spatial audio in virtual reality: a systematic review,” in *Proceedings of the 25th symposium on virtual and augmented reality*, 264–268.
- Finnegan, D., Petrini, K., Proulx, M. J., and O'Neill, E. (2022). Understanding spatial, semantic and temporal influences on audiovisual distance compression in virtual environments. doi:10.31234/OSF.IO/U3WC2
- Fisher, J. P., Young, C. N., and Fadel, P. J. (2015). Autonomic adjustments to exercise in humans. *Compr. Physiol.* 5, 475–512. doi:10.1002/CPHY.C140022
- Gao, M., and Boehm-Davis, D. A. (2023). Development of a customizable interactions questionnaire (CIQ) for evaluating interactions with objects in augmented/virtual reality. *Virtual Real.* 27, 699–716. doi:10.1007/s10055-022-00678-8
- Gnacek, M., Quintero, L., Mavridou, I., Balaguer-Ballester, E., Kostoulas, T., Nduka, C., et al. (2022a). AVDOS-VR: an affective video database with physiological signals and continuous self-ratings in VR. *Sci. Data* 9, Article 1. doi:10.1038/s41597-024-02953-6
- Gnacek, M., Broulidakis, J., Mavridou, I., Fatoorechi, M., Seiss, E., Kostoulas, T., et al. (2022b). emteqPRO—Fully integrated biometric sensing array for non-invasive biomedical research in virtual reality. *Front. Virtual Real.* 3, Article 781218. doi:10.3389/frvir.2022.781218
- Hendrix, C., and Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence Teleoperators Virtual Environ.* 5, 290–301. doi:10.1162/PRES.1996.5.3.290
- Hong, J. Y., He, J., Lam, B., Gupta, R., and Gan, W. S. (2017). Spatial audio for soundscape design: recording and reproduction. *Appl. Sci. Switz.* 7, 627. doi:10.3390/app7060627
- Huang, X. T., Wang, J., Wang, Z., Wang, L., and Cheng, C. (2023). Experimental study on the influence of virtual tourism spatial situation on the tourists' temperature comfort in the context of metaverse. *Front. Psychol.* 13, 1062876. doi:10.3389/fpsyg.2022.1062876
- Kaczmarek, L. D., Behnke, M., Enko, J., Kosakowski, M., Hughes, B. M., Piskorski, J., et al. (2019). Effects of emotions on heart rate asymmetry. *Psychophysiology* 56, e13318. doi:10.1111/PSYP.13318
- Kapralos, B., Jenkin, M. R. M., and Milios, E. (2008). Virtual audio systems. *Presence Teleoperators Virtual Environ.* 17, 527–549. doi:10.1162/pres.17.6.527
- Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., et al. (2016). Ten questions on the soundscapes of the built environment. *Build. Environ.* 108, 284–294. doi:10.1016/j.buildenv.2016.08.011
- Kern, A. C., and Ellermeier, W. (2020). Audio in VR: effects of a soundscape and movement-triggered step sounds on presence. *Front. Robot. AI* 7, 20. doi:10.3389/frobt.2020.00020
- Kim, H., Hernaggy, L., Jackson, P. J. B., and Hilton, A. (2019). “Immersive spatial audio reproduction for VR/AR using room acoustic modelling from 360° images,” in *26th IEEE conference on virtual reality and 3D user interfaces, VR 2019 - proceedings*. doi:10.1109/VR.2019.8798247
- Kotlyar, M., Chau, H. T., and Thuras, P. (2018). Effects of smoking and paroxetine on stress-induced craving and withdrawal symptoms. *J. Subst. Use* 23, 655–659. doi:10.1080/14659891.2018.1489008
- Mahrer, N. E., and Gold, J. I. (2009). The use of virtual reality for pain control: a review. *Curr. Pain Headache Rep.* 13, 100–109. doi:10.1007/s11916-009-0019-8
- Mavridou, I., Balaguer-Ballester, E., Seiss, E., and Nduka, C. (2021). Affective state recognition in virtual reality from electromyography and photoplethysmography using head-mounted wearable sensors. Available online at: <http://eprints.bournemouth.ac.uk/35917/> (Accessed August 16, 2022).
- Mavridou, I., Seiss, E., Hamed, M., and Balaguer-Ballester, E. (2018). “Towards valence detection from EMG for Virtual Reality applications,” in *12th international Conference on disability, virtual Reality and associated Technologies (ICDVRAT 2018) (Nottingham)*, 4–6. Available online at: <http://eprints.bournemouth.ac.uk/31022/>.
