

Document Version

Final published version

Licence

CC BY

Citation (APA)

Cucinella, S. L., de Winter, J., van den Berg, A., van den Berg, B., Hoogsteen, D., Poutsma, D., van Wingerden, F., & Marchal-Crespo, L. (2026). Shopping in immersive virtual reality: Effects of visual, auditory, and cognitive demands on mental workload. *Applied Ergonomics*, 137, Article 104781. <https://doi.org/10.1016/j.apergo.2026.104781>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

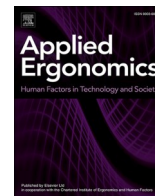
In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Shopping in immersive virtual reality: Effects of visual, auditory, and cognitive demands on mental workload

Salvatore Luca Cucinella^{a,b}, Joost de Winter^{a,*} , Alex van den Berg^a, Bibi van den Berg^a, Douwe Hoogsteen^a, Daan Poutsma^a , Floris van Wingerden^a, Laura Marchal-Crespo^{a,b} 

^a Department of Cognitive Robotics, Delft University of Technology, Delft, the Netherlands

^b Department of Rehabilitation Medicine, Erasmus Medical Center, Rotterdam, the Netherlands

ARTICLE INFO

Keywords:

Immersive virtual reality
Mental workload
Cognitive demands
Sensory demands
Virtual shopping task

ABSTRACT

Introduction: Immersive virtual reality applications are increasingly popular in entertainment, education, and professional training. While many aim for maximal realism, simplifying the virtual environment may offer benefits such as reducing mental workload and improving focus on core tasks. However, the impact of different types of demand on users' mental workload remains unclear.

Objective: This study explored the impact of visual, auditory, and cognitive demands on users' mental workload during a daily living activity in immersive virtual reality.

Methods: Twenty-four participants used a head-mounted display for a virtual shopping task, i.e., picking ten listed products from a shelf, under different conditions: visual demands (moving characters), auditory demands (background noise), cognitive demands (simultaneous arithmetic task), and a combination of all three. Mental workload measures included heart rate, pupil diameter, and self-reported mental demand & effort.

Results: The cognitively demanding secondary task induced the largest mental workload, significantly exceeding that of auditory and visual demands. For example, on a scale of 1 (low) to 10 (high), self-reported mental demand & effort was 4.40 for the moving characters, 5.00 for the background noise, 6.67 for the arithmetic task, and 7.17 for the combined condition. Biosignal differences were consistent within participants but were masked by high inter-individual variability.

Conclusions: In virtual shopping tasks, reducing enforced cognitive demands may be more effective for decreasing mental workload than reducing non-task-relevant visual or auditory demands.

1. Introduction

Head-mounted displays (HMDs) for immersive virtual experiences are popular not only for entertainment and gaming, as reflected in the large number of households that own or have used such a device (ARtillery Intelligence, 2024; Statista, 2023), but also across diverse professional domains, including architectural design (Hou et al., 2024), sports training (Faure et al., 2020; Pastel et al., 2023), surgical training (Mao et al., 2021; Munawar et al., 2024), exposure therapy for phobias (Krijn et al., 2004; Spyridonis et al., 2024), and neurorehabilitation (i.e., therapy aimed at restoring function after brain injury; Langhorne et al., 2011; Palacios-Navarro and Hogan, 2021; Specht et al., 2023).

HMDs can offer advantages over conventional 2D screens for experiencing virtual environments, for example by eliciting greater feelings of enjoyment and presence (Newman et al., 2022). Wenk et al. (2022)

found that participants made straighter, shorter, and smoother movements when their 3D reaching movements were visualized with HMDs compared to a 2D screen, likely due to stereoscopic vision, reduced need for visuo-spatial transformations, and better eye-hand coordination. In the context of rehabilitation, one advantage of immersive virtual reality is its ability to provide simulated environments that closely resemble real-world settings for assessment and intervention, thereby allowing patients to safely practice complex skills without the associated real-world risks (Saldana et al., 2020). Furthermore, the computerized nature of these tasks allows for tracking performance and task progress (De Pasquale et al., 2024; Nath et al., 2022; Stanica et al., 2020). Yet, the discussion on the necessary level of realism for immersive virtual experiences is still open (e.g., Gonçalves et al., 2022). Realism refers to the degree to which a virtual environment represents its real-world counterpart. This includes auditory, visual, and tactile feedback, which can

* Corresponding author.

