



OPEN ACCESS

EDITED BY

Maria Limniou,
University of Liverpool, United Kingdom

REVIEWED BY

Benjamin Stodt,
Leibniz Research Centre for Working
Environment and Human Factors (IfADo),
Germany
Max Blümer,
University Medical Center Hamburg-
Eppendorf, Germany

*CORRESPONDENCE

Hendrik Himmelein,
✉ hendrik.himmelein@th-koeln.de

RECEIVED 24 July 2025

ACCEPTED 30 September 2025

PUBLISHED 16 October 2025

CITATION

Himmelein H, von Berg M, Pörschmann C and
Steffens J (2025) Influence of mental effort on
sound evaluations in virtual and real
experimental environments.
Front. Virtual Real. 6:1672595.
doi: 10.3389/frvir.2025.1672595

COPYRIGHT

© 2025 Himmelein, von Berg, Pörschmann and
Steffens. 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.

Influence of mental effort on sound evaluations in virtual and real experimental environments

Hendrik Himmelein^{1,2*}, Markus von Berg^{2,3},
Christoph Pörschmann¹ and Jochen Steffens^{2,3}

¹Institute of Computer and Communication Technology, TH Köln - University of Applied Sciences, Cologne, Germany, ²Audio Communication Group, Technische Universität Berlin, Berlin, Germany, ³Institute of Sound and Vibration Engineering, HS Duesseldorf - University of Applied Sciences, Duesseldorf, Germany

Psychoacoustic research increasingly relies on virtual reality (VR) to account for the complexity of acoustic scenarios and enhance the ecological validity of laboratory findings. However, recent studies suggest that virtual environments can alter mental effort compared to real-world settings, for example, through increased perceptual complexity which in turn may affect auditory perception. This could bias experimental outcomes and compromise the ecological validity of studies conducted in VR. To investigate this, a 2 × 2 between-subjects experiment was conducted to assess whether VR environments increase mental effort and thereby influence auditory perception. A real office environment was visually reconstructed in Unity and presented to the participants via a head-mounted display (HMD) and compared to its real counterpart. Participants in both environments were asked to retrospectively rate the loudness and unpleasantness of dynamically rendered binaural office noise scenarios presented via headphones and to report perceived sound sources. Moreover, participants were divided into two groups to induce different levels of mental effort. One group was asked to listen only to the sounds, while the other performed the Stroop Color-Word interference test in parallel. The results show no significant difference in the overall induced mental effort between environment conditions. Furthermore, performing the Stroop test had an effect on loudness and unpleasantness that was mediated by subjective effort. The results also suggest that auditory judgment depends primarily on individual sound properties, regardless of the visual environment.

KEYWORDS

psychoacoustics, virtual reality (VR), mental effort, auditory perception, ecological validity

1 Introduction

Audiovisual virtual reality (VR) has become an increasingly important tool in psychoacoustic research. Since environmental simulations are vital to studying human behavior in controlled experiment settings (Bishop and Rohrmann, 2003), VR provides a manifold pool of possibilities and enables researchers to account for the complexity of recreated audiovisual surroundings (Higuera-Trujillo et al., 2017). However, as laboratory settings are deliberately designed to eliminate confounding variables, a transfer to real-life conditions may result in differences in experience, perception, behavior, or judgment (Tarlao et al., 2022). Hence, the generalizability of laboratory findings and congruence of

psychological processes to real-life situations, known as ecological validity, has become a common matter of interest (Aletta and Xiao, 2018). In this paper, we follow a definition by Keidser et al., who defined ecological validity as “the degree to which research findings reflect real-life hearing-related function, activity, or participation” (Keidser et al., 2020). To achieve a high degree of ecological validity in laboratory settings, it is necessary for the experimental design to incorporate both physical and behavioral realism (Ferwerda, 2003). Physical realism describes the congruity of identification and physical properties (e.g., surface structure) of objects or structures (Ferwerda, 2003), while behavioral realism describes response-behavior to an appropriate task that is similar to the real environment (Sheridan, 1992).

One possibility of achieving high behavioral and physical realism while maintaining experimental diligence is the use of VR (Meyer et al., 2012; de Kort et al., 2003). Generally, VR is widely applied in various disciplines, for example, medicine (Frederiksen et al., 2020), education (Makransky et al., 2019) or piloting (Dehais et al., 2013). The approach used in this study enables visual and auditory simulations to be presented via a head-tracked Head-Mounted Display (HMD) and synchronized dynamic binaural headphone reproduction, which is emerging as a popular method to achieve high ecological validity (Xu et al., 2021). HMDs, combined with binaural sound reproduction, can facilitate behavioral and physical realism, as well as a high sense of immersion in the virtual environment. However, while striving for realistic and ecologically valid VR environments, audiovisual interaction can significantly affect auditory perception (Tarlao et al., 2022). In particular, judgments about unpleasantness and loudness can differ between realistic VR laboratory and *in situ* settings due to differences in reaction to sound events (Hohmann et al., 2020; Li and Lau, 2020). Hohmann et al. (2020) showed that *in situ*, a clear reaction to a sound event could be observed, while in the lab, this behavior was not as pronounced. Given these findings, it is crucial to closely examine contextual aspects such as multimodal processes involved in auditory perception in simulated virtual environments (Hermida Cadena et al., 2017).

Additionally, in everyday life, the auditory environment is often not the primary focus of attention; it is perceived while executing various other activities and tasks. In this context, acoustic events can either be viewed as unconscious distractors of a primary task, possibly shifting attention, or as a deliberately monitored secondary cue parallel to another primary focus of attention. This study focuses on the latter multitasking scenario, where auditory perception is one of two tasks. Research focusing on these dual-task scenarios demonstrated that the addition of a secondary task can affect mental effort and shift the allocation of attention resources (Gagné et al., 2017; Picou and Ricketts, 2014). Thus, as attention resources are limited (Kahneman, 1973), the success of managing both tasks without interference depends on both tasks' combined required mental effort, the deliberate involvement of cognitive control resources (e.g., working memory) necessary to meet the cognitive demands of the tasks (Westbrook and Braver, 2015), or the cognitive load (Lavie et al., 2004; Fisher et al. (2019)). Since people generally tend to focus their attention on tasks which require less cognitive control (e.g., which are less demanding) (Kool and Botvinick, 2018), motivation can be a factor to regulate peoples' willingness to invest their cognitive resources in a given task at hand (Braver et al., 2014).

The terms perceptual load, cognitive load, and mental effort are sometimes used inconsistently and ambiguously in the literature. In an attempt to combine previous findings, this study employs the term *mental effort* to describe the deliberate use of mental resources to complete a specific task, following two specifications. First, the mental resources comprise perceptual (e.g., working memory capacities used to process sensory input in different modalities) and ‘higher-order’ cognitive resources (e.g., resources for the organization, evaluation, and storing of information over longer time spans) (Fisher et al., 2019). Second, the investment of mental resources (i.e., the mental effort) depends on the task-specific perceptual (i.e., amount and complexity of sensory information to be perceived) and cognitive load (i.e., the number of organizational and evaluative operations to complete the task) on the one hand, and on the subject-specific *willingness* to employ these mental resources on the other hand (Westbrook and Braver, 2015).

In a virtual environment, the demands to organize visual and auditory information (i.e., the perceptual load), are probably elevated in the course of, for instance, correcting mismatches of visual representations of real objects, compensating for the absence of the own body, or recognizing visual objects as sources of sound that are not well localizable through the applied auralization. A task's cognitive load, such as recalling auditory or visual information, would, on the other hand, not be expected to be raised in a virtual environment. In this context, Fisher et al. (2019) used different versions of a video game played on a computer to separately manipulate perceptual load by increasing the visual complexity of the displayed scene, and cognitive load by adding complexity to the video game's point-scoring logic. During this game, participants had to respond to sudden auditory and visual stimuli as a secondary task, and the response times were used as a measure of available mental resources in each condition. They concluded that additional perceptual load only affects modality-specific resources (i.e., response times only increased in case of a visual secondary task), whereas cognitive load draws on shared resources (response time increased for auditory and visual stimuli in the secondary task). Therefore, a virtual environment's increased perceptual load would rely on modality-specific resources, so that an increased visual load would not interfere with auditory perception and vice versa. By contrast, multiple tasks with high cognitive load would be prone to interference, both in virtual and real-world scenarios. As a consequence, the virtual environment might increase the perceived effort, but would not be expected to raise the risk of interference between audio and visual tasks compared to a real-world scenario.