- McCall, C., Schofield, G., Halgarth, D., Blyth, G., Laycock, A., and Palombo, D. J. (2022). The underwood project: a virtual environment for eliciting ambiguous threat. *Behav. Res. Methods* 55, 4002–4017. doi:10.3758/s13428-022-02002-3
- Naef, M., Staadt, O., and Gross, M. (2002). Spatialized audio rendering for immersive virtual environments. *ACM Symposium Virtual Real. Softw. Technol. Proc. VRST*, 65–72. doi:10.1145/585740.585752
- Potter, T., Cvetković, Z., and De Sena, E. (2022). On the Relative importance of visual and spatial audio rendering on VR immersion. *Front. Signal Process.* 2. doi:10.3389/frsip.2022.904866
- Rosén, J., Kastrati, G., Reppling, A., Bergkvist, K., and Åhs, F. (2019). The effect of immersive virtual reality on proximal and conditioned threat. *Sci. Rep.* 9, 17407. doi:10.1038/s41598-019-53971-z
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi:10.1037/h0077714
- Schiza, E., Matsangidou, M., Neokleous, K., and Pattichis, C. S. (2019). Virtual reality applications for Neurological disease: a review. *Front. Robot. AI* 6, 100. doi:10.3389/frobt.2019.00100
- Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N. C., and Nordahl, R. (2018). Sonic interactions in virtual reality: state of the art, current challenges, and future directions. *IEEE Comput. Graph Appl.* 38, 31–43. doi:10.1109/MCG.2018.193142628
- Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5, 258. doi:10.3389/fpubh.2017.00258
- Slater, M. (1999). Measuring presence: a response to the witmer and singer presence questionnaire. *Presence Teleoperators Virtual Environ.* 8, 560–565. doi:10.1162/105474699566477
- Slater, M. D. (2002). Entertainment-education and elaboration likelihood: understanding the processing of narrative persuasion. *Commun. Theory* 12, 173–191. doi:10.1093/CT/12.2.173
- Standen, B., Anderson, J., Sumich, A., and Heym, N. (2023). Effects of system- and media-driven immersive capabilities on presence and affective experience. *Virtual Real.* 27, 371–384. doi:10.1007/s10055-021-00579-2
- Tcha-Tokey, K., Loup-Escande, E., Christmann, O., and Richir, S. (2016). “A questionnaire to measure the user experience in immersive virtual environments,” in *Proceedings of the 2016 virtual reality international conference* (New York, NY, USA: ACM), 1–5. doi:10.1145/2927929.2927955
- van Baren, J., and IJsselstein, W. (2004). Deliverable 5: Measuring presence: A guide to current measurement approaches. OmniPres Project IST-2001-39237, Project Report. Available online at: <https://orbilu.uni.lu/handle/10993/63956>
- Van Elzakker, M. B., Staples-Bradley, L. K., and Shin, L. M. (2017). The neurocircuitry of fear and PTSD. *Sleep Combat-Related Post Trauma. Stress Disord.*, 111–125. doi:10.1007/978-1-4939-7148-0_10
- van Boxtel, A. (2010). Facial EMG as a tool for inferring affective states. *Proc. Meas. Behav.* 2010, 104–108. doi:10.1007/978-1-4939-7148-0_10
- Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., et al. (2020). Heart disease and stroke statistics—2020 update a report from the American Heart Association. *Circulation* 141, E139–E596. doi:10.1161/CIR.0000000000000757
- Warp, R., Zhu, M., Kiprijanovska, I., Wiesler, J., Stafford, S., and Mavridou, I. (2022). “Moved by Sound: how head-tracked spatial audio affects autonomic emotional state and immersion-driven auditory orienting response in VR Environments,” in *AES europe spring 2022 - 152nd audio engineering society convention 2022*.
- Weber, S., Weibel, D., and Mast, F. W. (2021). How to get there when you are there already? Defining presence in virtual reality and the importance of perceived realism. *Front. Psychol.* 12, 628298. doi:10.3389/fpsyg.2021.628298
- Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* 10, 158. doi:10.3389/fpsyg.2019.00158
- Wiederhold, B. K., Jang, D. P., Kaneda, M., Cabral, I., Lurie, Y., May, T., et al. (2001). “An investigation into physiological responses in virtual environments: an objective measurement of presence,” in *Towards cyberpsychology: mind, cognitions and society in the internet age*. Editors G. Riva and C. Galimberti (Amsterdam: IOS Press), 175–183. Available online at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a5263deaba74e65ae7b847135c53bfe0f5ea0a90> (Accessed October 3, 2023).
- Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence Teleoperators Virtual Environ.* 7, 225–240. doi:10.1162/105474698565686
- Zhang, S., Feng, X., and Shen, Y. (2021). Quantifying auditory presence using electroencephalography. *Appl. Sci.* 11, 10461. doi:10.3390/AP112110461