E-mail address: j.c.f.dewinter@tudelft.nl (J. de Winter).

<https://doi.org/10.1016/j.apergo.2026.104781>

Received 12 February 2025; Received in revised form 6 January 2026; Accepted 29 March 2026

Available online 7 May 2026

0003-6870/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

help simulate key perceptual and environmental features of real-life experiences (Melo et al., 2022).

It is commonly assumed that the more realistic the virtual environment, the greater the transfer of learning to the real world, although evidence suggests that the relationship between fidelity and transfer is more nuanced (Levac et al., 2019). Moreover, previous studies (Cucinella et al., 2025; Elghoul et al., 2022; Guadagnoli and Lee, 2004; Levin and Demers, 2021; Sweller, 2020; Van Merriënboer and Sweller, 2005; Wiskerke et al., 2022) indicate that learners should not be overstimulated, but rather adequately challenged so that they can perform at an optimal level of cognitive demand. In the current study, we use the term “mental workload” to describe the mental cost to the user, a variable we operationalized through self-reports and physiological indices (Hart and Staveland, 1988). In this context, simplifying the environment may prove to be a useful solution to increase training effectiveness (cf. Fisher et al., 2014; Dahlstrom et al., 2009). For example, it is conceivable that when learning to drive, removing visual information (such as trees and buildings) may enable a reduction of mental workload and allow the learner to better focus on the core task of keeping the car on the road, thus learning it more effectively. However, it remains an open question as to how immersive virtual environments should be simplified to reduce the mental workload of participants.

Understanding how different task demands influence mental workload is important for designing virtual reality systems that adjust environmental complexity or task difficulty based on the user's needs. To examine how different types of supplementary demands affect mental workload, participants performed a daily activity task in immersive virtual reality. The selected environment was a digital replica of a supermarket, and the chosen task involved shopping. Shopping in immersive virtual reality has previously been used to assess cognitive functions, with users memorizing items in a shopping list and then finding and collecting them (Ouellet et al., 2018), as well as cognitive and motor functions related to instrumental activities of daily living, with users navigating a virtual grocery store on an omnidirectional treadmill while searching for items on a shopping list (Lewis et al., 2023).

In our study, we chose to manipulate auditory, visual, and cognitive demands, as shopping typically entails all three types of demands. Specifically, stores can be noisy and contain various visual distractions from the primary task, such as other people walking around. Shopping for products can also be cognitively challenging, such as when having to remember product prices, the amount of money spent, or, if not available as a shopping list, which products to purchase.

Wickens' multiple resource theory (2002) offers a framework for predicting how these three forms of demands might affect performance. This theory posits that multitasking becomes challenging when two tasks draw on the same type of resources (e.g., visual resources, verbal resources). While research indicates that background sounds can negatively impact task performance (e.g., Doggett et al., 2021) and that visual complexity can reduce performance in visual tasks (Ragan et al., 2015), humans may attempt to mitigate the effects of such distractors by directing their attention toward the primary task. In other words, visual and auditory demands, such as other people in a shopping environment, have the potential to compete with the primary visual demands of shopping (e.g., finding products on the shelves), but active processing of these distractors is not required, as humans can selectively attend to relevant task cues (see Wickens, 2021, for a review of attentional mechanisms). However, retaining items in working memory is inherently demanding and likely to interfere with the primary central processing resources required for memorizing and identifying products. Thus, it can be expected that non-task-related visual and auditory stimuli will have less impact on task performance, whereas an imposed cognitive secondary task is likely to be more inhibitory.