So far, only a limited amount of research has explicitly compared differences in mental effort in classical 2D monitor-based environments and VR. In addition, conflicting indications can be found as to whether mental effort increases or decreases in virtual environments. In particular, while comparing traditional computer-based studies to those in VR environments, an increased mental effort has been observed in various domains, for instance, for complex motor tasks (Juliano et al., 2022), surgical training (Frederiksen et al., 2020), or education (Makransky et al., 2019). By contrast, more recent studies found that comparing virtual environments to real-world counterparts like navigation in a subway station (Armougum et al., 2019), perceiving daylight spaces (Chamilothori et al., 2019), as well completing the Wisconsin Card

Sorting Test (Maneuvrier et al., 2023) and the N-back task (Luong et al., 2019), or reaching for objects (Wenk et al., 2023) did not significantly alter mental effort in the virtual environment. Furthermore, a study by Wenk et al. (2019) comparing motor training in VR with classical screen-based environments found reduced mental effort in VR conditions. These findings are corroborated by a recent study by da Silva Soares et al. (2024), who used neuroimaging to access mental effort of a puzzling task in VR and screen-based environments and observed higher neural efficiency in the prefrontal cortex (i.e., lower task load) in VR.

Thus, we hypothesized that the VR environments would impact auditory perception and evaluation of sound due to additional induced mental effort, possibly resulting in increased interference between modalities and threatening the ecological validity of empirical findings. Therefore, the current study aims to compare auditory perception in an everyday life-like situation re-created in the lab (i.e., working in an office) and its counterpart in an audio-visual VR environment. Importantly, this study deliberately chose to vary only the visual environment while maintaining a consistent head-tracked dynamic binaural auralization. This aligns with widely used psychoacoustic experiments using audio-visual virtual environments to research auditory processing under controlled, ecologically valid experimental settings without introducing uncontrollable confounding factors inherent to real auditory environments. To do so, we performed a 2×2 between-subject experiment under two visual environments (VR vs real) and two load (high vs low) conditions. In detail, we reconstructed a real office space at TH Cologne in Unity and presented it to the participants in the VR conditions via a HMD. Participants in all conditions were asked to retrospectively rate dynamic binaural office soundscapes presented via headphones concerning their loudness and unpleasantness, and to recall perceived sounds. To impose a constant cognitive load for the sustained investment of mental effort, participants in the high-load condition additionally performed the Stroop Color-Word interference test (Stroop, 1935). By contrast, participants of the low-load condition solely listened to the presented sounds.

2 Methods

2.1 Participants

Sixty-one participants with self-reported normal hearing and a mean age of 31.1 years ($SD = 12.2$ years, 18.7% female) were selected from a sample of 63 (two participants were excluded due to either missing data or incorrect experiment execution). Most participants were undergraduate students recruited from the TH Köln - University of Applied Sciences mailing lists. Prerequisites to participate were fluency in German and either no visual impairment or wearing contact lenses (wearing glasses was not admissible). Participants were randomly assigned to one of the four experimental conditions. A prior power analysis indicated that a total sample size of 60 participants was required for a repeated measure ANOVA with a small effect size of $d = 0.20$, standard alpha level of 0.05, and power of 0.95 with four groups and 16 repeated measures. Since the low load condition was implemented as a baseline condition, a slight imbalance can be

observed in the participant distribution: Twenty participants in the VR high load condition (mean age: 30.5, $sd = 10.1$ yrs, 20% female), 20 participants in the real high load condition (mean age: 27.2, $sd = 8.5$ years, 20% female), 12 participants in the VR low load condition (mean age: 33.4, $sd = 13.4$ years, 33% female) and nine participants in the real low load condition (mean age: 37.8, $sd = 16.4$ years, 34% female). In each of the four conditions, participants could get one of five cinema vouchers for a popular cinema in Cologne. In the high-load conditions, the vouchers were given as a reward-based performance in the Stroop test. This was done to increase participants willingness to invest mental effort in the presented task. In the low-load conditions, the vouchers were randomly distributed among participants.

Before the experiment, informed consent was obtained from all participants. This study was approved by the Commission on Responsibility in Science Ethics Committee at TH Cologne (Application number THK-2023-0003).

2.1.1 Questionnaires

Before the main part of the experiment, participants were asked to report on several demographic variables, such as gender, age, and level of education. Normal hearing was assessed based on participants' self-report. Additionally, participants in the VR environment filled out the Slater-Usch-Steed (SUS) presence questionnaire (Usoh et al., 2000) to rate their immersion in the virtual environment.

Subjective mental effort was obtained at the end of the experiment using the NASA Task Load Index (NASA TLX) (Hart and Staveland, 1988). The test consists of six items: 'mental demand', 'physical demand', 'temporal demand', 'performance', 'effort', and 'frustration', which are retrieved from six continuous 20-point Likert scales from 0 to 1 and which are averaged to obtain an overall score for mental effort (Augereau et al., 2022; Bustamante and Spain, 2008). In this study, the 'performance' item was excluded because there was no substantial correlation with the other items and most participants reported high difficulties in self-evaluating their performance in the given tasks and/or misread the scale description, as this is the only reversely-scored item in the NASA TLX.

2.2 Visual environments

2.2.1 Real environment

The selected real environment was a common two-person office at TH Cologne, and we assumed that most participants would be reasonably familiar with such a space. In addition, virtual office spaces have been established in the literature, facilitating behavioral realism (Alvarez et al., 2008; Klinger et al., 2005; Macedonio et al., 2007). In preparation for the experiment, the office space was equipped with typical office items, which functioned solely as visual sources for auditory stimuli (see Figures 1, 3). In addition, the relevant acoustics and geometries of the room were measured. The total size of the office space was $6.89 \text{ m} \times 5.85 \text{ m} \times 3 \text{ m}$ and the reverberation time averaged over frequency was $T_{60} = 0.43 \text{ s}$ ($T_{60@200\text{Hz}} = 0.49 \text{ s}$, $T_{60@1000\text{Hz}} = 0.37 \text{ s}$, $T_{60@5000\text{Hz}} = 0.58 \text{ s}$, $t_{60@12500\text{Hz}} = 0.45 \text{ s}$). Since the sound sources were placed rather far away from the listener, their direct-to-



FIGURE 1
Participants' view of the real-life office space at TH Köln, including the experimenter (first author).



FIGURE 2
Participants' view of the virtual recreated office space, including a virtual experimenter.

reverberant ratios (DRR) were relatively low (e.g., $DRR_{\text{Clock}} = -4.4$ dB, $DRR_{\text{Printer}} = -1.9$ dB, $DRR_{\text{Window}} = -7.7$ dB).

2.2.2 VR environment

A virtual counterpart of the actual office space was created following the suggestions by Fuchs et al. (2011), who highlighted the role of immersion, interaction, and interface. According to the authors, immersion is achieved through multimodal representation, mainly vision and hearing, while incorporating haptic feedback and changeability through an intuitive and easily understandable interface that also promotes interaction (Souza et al., 2022). Consequently, the office space was carefully recreated in VR using the 3D software Blender (Blender, 2022) and the Unity Asset Store (Unity Technologies, 2022) to match its real-life counterpart. Pictures from the real environment (e.g., view outside the window) and natural virtual lighting were included to enhance immersion. Figure 2 depicts the virtual office space, including the experimenter.

During the VR experiment, the virtual environment was presented to the participants using the HTC Vive Pro (VIVE, 2022) with a low-persistence resolution of 2880×1600 px set at 90 Hz and a field of view of 110° . The simulation was rendered on a Desktop PC Win10, Intel Core i9 9900K (3.6 GHz), 24GB RAM, and a Nvidia GeForce RTX 2080 graphic card.

To incorporate haptic feedback, HTC Vive Trackers (VIVE, 2022) were taped to a chair, keyboard, and trackpad, allowing participants to interact with these components in the virtual duplicates freely. Although the virtual environment was designed to enable 6-Degree-of-Freedom (6DoF) movement, participants were instructed to sit in front of the virtual monitor and only move their head (3DoF) throughout the experiment.

Since participants were still physically present in the real office, only the visual environment changed when wearing the HMD. At the same time, other factors such as temperature remained the same as in the real environment conditions.

2.3 Virtual auditory environment

In both environments, participants were presented with auditory stimuli via headphones consisting of typical office

sounds reported in the literature (Hochschule Luzern -Technik und Architektur, 2010; Kim and de Dear, 2013; Kreizberg, 2021), namely, a clock, a coffee maker, two talking colleagues (unintelligible babble noise), a keyboard and mouse, an office printer, air conditioning (fan), and traffic noise coming from two open windows, which were recorded at a distance of 1.5 m in the anechoic chamber at TH Cologne using a TLM 102 microphone and an RME Babyface Pro. Outdoor noise was recorded *in situ* from the target office space using the same equipment and at a similar distance from the windows.

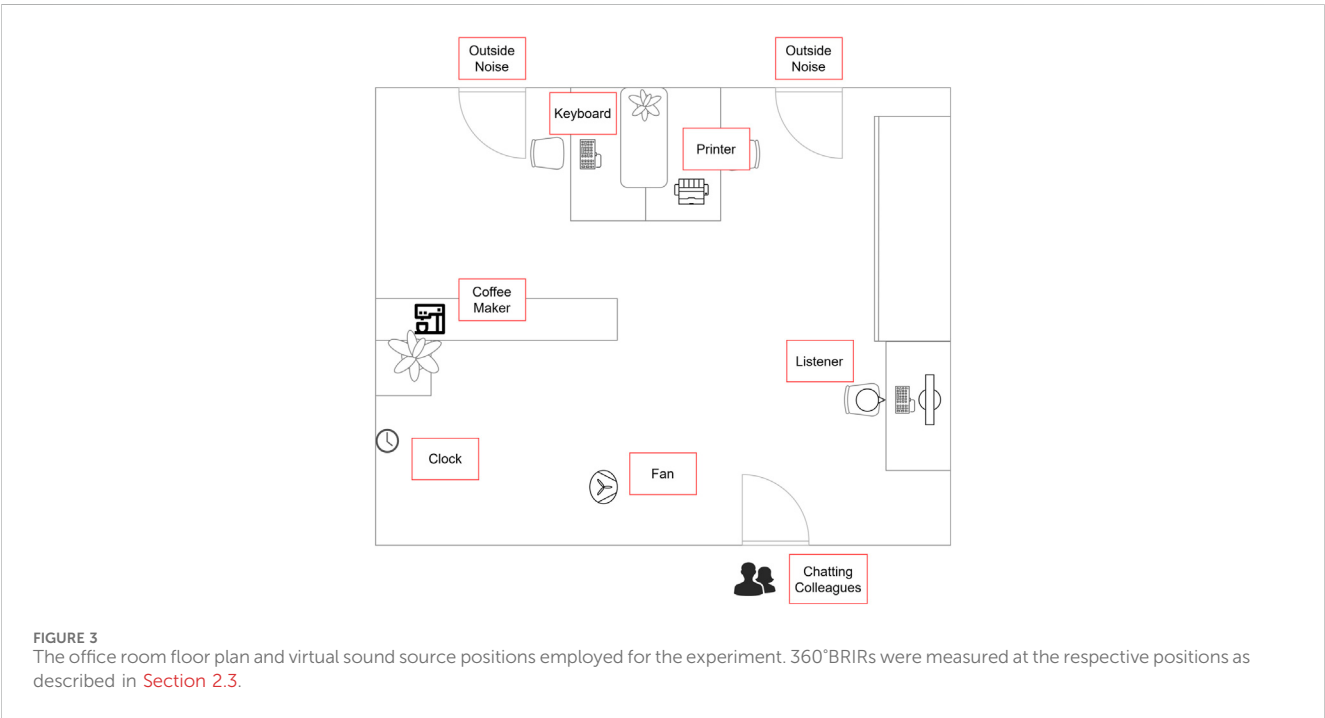
The babble speech had a broadband spectral distribution between 500 Hz and 5 kHz and showed moderate temporal fluctuation rates typical of office dialogue. The printer and keyboard exhibited transient-rich temporal characteristics, while the fan and the outdoor noise were more stationary with low modulation depth.

The individual sounds were combined to create a plausible 40-s auditory scene while preserving spectral and temporal diversity (see Table 1). Each stimulus included street noise as a constant background noise, while other sources were varied to manipulate acoustic complexity and mental effort (Pichora-Fuller et al., 2016). Sources were divided in different categories like mechanical devices (printer, fan), human-related sources (babble noise, coffee maker, keyboard), and background noise (outdoor noise, clock), and combined to reflect the complexity of a typical office environment.

For dynamic (in real-time updated binaural representation based on listeners head movement) binaural synthesis (Vorländer, 2020), a set of Binaural Room Impulse Responses (BRIRs) was recorded using a Genelec 8020D loudspeaker placed at the sound source position of each stimulus (as depicted in Figure 3) and a Neumann KU100 dummy head mounted on the VariSpear device (Bernschütz et al., 2010) placed at the listener's position. The VariSpear is a fully automated device that can horizontally turn the dummy head in steps of 1° horizontal direction. It sequentially measures the binaural room impulse response for each of these directions. Exponential sine sweeps with a length of 2^{19} samples at a sampling rate of 48 kHz were used for all measurements played back from Genelec 8020. The deviating loudspeaker directivity compared to natural objects was neglected for this study. Moreover, according to Blau et al. (2021),

TABLE 1 Overview of the auditory scenes used in the experiment. Each column indicates whether a specific source was present in a given scene. The rightmost column shows the combined mean dB_{Aeq} level for each scene.

Scene	Clock	Coffee maker	Babble noise	Keyboard	Printer	Fan	Outdoor noise	Mean dB_{Aeq}
1	—	—	—	✓	✓	—	✓	50.0
2	—	✓	—	—	✓	—	✓	52.1
3	—	—	—	✓	—	✓	✓	54.3
4	✓	—	✓	✓	✓	—	✓	55.8
5	—	—	—	✓	—	✓	✓	57.9
6	✓	✓	—	—	—	✓	✓	60.1
7	✓	✓	—	—	✓	—	✓	62.9
8	✓	✓	✓	✓	✓	✓	✓	65.2



and [Bau et al. \(2024\)](#) generic BRIRs are sufficient for binaural presentation in dynamic virtual acoustic environments. Thus, no further individualization was applied.

The unprocessed acoustic stimuli were convolved with the measured BRIR sets using the SoundScape Renderer (SSR) in Binaural Room Synthesis (BRS) mode ([Ahrens et al., 2008](#)). Using a Polhemus Fastrak ([PohlemusFastrak, 2022](#)) dynamic headtracking was enabled on the azimuth plane (latency below 60 ms). Auditory stimuli in Reaper were routed using Jack Audio and were auralized in the SSR. The binaural stimuli were presented to the participants using Sennheiser HD600 headphones (including headphone equalization) and a Babyface Pro interface (both conditions). A binaural presentation via headphones was chosen for all conditions since a representation via adjacent loudspeakers could have hindered auditory immersion due to missing visual congruency. All audio software components were rendered on an iMac 2013 OSX 10.10.5. [Table 1](#) lists the eight presented auditory scenes with

separate binaural stimuli and, respectively, measured scene dB_{Aeq} levels (the equivalent free-field sound level at the listener position).

2.4 Load conditions

Beyond manipulating the visual environment (real vs VR), two load conditions, a high-load and a low-load condition, were implemented to investigate the influence of task-induced attentional demands on sound evaluations compared to focused listening in both environments.

In detail, participants were asked to perform the Stroop Color-Word Interference Test during stimulus presentation in the high-load condition. In contrast, participants in the low-load condition merely listened to the sounds. The Stroop task was chosen since it has been shown to induce a constant high mental effort and can plausibly reflect the effort imposed by everyday work situations



FIGURE 4
Comparison of the experimental display in the visual-virtual (left) and real (right) environment.

(Laeng et al., 2012). Additionally, the Stroop task has already been adapted and evaluated in virtual reality, which makes it a promising and well-suited option (Parsons et al., 2011).

During the task, participants were shown random German color words (“rot”, “grün”, “blau”) from a randomly generated list that were displayed in either the matching or a different font color. The task was to click one of the three always-present buttons corresponding to the word’s font color as quickly as possible. The buttons were placed in a triangle at the bottom of the screen, directly underneath the presented colored word, and showed the words red, green, and blue in black font color (cf. Figure 4). Regardless of the correctness of the selection, the next word appeared after the button was clicked or after 3 seconds passed without a response. Participants were familiarized with the task through a trial run at the beginning of the experiment.