In this study, various dependent variables known to be associated with mental workload were employed. First, we used physiological measures, including heart rate and pupil diameter. Heart rate is

traditionally recognized as a measure of physical workload (Hoover et al., 2012; Tao et al., 2019) and is likely to be elevated during tasks that require physical movement, such as shopping. In the current study, we investigated whether simplifying the environment by reducing non-physical demands has a detectable effect on heart rate. Previous research has yielded mixed results regarding the impact of cognitive tasks on heart rate. Some studies, such as those examining physiological responses during highway driving and when additional cognitive tasks are introduced, have found that non-physical cognitive tasks did increase heart rate, while other studies have observed related cardiovascular changes such as decreased heart rate variability (Charles and Nixon, 2019; Hoover et al., 2012; Mehler et al., 2012; Novak et al., 2010). In contrast, a study involving a computer-based task that combined mental and physical demands found that mental demand did not significantly affect heart rate variability indices of autonomic activity beyond the effect of physical demand alone (Garde et al., 2002). The current study aims to further understand the relationship between non-physical demands and heart rate, and to provide a clearer understanding of physiological responses in daily living scenarios.

Pupil diameter is a commonly used measure of mental workload. Specifically, pupils dilate during tasks such as performing calculations or retaining items in working memory, and constrict when those items are released (De Winter et al., 2021; Kahneman and Beatty, 1966; Marquart and De Winter, 2015). Pupil diameter also changes in response to visual (Bradley et al., 2008) and auditory (Partala and Surakka, 2003) stimuli of an emotionally arousing nature. However, so far, little is known about the utility of pupil diameter in dynamic tasks like shopping (for exceptions, see Ladeira et al., 2020; and for naturalistic visual search, see Medathati et al., 2020). Using physiological measures such as heart rate and pupil diameter can potentially offer advantages in future applications, especially when participants cannot immediately express their perceived mental workload or in cases where automated adaptation of the virtual world is desired (Kritikos et al., 2021; Luong et al., 2020). In addition to these two mental workload measures, we used two standard measures: self-reported mental demand & effort from the NASA-TLX, and task completion time as a performance measure (Hart and Staveland, 1988; Hertzum, 2022).

In summary, in this experiment, we exposed participants to visual, auditory, and cognitive demands in a virtual shopping task to investigate the effects on task performance, perceived mental demand and effort, and physiological consequences. We expected that the combined visual-auditory-cognitive demand condition would yield the longest task completion times, the highest self-reported mental demand and effort, and the greatest increases in heart rate and pupil dilation compared to the other conditions. Furthermore, as mentioned above, cognitive demands were expected to be more demanding than visual or auditory demands, which can be filtered out by the participant.

2. Materials and methods

2.1. Participants

Participants were recruited from TU Delft via word of mouth and did not receive compensation. All participants provided written informed consent. The study was approved by the Human Research Ethics Committee (HREC) of the TU Delft, approval number 2615.

The sample size was determined heuristically, aiming to achieve complete counterbalancing (all 24 permutations; $4 \times 3 \times 2 \times 1$) of the four experimental conditions. After testing all 24 participants, errors and incomplete data affected the results for three participants (Participants 2, 4, and 5). Specifically, Participant 2's heart rate could not be measured for all trials, likely due to incorrect electrode placement; sound was not activated for Participant 4; and pupil diameter data were missing for one trial of Participant 5. To ensure complete datasets for all 24 participants, these three were replaced with newly recruited participants. The group of 24 participants was composed of university

students aged 20–25 years (12 women, 12 men).

2.2. Experimental setup

The experiment was conducted in a laboratory room with sufficient space for participants to move without risk of colliding with objects. The door was closed and the room lights were on. The room admitted daylight. The experimental setup included an HMD (HTC Vive Pro Eye, HTC Corporation, Taiwan) with a field of view of 110° and dual OLED screens with a resolution of 1440 × 1600 pixels per eye (2880 × 1600 pixels combined), and an HTC Vive tracked controller. The HTC Vive Pro Eye featured built-in eye-tracking technology; pupil diameter data were accessed at a frequency of approximately 90 Hz.