There is some heterogeneity in Stroop test scoring in previous literature (see Scarpina and Tagini (2017), for a review). Here, both the relative number of correct responses and the response times were recorded. While the former is a straightforward measure of participants’ ability to complete the task, the latter indicates how easily they can overcome the interference between the word and its font color (Long and Prat, 2002). The ratio of reaction time in ms to the relative amount of correct Stroop trials were combined in the Inverse Efficiency Score (IES) proposed by Townsend and Ashby (1978) who suggested to determine capacity in cognitive psychology as a combination of accuracy and latency.

The Stroop task was designed using PsychoPy 2022.2.4 (Peirce et al., 2019) and Python 3.8 and presented on the same Desktop PC mentioned above. In the VR condition, the task’s user interface was streamed to Unity using SpoutCapture (SpoutCapture, 2021) and presented on a virtual computer screen directly in front of the participants as shown in Figure 4.

2.5 Procedure

The experiment started with a brief disclosure of data privacy and consent to participate in the study. Participants were randomly

assigned to an environmental condition (real vs VR) and then filled out the questionnaire while already present in the office space. Further, participants in both environments were divided into two conditions (high vs low-load), resulting in four experimental conditions (2×2 between-subjects factorial design). Regarding the next step, the procedure differed depending on the experimental environment. In the real environment, participants were seated in front of a Desktop PC. In the VR counterpart, participants sat at the same spot but put on the HTC Vive HMD. Figure 4 shows the comparison between the real desktop PC (right) and its recreation in VR (left). Subsequently, the participants received written (monitor screen) and verbal instructions on how to perform the task. From this point on, participants in the VR conditions remained in the virtual environment until the experiment was ultimately finished. All instructions were presented on a virtual monitor screen, including ratings and questionnaires.

Once participants successfully passed a Stroop Task trial run, the main part of the experiment started. In all conditions, participants were presented one of the acoustical stimuli listed in Table 1 for 40 s. Additionally, participants in the high-load condition performed the Stroop task simultaneously. Retrospectively, participants had to rate the perceived unpleasantness and loudness of the presented soundscape on a continuous five-point Likert scale according to ISO 12913-2 (International Organization for Standardization, 2018). In addition, participants were asked to recall any perceived sound (source) using a randomly generated list of twelve words, including all presented sounds as well as at least 33% distractors. Especially in the high-load task, participants were explicitly instructed to devote all their attention to completing the Stroop test and not to the sounds presented. In the low-load condition, they were instructed to relax and observe the audiovisual virtual environment and not the acoustical scene, depending on the experimental condition.

After each trial, the score achieved for the Stroop task was displayed in the high-load condition before starting the subsequent trial. In total, all eight acoustical stimuli were presented twice to obtain the test-retest reliability, resulting in 16 randomly ordered trials per participant. As described above, the experienced mental effort was

measured using the NASA-TLX questionnaire at the end of the experiment. The experiment took approximately 15–20 minutes per participant.

2.6 Statistical analysis

First, the potential effects of the independent variables *load* and *environment* on the dependent variable subjective effort, as measured by the NASA TLX, were tested. The Stroop test's Inverse Efficiency Scores were calculated as described above (note that low IES values indicate good task performance). Finally, the impact of the independent variables, load, and environment, on the dependent variables unpleasantness, loudness, recall, and test-retest reliability, was measured.

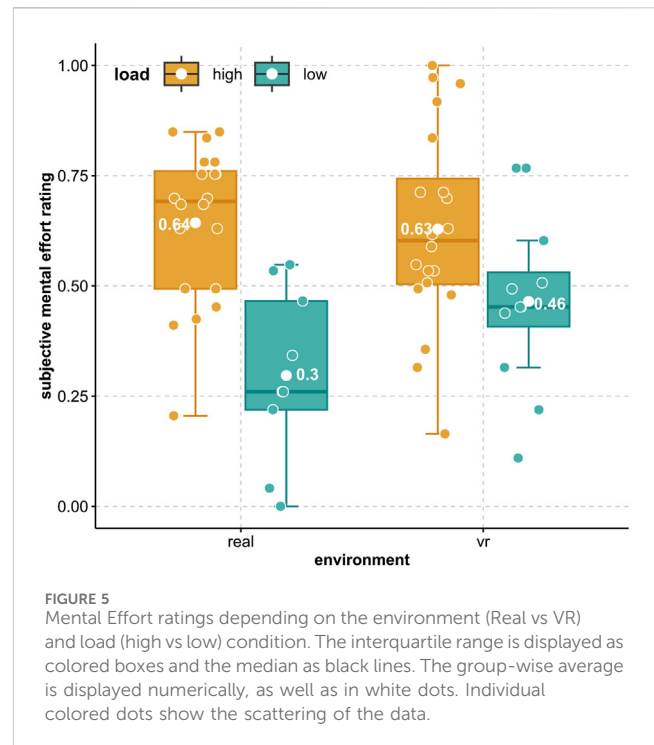
Most analyses were based on Bayesian inference. Compared to frequentist null-hypothesis significance testing (NHST), Bayesian methods and Markov chain Monte Carlo (MCMC) simulations provide credibility intervals instead of point estimates, replace binary decisions for significance with comparing the probabilities of the specified model to a null model by means of the Bayes factor, and are better equipped to handle small sample sizes—at least if appropriate prior distributions are specified (McNeish, 2016). For the dependent variable, subjective effort, where there was only one observation per participant, a Bayesian analysis of variance (ANOVA) as described in Rouder et al. (2012) was employed, which is a linear regression with categorical predictors. For dependent variables with more than one observation per participant, hierarchical linear models with an additional random intercept for participants were chosen. Weakly informative priors were applied to regularize estimates without imposing strong assumptions. Standard normal distributions were selected as priors for the dummy-coded fixed effects, and gamma distributions were chosen for the standard deviation of the random intercepts (for participant and played sound) and the residual error. For the Bayesian ANOVA, Cauchy-priors were supplied for the dummy-coded fixed effects following Rouder et al. (2012), but a weakly informative gamma distribution was chosen instead of the proposed uninformative Jeffrey's prior. Calculations were performed in Posit team (2025) using the 'brms' package (Bürkner, 2021) for Bayesian modeling. Unless noted otherwise, the models were run with four Monte-Carlo chains with 10,000 iterations including 8000 burn-in samples each. Testing for possible prior-data conflicts was realised using the 'priorsense' package (Kallioinen et al., 2024). Lastly, additional causal mediation analyses were performed using the 'mediation' package (Tingley et al., 2014). These analyses are based on quasi-Bayesian Monte Carlo and provide both credibility intervals and well as p-values for NHST. Graphical plots were created using 'ggplot2' and 'ggpubr' (Wickham, 2016).

3 Results

3.1 Mental effort

3.1.1 Environment and load conditions

In the first step, differences in subjective mental effort ratings between conditions were analyzed. Figure 5 depicts the mean ratings



of subjective mental effort across both environment and load conditions.

The Bayesian ANOVA indicated a clear effect of load on subjective mental effort, with higher ratings in the high-load condition ($\beta = 1.068$, 95% CI: [0.595, 1.541]). However, no evidence for a general effect of environment ($\beta = 0.302$, 95% CI: [-0.167, 0.759], with the real environment as reference condition) nor for an interaction between environment and load ($\beta = -0.496$, 95% CI [-1.137, 0.1445]) was supported by the model. A Bayes factor of 68.875 (meaning that the model is nearly 69 times more probable for the observed data than the null model) favored this model over the null model. All $\hat{R}$ values were equal to 1.00, indicating good convergence. The prior and likelihood sensitivity analysis indicated no prior-data conflicts.

3.1.2 Task performance

In the next step, the effect of the environmental condition on performance (IES) in the Stroop task (high-load conditions) was tested. In the real environment condition, the mean of this inverse performance score was at $IES_{\text{Real}} = 1032\text{ms}$, whereas in the VR condition, the mean score was at $IES_{\text{VR}} = 850\text{ms}$. Performance in both conditions showed no considerable change over trial and time.

A hierarchical model was fitted to predict IES from the environment as fixed and participant as random effect. This model drew on the data of the high load conditions only, and the default setting with 10,000 had resulted in a low effective sample size (ESS). Thus, the number of iterations was increased to 12,000 (with 8000 burn-in samples), resulting in all MCMC chains showing good convergence ($\hat{R} = 1.00$). The model showed an effect of environment on performance, with IES scores being lower (i.e., performance being higher) in the VR condition ($\beta = -0.885$, 95% CI: [-1.349, -0.410]). The effect was further supported by a Bayes factor of 4.669, comparing this model against a null model

containing only the random intercept for each participant. Again, prior and likelihood sensitivity analysis indicated no prior-data conflicts.

3.2 Unpleasantness and loudness evaluations

In the next step, the effect of environment and load conditions on differences in unpleasantness and loudness evaluations were assessed. Results are illustrated by Figure 6 for averaged unpleasantness ratings and Figure 7 for averaged loudness ratings. Slight relative increases in unpleasantness and loudness judgments in all conditions are depicted for sounds containing speech signals (Sound 4 [56 dB] and 8 [65 dB], see Table 1).

Two separate hierarchical models were calculated for unpleasantness and loudness judgments, respectively, with environment and load as fixed effects and participant and sound as random effects. The loudness model revealed no clear positive or negative effects of either environment ($\beta = 0.250$, 95% CI $[-0.327, 0.849]$), load ($\beta = 0.281$, 95% CI $[-0.273, -0.836]$) condition, or their interaction ($\beta = -0.09$, 95% CI $[-0.82, 0.60]$); unpleasantness: $\beta = -0.211$, 95% CI $[-0.933, 0.525]$; also, the Bayes factor of 0.065 clearly favored the null model. Similarly, the unpleasantness judgments indicated no effects of environment, load condition, or interaction. Again, the Bayes factor of 0.031 suggested retaining the null model. For both the loudness and the unpleasantness model (including the fixed effects predictors), prior and likelihood sensitivity analysis indicated no prior-data conflicts or too strong priors.

Although unpleasantness and loudness judgment showed no significant difference for either environment or load conditions, some effects were expected due to subjective participants' feedback experiencing high load and unpleasantness in the VR environment. Thus, a mediation analysis similar to (von Berg et al. (2024)) was conducted. Mediation analysis is a statistical approach used to investigate whether the effect of an independent variable on a dependent variable is realised (i.e., mediated) by a third intervening variable. It allows for quantifying the direct and indirect effects of underlying relationships between variables. Figure 8 shows the employed mediation analysis with the independent variables load and environment on the dependent variable mediated by mental effort. The applied quasi-Bayesian Monte Carlo method (Tingley et al., 2014) indeed revealed statistically significant mediations of load condition on loudness and unpleasantness through subjective effort (loudness: $\beta = 0.371$, 95% CI $[0.147, 0.660]$, $p < .001$; unpleasantness: $\beta = 0.318$, 95% CI $[0.137, 0.545]$, $p < .001$). As in the previous hierarchical models on single trial ratings, the direct effects of load of loudness and unpleasantness were ambiguous in magnitude and non-significant (loudness: $\beta = -0.208$, 95% CI $[-0.576, 0.153]$, $p = 0.264$; unpleasantness: $\beta = -0.352$, 95% CI $[-0.723, 0.024]$, $p = 0.069$). However, for unpleasantness, both the credibility interval and the p-value imply a tendency of load condition on mean unpleasantness per participant with unpleasantness being lower in the high load condition. Regarding the environment, no meaningful mediation effects of subjective effort were found for either loudness ($\beta = 0.041$, 95% CI $[-0.117, 0.217]$, $p = 0.612$) or unpleasantness ($\beta = 0.029$,

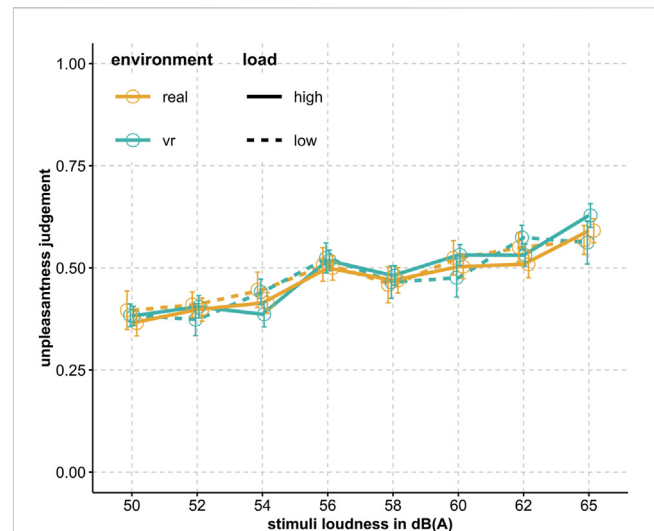


FIGURE 6
Averaged normalized unpleasantness judgments per auditory stimuli in response to different environment and load conditions.

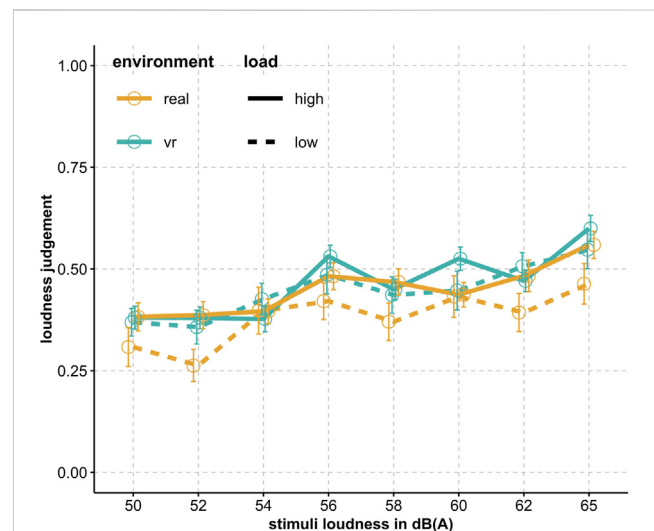
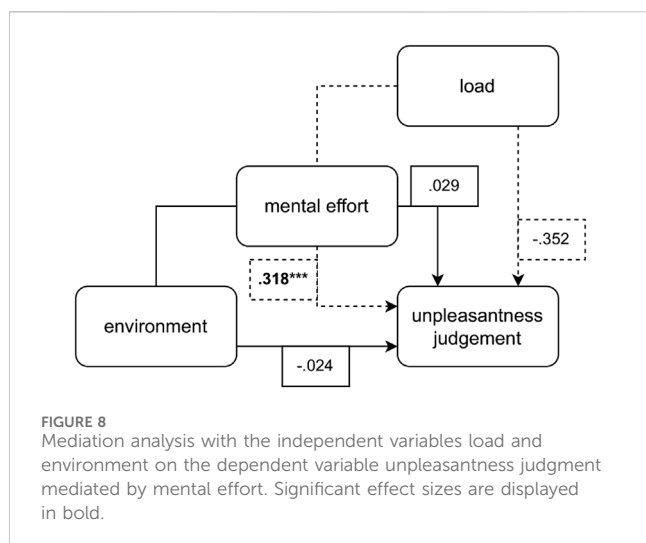


FIGURE 7
Averaged normalized loudness judgments per auditory stimuli in response to different environment and load conditions.

95% CI $[-0.092, 0.159]$, $p = 0.612$), and there were also no significant direct effects in both cases (loudness: $\beta = 0.072$, 95% CI $[-0.282, 0.424]$, $p = 0.693$; unpleasantness: $\beta = -0.024$, 95% CI $[-0.357, 0.323]$, $p = 0.885$).

Furthermore, the effect of the individual sounds on unpleasantness ratings and loudness judgment was assessed. As Figures 6, 7 show, loudness and unpleasantness judgments tend to increase with the overall sound level. Also, all ratings across all experimental conditions increase for scene 4 (level: 56 dBA) which was one of the two scenes that contained speech-alike babble noise (see Table 1). Thus hierarchical models were tested to predict loudness and unpleasantness judgments from the overall sound level as well as the presence or absence of babble noise as fixed effects and the



participants as random effect. The models confirmed positive effects of similar magnitudes for the sound level and the presence of speech on both loudness (level: $\beta = 0.209$, CI = [0.164, 0.254], babble noise: $\beta = 0.312$, CI = [0.211, 0.412]) and unpleasantness ratings (level: $\beta = 0.295$, CI = [0.249, 0.341] babble noise: $\beta = 0.255$, CI = [0.148, 0.361]). The Bayes factors (comparing the respective models against a null model to containing only random effects for the participant and the sound) were at 3852.70 for the loudness model and at 14,829.60 for unpleasantness, providing further evidence for strong effects. The prior and likelihood sensitivity analysis did not indicate any prior-data conflicts or weak likelihoods.