The virtual environment was developed using the Unity game engine (Unity Technologies, USA), version 2021.3.11f1. To connect the HTC Vive with the Unity software, we used the SteamVR plugin (version 2.7.3, Valve Corporation, USA). The Tobii XR SDK and SRanipal SDK packages were downloaded and installed in Unity for collecting the eye-related data (developer.vive.com and hub.vive.com). Shopping products (e.g., onions, loaves of bread, cases of beer, cartons of milk) and character avatars (NPC Populator package) were obtained from the Unity Asset Store. Additionally, two soundtracks were incorporated as auditory stimuli; these were acquired from the Envato Elements store (elements.envato.com). The headphones of the HTC Vive Pro Eye were used to play the auditory stimuli.

A TMSi Porti 7 device (TMSi, the Netherlands) using a Bluetooth connection was used to obtain electrocardiogram (ECG) recordings at a frequency of 1024 Hz. A lead II configuration was used for electrode placement. Data were recorded using MATLAB (version R2021b). The ECG recordings and eye-related recordings were conducted using separate software (MATLAB version R2021b for the ECG data, and Unity for the HTC Vive data). Recordings were approximately synchronized by manually starting them simultaneously.

A computer, running Windows 10 64-bit edition, executed the task within the Unity Editor. The computer was equipped with 32 GB of DIMM DDR4 RAM, an NVIDIA GeForce RTX 3080 GPU (NVIDIA Corporation, USA), and an AMD Ryzen 9 5900X 12-Core Processor at 3.70 GHz (Advanced Micro Devices, USA).

2.3. Virtual grocery shopping task

The task required participants to (1) read ten products from a shopping list located in the virtual environment, (2) locate and pick up each product from the shelves in the order displayed on the list, (3) place the products in a shopping cart situated beneath the shopping list, and, once the last product was collected, (4) say aloud “Stop”, so the researcher could end the trial. Participants could move around a walkable area that included the shelves with the products, the shopping list, and the cart (Fig. 1).

Participants grabbed an object by pressing a controller button and released it by letting go of the same button. The participant's right hand was displayed as a virtual object in the environment (Fig. 2). A marker in the virtual hand turned from red to white to indicate to the participants that they could grab the product, and from white to red once the object was grasped. Fig. 1 depicts the virtual environment from a remote perspective, Fig. 2 depicts the virtual environment from the participant's perspective, while Fig. 3 shows a top view of the virtual environment.

Participants completed the task in four trials. In each trial, ten designated areas on the shelves were used to position products. For each of the four conditions, the product sequence on the shopping list and the product arrangement on the shelves were fixed across participants but differed between conditions. Participants accessed the same shelf areas in each trial, but the order in which they collected the ten products differed between trials, preventing them from memorizing the sequence of products.



Fig. 1. The task environment, from a remote perspective.



Fig. 2. The task environment from the participant's perspective. This is a screenshot from a video recording of the monitor used to display what the participant saw in the HMD. In this screenshot, the participant has just grabbed a loaf of bread (product 2 from the list) and is in the process of placing it into the shopping cart.

2.4. Experimental conditions

The experiment consisted of four different conditions (see Table 1 for an overview).

Condition “Characters”. Multiple human characters (avatars) walked through the aisles and behind the shelves. Twelve characters (6 female, 6 male) were placed in the virtual environment. Characters navigated within six walkable areas delineated by the shelves. The characters did not interfere with the participant's task but were behind the shelves and in the neighboring aisle (see Fig. 3).