3.3 Recall of perceived sounds

In the next step, we tested whether the two experimental conditions would affect the recall of single sounds in the presented acoustic environments. Similar to previous analyses, a hierarchical model including the load condition and environment as fixed effects and the participant and sound as random effects on the number of correctly remembered sounds was sampled. The model showed an effect towards worse sound recall in the high load condition ($\beta = -0.370$, 95% CI [-0.722, -0.020]), but no evidence for an effect of environment ($\beta = -0.163$, 95% CI [-0.548, 0.229]) nor for an interaction between environment and load ($\beta = -0.05$, 95% CI [-0.531, 0.434]). The Bayes factor comparing this model to a null model was at 1.127, indicating only small improvement in model fit by the inclusion of the fixed effect predictors. Model convergence was good ($\hat{R} = 1.00$), and sensitivity analysis indicated no prior-data conflicts.

3.4 Presence in VR

The subjectively reported presence acquired via the SUS questionnaire implied moderate to high levels of immersion in the VR environment (mean: 0.618, SD: 0.185, on a scale from 0 to 1). Concerning the load conditions, presence ratings tended to be marginally higher in the low load condition (mean: 0.653, SD:

0.083) than in the high load condition (mean: 0.598, SD: 0.225). A Bayesian ANOVA testing for an effect of load condition on the SUS score, however, showed no significant effect ($\beta = -0.189$, 95% CI [-0.689, 0.309]) and the Bayes factor of 0.392 confirmed that this model was clearly less likely than the null model.

3.5 Test-retest reliability

The two ratings of each of the eight presented sounds per participant were used to assess test-retest reliabilities of the loudness and unpleasantness ratings. For the entire sample, test-retest reliabilities were at moderate levels of 0.591 for loudness and 0.540 for unpleasantness. Test-retest reliabilities in the real environment were slightly higher (loudness: 0.606, unpleasantness: 0.557) than in VR (loudness: 0.576, unpleasantness: 0.526). By contrast, there were only marginal differences between the high load condition (loudness: 0.584 unpleasantness: 0.552) and the low load condition (loudness: 0.618, unpleasantness: 0.533).

4 Discussion

The present study examined the influence of different visual environments (VR vs Real) and imposed load (high vs low) on perceived mental effort and the perception of the acoustic environment in an office space, thereby aiming to gain insights on the influences of complex multimodal experimental settings in general.

4.1 Mental effort

First, differences in subjective mental effort ratings dependent on the experimental conditions, were assessed. The results showed significant differences for the load, but not for the environmental condition. This was not expected, as behavioral and physiological results from previous research provide evidence for increased mental effort when performing tasks in virtual environments (Frederiksen et al., 2020; Juliano et al., 2022; Makransky et al., 2019). However, especially in the high load condition, such effects seem absent in this study's effort ratings (see Figure 5). Several reasons might explain these differences to previous research that are rooted in the employed task and the fact that this study collected subjective ratings instead of physiological measurements. Generally, task performance in both high-load conditions was relatively high, indicating that the administered Stroop task may not have led to the expected mental overload. Furthermore, since the Stroop task is related to processing speed and interference control (Periáñez et al., 2020), possible visual complexity in the virtual environment could have inhibited the immediate lexical interpretation of the displayed words and therefore required less interference control, resulting in an easier task in VR. The differences in the IES further support this assumption. In the virtual environment, the IES was lower, although participants kept a similar error rate, indicating that smaller mental demands were counterbalanced with higher processing speed. This finding may indicate that, overall, participants were equally committed to achieving a certain task performance in both

conditions, given the chance to win a voucher. Furthermore, effort cannot just be regarded as a load imposed by a task but is also affected by actively deploying resources towards a task (Westbrook and Braver, 2015), which is affected by previous experiences and skills (Maneuvrier et al., 2023), the interest in the task at hand (Horrey et al., 2017) and the prospect of reward (Fisher et al., 2019). Especially the latter might have mitigated similar investment of effort in both high-load conditions, even though task difficulty might have been different.

By contrast, if only considering the low-load condition, a tendency towards higher subjective mental effort in the virtual environment was observed. This finding would support the assumption of an increased perceptual load due to higher visual abstraction levels (e.g., to correct for mismatches of visual representations and compensate for the absence of the own body), leading to increased working memory demands in VR (Makransky et al., 2019). This trend seemed to vanish when participants completed the Stroop test. Similar effects were observed for a dual motor-cognitive task, where the visualization mode affected neither the performance in the cognitive task (Wenk et al., 2019), nor the subjective rating of overall cognitive load (Wenk et al., 2023). A possible explanation might be the nature of perceptual and cognitive load. Employing resources to meet the VR environment's increased perceptual load is indispensable for processing the outside world and mostly beyond conscious control (Fisher et al., 2019) - contrary to executing mental operations needed to meet the Stroop test's cognitive load, which is a deliberate action (Westbrook and Braver, 2015). Therefore, when asked to rate the overall mental effort, participants might give more weight to the cognitive load component, which was more similar in both environments and, as mentioned above, was even lower in VR. This notion would imply that, in the subjective effort ratings, VR's increased perceptual load may be neglected in the presence of high cognitive load.

4.2 Sound evaluation

Furthermore, the subjective evaluation of the acoustic (office) environment, depending on the experimental factors, was analysed. Here, the analysis indicated no effect of the environment on neither the ratings of loudness and unpleasantness - which involve monitoring and recalling the overall perceptual impression and presumably involve little cognitive load - nor the cognitively more demanding task to identify and remember distinct sound sources. Considering that the VR environment is assumed to predominantly impact the visual perceptual load, it is consistent with the assumption of perceptual load drawing on modality-specific resources (Fisher et al., 2019) that there were apparent effects on the perception and recall of auditory information. By contrast, completing the Stroop test which - albeit an overall manageable task - was shown to impose cognitive load seemed to interfere with the sound evaluation. In detail, the ability to recall the presented sound slightly decreased in high-load conditions. The recall presumably relies more heavily on the higher processing stages (source recognition and memorizing), that are expected to share common resources (Fisher et al., 2019) with completing the Stroop test. Conclusively, performance in these tasks is highly likely to be

prone to interference. Thus, a less accurate sound recall in the dual task condition compared to a listening-only condition could be attributed to either participants failing to execute the primary task (Stroop test) and secondary sound source monitoring task in parallel, resulting in sounds remaining unnoticed (Molloy et al., 2015) or at least unidentified. Also, it is possible that participants deliberately prioritized the Stroop test over paying attention to the sound sources as they would otherwise expect a performance decrease. Second, as the mediation effect of subjective effort showed, only participants who reported to invest much effort in the Stroop task performed higher loudness and unpleasantness ratings, which again implies task interference - at least among those who perceived the Stroop test as demanding. However, these results contradict previous studies which found a significant negative influence of performing the Stroop test on unpleasantness perception (Steffens et al., 2020; Steffens and Himmelein, 2022). These discrepancies might be explained by the overall higher salience of the presented office stimuli with distinct source positions in this study - compared to pink noise used in Steffens and Himmelein (2022) and the more complex, blended stereo sounds used in Steffens et al. (2020). This saliency could have triggered disturbing attentional shifts toward the sounds among those participants who rated completing the task as more effortful than others. Moreover, another study performed on this experiment's data showed that individual factors such as subjectively assessed noise sensitivity do affect unpleasantness and loudness judgment as well as the recall task performance (von Berg et al., 2024).

4.3 Limitations

In the course of this discussion, several limitations have to be addressed, in particular, those associated with the experimental design itself. Firstly, one limitation is the measurement of the subjective mental effort. Since the NASA-TLX only assesses the subjective effort retrospectively and does not discriminate between cognitive and perceptual load, no differentiated statements can be made about this. Although some objective measurements for mental effort like pupillometry (Mathôt, 2018) exist, these are only suitable to a limited extent, especially in VR. Another factor needing additional research might be the incorporation of speech signals. As shown by Kidd et al. (2017) the inclusion of speech stimuli, especially those with linguistic content or few talkers, is likely to increase informational masking due to attentional diversion, linguistic processing, and source confusion. This should be further addressed by further research. Moreover, the length of the individual trials is a limiting factor. Results are likely to differ from actual office scenarios since mental effort might be decreased over time based on listeners adaption to the task (Pichora-Fuller et al., 2016). Additionally, over longer time periods the effort-reward ration might shift leading to a decrease in motivation to expend mental effort on the task (Pichora-Fuller et al., 2016). Thus additional research is needed to confirm the assumption over longer time periods.

Finally, additional limitations lie in the context of ecological validity itself. Overall, ecological validity is controversially discussed in literature. Several studies rely on different and sometimes unclear definitions of ecological validity, thus reducing comparability

(Hohmann et al., 2020). Moreover, since ecological validity is a rather vague definition for a complex phenomenon, the lack of a quantitative measurement method complicates comparability as well (Keidser et al., 2020). Additionally, measurement methods per se might get in the way of the concept of ecological validity. Participants who are aware that their behavior is being measured might involuntarily alter their behavior compromising ecological validity (Keidser et al., 2020). Therefore, researchers are advised to always carefully analyses the demands regarding ecological validity and widen their results with *in-situ* data as well.

Nonetheless, the present study highlights the complex relationship between subjective mental effort, environmental modality (VR vs real), and auditory perception in multimodal settings. These results have important implications for ecological validity and suggest that under increased task load and in VR environments, basic auditory judgments such as loudness and unpleasantness only increase among those participants who report particularly high mental effort, whereas higher-order processes like sound recall are generally more susceptible to interferences. Future psychoacoustic experiments in multimodal (VR) environments should thus carefully control for mental effort and account for modality-specific resource allocation. Here, objective effort measures and stimulus control will be crucial for isolating perceptual phenomena from higher-order cognitive interferences.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Commission on Responsibility in Science Ethics Committee at TH Cologne. 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. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

HH: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources,

Software, Validation, Visualization, Writing – original draft, Writing – review and editing, Funding acquisition. MvB: Conceptualization, Data curation, Formal Analysis, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing. CP: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing. JS: Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The research was conducted as part of the project “Tieffrequente Immissionen im Freizeitlärm” (13FH547KA0), funded by the German Federal Ministry of Education and Research.

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.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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.

References

- Ahrens, J., Geier, M., and Spors, S. (2008). *The soundscape renderer: a unified spatial audio reproduction framework for arbitrary rendering methods*. Amsterdam, Netherlands: Journal of the Audio Engineering Society.
- Aletta, F., and Xiao, J. (2018). *What are the current priorities and challenges for (urban) soundscape research?* Challenges, 9, 16. doi:10.3390/challe9010016
- Alvarez, R. P., Biggs, A., Chen, G., Pine, D. S., and Grillon, C. (2008). Contextual fear conditioning in humans: cortical-hippocampal and amygdala contributions. *J. Neurosci. official J. Soc. Neurosci.* 28, 6211–6219. doi:10.1523/jneurosci.1246-08.2008
- Armougum, A., Orriols, E., Gaston-Bellegarde, A., Marle, C. J. L., and Piolino, P. (2019). Virtual reality: a new method to investigate cognitive load during navigation. *J. Environ. Psychol.* 65, 101338–15. doi:10.1016/j.jenvp.2019.101338
- Augereau, O., Brocheton, G., and Do Prado Neto, P. P. (2022). “An open platform for research about cognitive load in virtual reality,” in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (IEEE), 54–55. doi:10.1109/VRW55335.2022.00020
- Bau, D., Himmelein, H., Lambert, J., Lübeck, T., and Pörschmann, C. (2024). “Investigating the plausibility of spatially interpolated binaural room impulse

responses,” in AES 5th international conference on audio for virtual and augmented. (Redmont, United States: Audio Engineering Society).

Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2010). *Entwurf und Aufbau eines variablen spärlichen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio*. 36th DAGA, 717–718.

Bishop, I., and Rohrmann, B. (2003). Subjective responses to simulated and real environments: a comparison. *Landsc. Urban Plan.* 65, 261–277. doi:10.1016/s0169-2046(03)00070-7

Blau, M., Budnik, A., Fallahi, M., Steffens, H., Ewert, S. D., and van de Par, S. (2021). Toward realistic binaural auralizations – perceptual comparison between measurement and simulation-based auralizations and the real room for a classroom scenario. *Acta Acust.* 5, 8. doi:10.1051/aacus/2020034

Blender. Blender.org free and open 3d creation software (2022).

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., et al. (2014). Mechanisms of motivation-cognition interaction: challenges and opportunities. *Cognitive, Affect. & Behav. Neurosci.* 14, 443–472. doi:10.3758/s13415-014-0300-0

Bürkner, P. C. (2021). Bayesian item response modeling in R with brms and stan. *J. Stat. Softw.* 100, 1–54. doi:10.18637/jss.v100.i05

Bustamante, E. A., and Spain, R. D. (2008). Measurement invariance of the nasa tlx. *Proc. Hum. Factors Ergonomics Soc. Annu. Meet.* 52, 1522–1526. doi:10.1177/154193120805201946

Chamilothori, K., Wienold, J., and Andersen, M. (2019). Adequacy of immersive virtual reality for the perception of daylight spaces: comparison of real and virtual environments. *LEUKOS* 15, 203–226. doi:10.1080/15502724.2017.1404918

da Silva Soares, R. Jr., Ramirez-Chavez, K. L., Tufanoglu, A., Barreto, C., Sato, J. R., and Ayaz, H. (2024). Cognitive effort during visuospatial problem solving in physical real world, on computer screen, and in virtual reality. *Sensors Basel, Switz.* 24, 977. doi:10.3390/s24030977

de Kort, Y. A. W., Ijsselstein, W. A., Kooijman, J., and Schuurmans, Y. (2003). Virtual laboratories: comparability of real and virtual environments for environmental psychology. *Presence Teleoperators Virtual Environ.* 12, 360–373. doi:10.1162/105474603322391604

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., and Tremblay, S. (2013). Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. *Hum. factors* 56, 631–644. doi:10.1177/0018720813510735

Ferwerda, J. A. (2003). “Three varieties of realism in computer graphics,” in Proceedings of Human Vision and Electronic Imaging VIII, Santa Clara, California, USA, January 21, 2003, 5007 290–297. doi:10.1117/12.473899

Fisher, J. T., Hopp, F. R., and Weber, R. (2019a). Modality-specific effects of perceptual load in multimedia processing. *Media Commun.* 7, 149–165. doi:10.17645/mac.v7i4.2388

Fisher, J. T., Hopp, F. R., and Weber, R. (2019b). “Disentangling the roles of reward and process type on resource allocation in an interactive task,” in 69th Annual Conference of the International Communication Association.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., et al. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surg. Endosc.* 34, 1244–1252. doi:10.1007/s00464-019-06887-8

Fuchs, P., Moreau, G., and Guitton, P. (2011). *Virtual reality: concepts and technologies. A balkema book*. Boca Raton, FL: CRC Press. doi:10.1201/b11612

Gagné, J. P., Besser, J., and Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: a review. *Trends Hear.* 21, 2331216516687287. doi:10.1177/2331216516687287

Hart, S. G., and Staveland, L. E. (1988). “Development of nasa-tlx (Task load index): results of empirical and theoretical research,” in *Human mental workload (North-Holland), advances in psychology*. Editor P. A. Hancock, 52, 139–183. doi:10.1016/s0166-4115(08)62386-9

Hermida Cadena, L. F., Lobo Soares, A. C., Pavón, I., and Coelho, L. B. (2017). Assessing soundscape: Comparison between *in situ* and laboratory methodologies. *Noise Mapp.* 4, 57–66. doi:10.1515/noise-2017-0004

Higuera-Trujillo, J. L., López-Tarruella Maldonado, J., and Llinas Millán, C. (2017). Psychological and physiological human responses to simulated and real environments: a comparison between photographs, 360° panoramas, and virtual reality. *Appl. Ergon.* 65, 398–409. doi:10.1016/j.apergo.2017.05.006

Hochschule Luzern -Technik und Architektur (2010). SBIB-studie schweizerische befragung in büros.

Hohmann, V., Paluch, R., Krueger, M., Meis, M., and Grimm, G. (2020). The virtual reality lab: realization and application of virtual sound environments. *Ear Hear.* 41 (Suppl. 1), 31S–38S. doi:10.1097/aud.0000000000000945

Horrey, W. J., Lesch, M. F., Garabet, A., Simmons, L., and Maikala, R. (2017). Distraction and task engagement: how interesting and boring information impact driving performance and subjective and physiological responses. *Appl. Ergon.* 58, 342–348. doi:10.1016/j.apergo.2016.07.011

International Organization for Standardization (2018). Acoustics - Soundscape - Definition and conceptual framework (ISO Standard No. 12913-2:2018). Available online at: <https://www.iso.org/standard/75267.html>.