Condition “Noise”. This condition used two monophonic audio fragments: one capturing the ambience of a busy supermarket

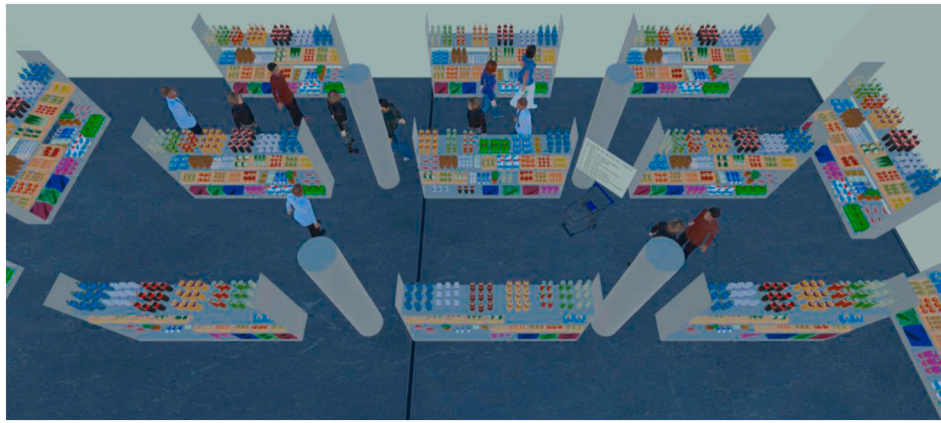


Fig. 3. The task environment from a top view. This screenshot, taken from the condition Characters, shows the twelve human characters who continuously walked along predefined paths in the aisles behind and beside the main shelves. These paths consisted of invisible waypoints, guiding their movement from point to point throughout the experiment.

Table 1
Overview of the four experimental conditions.

	Human characters walking through the aisles	Background sounds of a busy supermarket	Counting backward task
Condition “Characters”	Yes	No	No
Condition “Noise”	No	Yes	No
Condition “Arithmetic”	No	No	Yes
Condition “Combined”	Yes	Yes	Yes

(conversations, door openings, etc.) and one containing sounds of cash registers. The supermarket ambience was presented as non-spatialized audio, without a defined spatial location. The cash register audio was rendered as spatialized audio, positioned behind the participant on the right side at an approximate distance of 4 m. Both fragments were continuously looped through the HMD headphones. The volume was maintained at a consistent, comfortable level for all participants.

Condition “Arithmetic”. Each time participants grabbed a product from the shelves, they were required to mentally count backward from 99, subtract 7 each time, and report the result aloud (for similar procedures, see e.g., Chen et al., 2023; Runswick et al., 2018). Each trial required exactly 10 spoken subtraction responses (one per product grasp).

Serial subtraction creates cognitive demand on working memory and attention (Bristow et al., 2016), as it requires individuals to hold a number in mind, perform repeated mental calculations, and sustain focus over time. Compared with simpler arithmetic tasks (e.g., subtracting by three), subtracting by seven involves more frequent borrowing operations, which are cognitively more demanding. For example, the ten steps starting from 99 are: 92 (basic), 85 (borrowing), 78 (borrowing), 71 (basic), 64 (borrowing), 57 (borrowing), 50 (basic), 43 (borrowing), 36 (borrowing), and 29 (borrowing) (for research on the demands of carry/borrow operations, see Imbo et al., 2007; Raghubar et al., 2010). Although this task does not replicate memorizing a shopping list, it imposes a general working-memory and attention load in a standardized and controlled way across participants. We selected this task to ensure consistency and to avoid the variability introduced by more ecologically valid activities (e.g., making phone calls), which can introduce confounding factors such as timing, language proficiency, or emotional responses.

Condition “Combined”. This condition combined the previous three conditions (Characters, Noise, and Arithmetic). In the arithmetic

Table 2
Characteristics of the Characters and Noise in the Combined condition.

Characters	
Character type	12 characters (6 male, 6 female) dressed casually.
Character actions	Characters walking through the aisles and behind the shelves.
Character location	Characters behind and next to the main shelves, at least 4 m from the participant’s starting point.
Noise	
Audio type	Ambient busy supermarket, including conversations, laughs, crowd talk, checkout beeps, footsteps, plastic bags opening, and door openings.
Audio motion/state	Static (non-moving) audio fragments.
Audio location	Non-spatialized ambient audio and a spatialized sound source (cashier beeps) located approximately 4 m behind the participant on the right side.

task, participants started at 101 and had to subtract 7 each time they took a product. The Combined condition aimed to simulate a simplified version of a real supermarket experience, incorporating typical sounds of a busy supermarket and human-like characters walking (see Table 2).