Juliano, J. M., Schweighofer, N., and Liew, S. L. (2022). Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer. *J. neuroengineering rehabilitation* 19, 106. doi:10.1186/s12984-022-01084-6

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice-Hall Inc.

Kallioinen, N., Paananen, T., Bürkner, P., and Vehtari, A. (2024). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *Statistics Comput.* 34 1–27. doi:10.1007/s11222-023-10366-5

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear Hear.* 41 (Suppl. 1), 5S–19S. doi:10.1097/aud.0000000000000944

Kidd, G., and Colburn, H. S. (2017). “Informational masking in speech recognition,” in *The auditory system at the cocktail party*. Editors J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Cham: Springer International Publishing), 75–109. doi:10.1007/978-3-319-51662-2_4

Kim, J., and de Dear, R. (2013). Workspace satisfaction: the privacy-communication trade-off in open-plan offices. *J. Environ. Psychol.* 36, 18–26. doi:10.1016/j.jenvp.2013.06.007

Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. *Cyberpsychology & Behav. impact Internet, multimedia virtual Real. Behav. Soc.* 8, 76–88. doi:10.1089/cpb.2005.8.76

Kool, W., and Botvinick, M. (2018). Mental labour. *Nat. Hum. Behav.* 2, 899–908. doi:10.1038/s41562-018-0401-9

Kreitzberg, K. (2021). “Technische Regeln für Arbeitsstätten,” in *Betriebliche prävention*.

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci. Assoc. Psychol. Sci.* 7, 18–27. doi:10.1177/1745691611427305

Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. General* 133, 339–354. doi:10.1037/0096-3445.133.3.339

Li, H., and Lau, S. K. (2020). A review of audio-visual interaction on soundscape assessment in urban built environments. *Appl. Acoust.* 166, 107372. doi:10.1016/j.apacoust.2020.107372

Long, D. L., and Prat, C. S. (2002). Working memory and stroop interference: an individual differences investigation. *Mem. & Cognition* 30, 294–301. doi:10.3758/BF03195290

Luong, T., Martin, N., Argelaguet, F., and Lécuyer, A. (2019). “Studying the mental effort in virtual versus real environments,” in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 809–816. doi:10.1109/VR.2019.8798029

Macedonio, M. F., Parsons, T. D., DiGiuseppe, R. A., Weiderhold, B. K., and Rizzo, A. A. (2007). Immersiveness and physiological arousal within panoramic video-based virtual reality. *Cyberpsychology & Behav. Behav. Soc.* 10, 508–515. doi:10.1089/cpb.2007.9997

Makransky, G., Terkildsen, T. S., and Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* 60, 225–236. doi:10.1016/j.learninstruc.2017.12.007

Maneuverier, A., Ceyte, H., Renaud, P., Morello, R., Fleury, P., and Decker, L. M. (2023). Virtual reality and neuropsychological assessment: an analysis of human factors influencing performance and perceived mental effort. *Virtual Real.* 27, 849–861. doi:10.1007/s10055-022-00698-4

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *J. cognition* 1, 16. doi:10.5334/joc.18

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Struct. Equ. Model. A Multidiscip. J.* 23, 750–773. doi:10.1080/10705511.2016.1186549

Meyer, G. F., Wong, L. T., Timson, E., Perfect, P., and White, M. D. (2012). Objective fidelity evaluation in multisensory virtual environments: auditory cue fidelity in flight simulation. *PLoS one* 7, e44381. doi:10.1371/journal.pone.0044381

Molloy, K., Griffiths, T. D., Chait, M., and Lavie, N. (2015). Inattentional deafness: visual load leads to time-specific suppression of auditory evoked responses. *J. Neurosci.* 35, 16046–16054. doi:10.1523/jneurosci.2931-15.2015

Parsons, T. D., Courtney, C. G., Arizmendi, B., and Dawson, M. (2011). Virtual reality stroop task for neurocognitive assessment. *Stud. health Technol. Inf.* 163, 433–439.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). “PsychoPy2: experiments in behavior made easy,” *Behav. Res. Method.* 51, 195–203. doi:10.3758/s13428-018-01193-y

Periáñez, J. A., Lubrini, G., García-Gutiérrez, A., and Ríos-Lago, M. (2020). Construct validity of the stroop color-word test: influence of speed of visual search, verbal fluency, working memory, cognitive flexibility, and conflict monitoring. *Archives Clin. Neuropsychology* 36, 99–111. doi:10.1093/arclin/aaaa034

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi:10.1097/AUD.0000000000000312

- Picou, E. M., and Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear Hear.* 35, 611–622. doi:10.1097/AUD.0000000000000055
- Polhemus (2012). FASTRAK 3D tracking system. Colchester, United States: Polhemus Incorporated.
- Posit team (2025). RStudio: Integrated Development Environment for R. Boston, MA: Posit Software, PBC. Available online at: <http://www.posit.co/>.
- Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default bayes factors for anova designs. *J. Math. Psychol.* 56, 356–374. doi:10.1016/j.jmp.2012.08.001
- Scarpina, F., and Tagini, S. (2017). The stroop color and word test. *Front. Psychol.* 8, 557. doi:10.3389/fpsyg.2017.00557
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence Teleoperators Virtual Environ.* 1, 120–126. doi:10.1162/pres.1992.1.1.120
- Souza, V., Maciel, A., Nedel, L., and Kopper, R. (2022). Measuring presence in virtual environments: a survey. *ACM Comput. Surv.* 54, 1–37. doi:10.1145/3466817
- SpoutCapture (2021). Leadedge spoutcapture. Available online at: <https://github.com/leadedge/SpoutCapture>.
- Steffens, J., and Himmelein, H. (2022). “Induced cognitive load influences unpleasantness judgments of modulated noise,” in Proceedings of the 24 th International Congress on Acoustics.
- Steffens, J., Müller, F., Schulz, M., and Gibson, S. (2020). The effect of inattention and cognitive load on unpleasantness judgments of environmental sounds. *Appl. Acoust.* 164, 107278. doi:10.1016/j.apacoust.2020.107278
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi:10.1037/h0054651
- Tarlao, C., Steele, D., and Guastavino, C. (2022). Assessing the ecological validity of soundscape reproduction in different laboratory settings. *PloS one* 17, e0270401. doi:10.1371/journal.pone.0270401
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). Mediation: r package for causal mediation analysis. *J. Stat. Softw.* 59, 1–38. doi:10.18637/jss.v059.i05
- Townsend, J., and Ashby, F. (1978). Methods of modeling capacity in simple processing systems. *Cogn. Theory* 3, 199–239.
- Unity Technologies (2022). The leading platform for creating interactive, real-time content.
- Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality. *Presence Teleoperators Virtual Environ.* 9, 497–503. doi:10.1162/105474600566989
- VIVE. Discover virtual reality beyond imagination (2022).
- von Berg, M., Himmelein, H., and Steffens, J. (2024). Effects of noise sensitivity and listening effort on perceptual ratings of background noise. *JASA Express Lett.* 4, 084401. doi:10.1121/10.0028201
- Vorländer, M. (2020). “Convolution and binaural sound synthesis,” in auralization. *RWTHedition*. (Cham: Springer). doi:10.1007/978-3-030-51202-6_9
- Wenk, N., Penalver-Andres, J., Palma, R., Buetler, K. A., Müri, R., Nef, T., et al. (2019). “Reaching in several realities: motor and cognitive benefits of different visualization technologies,” in 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), 1037–1042. doi:10.1109/ICORR.2019.8779366
- Wenk, N., Penalver-Andres, J., Buetler, K. A., Nef, T., Müri, R. M., and Marchal-Crespo, L. (2023). Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Real.* 27, 307–331. doi:10.1007/s10055-021-00565-8
- Westbrook, A., and Braver, T. S. (2015). Cognitive effort: a neuroeconomic approach. *Cognitive, Affect. & Behav. Neurosci.* 15, 395–415. doi:10.3758/s13415-015-0334-y
- Wickham, H. (2016). *Ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag New York. doi:10.1007/978-3-319-24277-4
- Xu, C., Oberman, T., Aletta, F., Tong, H., and Kang, J. (2021). Ecological validity of immersive virtual reality (ivr) techniques for the perception of urban sound environments. *Acoustics* 3, 11–24. doi:10.3390/acoustics3010003