2.5. Dependent measures

The following dependent measures were calculated per completed trial:

- **Self-reported mental demand & effort (1 to 10).** After each trial, self-reported workload was measured using the NASA Task Load Index (TLX). It consisted of the following six items:
 - Mental demand: How mentally demanding was the task?
 - Physical demand: How physically demanding was the task?
 - Temporal demand: How hurried or rushed was the pace of the task?
 - Performance: How successful were you in accomplishing what you were asked to do?
 - Effort: How hard did you have to work to accomplish your level of performance?
 - Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Participants rated these dimensions on a 10-point scale, ranging from 1 (low) to 10 (high). In the current study, only the mental demand and effort items were evaluated. The reason for this is that these two items

consider the mental or cognitive aspects of the task, which are our main interest.¹ The two items were averaged to reduce random measurement error while preserving their shared variance. The decision to combine mental demand and effort is supported by a meta-analysis indicating that these two items exhibit the strongest correlation among all fifteen item pairs (Hertzum, 2022), as well as by other studies that have similarly aggregated these two items (Benoit et al., 2019; Bertrand et al., 2023; Hess et al., 2016; Johnston et al., 1997; Teodoro et al., 2023). For completeness, we provide the results of all six TLX items separately in the Appendix of this paper.

- *Task completion time (s)*. For each trial, we determined the task completion time, measured from the start of the trial until the tenth item was placed in the cart.
- *Pupil diameter (mm)*. The mean pupil diameter was determined by averaging the recorded pupil diameter for both the left and right eyes and then averaging over the trial duration.
- *Heart rate (bpm)*. A MATLAB implementation of the Pan-Tompkins QRS Detector, as developed by Sedghamiz (2014), was used to extract heartbeats from the ECG signal. The Pan and Tompkins (1985) method involves a sequence of signal processing steps designed to identify the QRS complexes in an ECG signal. Mean heart rate per trial was calculated as 60 times the number of detected beats divided by the ECG recording duration in seconds.

Additionally, after the final trial, participants were verbally administered eight open-ended questions asking about the clarity of instructions (1. *Was everything clear? What do we need to improve?*), perceived difficulty (2. *What did you experience as the hardest part?* and 3. *Which was the most difficult condition?*), presence of patterns across conditions (4. *Did you see a pattern?*), technical issues (5. *Did you encounter any bugs?*), motion sickness and general discomfort (6. *Did you experience motion sickness? What do you think caused it?* and 7. *Did you experience any other discomfort during or after the experiment?*), and overall impressions (8. *Do you have any other remarks about the experiment?*). The experimenter summarized and typed the participants' verbal responses.

2.6. Experimental protocol

Upon arrival at the experimental room, participants read and signed the consent form. They were then provided with instructions supported by a presentation about how to perform during the experiment. In particular, participants were instructed to collect ten items from a supermarket shelf and place them in a cart located on their right, adhering to a sequence presented on a shopping list positioned above the cart. The presentation also explained how to use the controller by pointing out that objects could be grabbed and released using a button. Participants were verbally instructed to use only their right hand and to keep their left hand alongside their body throughout the experiment to reduce artifacts in the ECG recordings. In addition, participants were instructed to complete the task as accurately and as fast as possible, prioritizing

¹ The other four items do not measure the facets of workload that directly relate to our research question. Specifically, for *Physical demands*, there is no major difference expected between conditions, as the physical actions required are the same for all participants (i.e., picking up and placing the ten objects). Similarly, regarding *Temporal demands*, the task instructions imposed are identical for all participants. The *Performance* item measures how well participants believe they are performing the task, which is an outcome measure that has already been assessed objectively (task completion time). As for *Frustration*, this reflects the extent to which someone feels hindered/obstructed, which is more of an individual evaluation or appraisal of the task, rather than the workload as we intend to measure it. It can be expected that these facets may be affected by the experimental conditions, but not in the direct sense that we are interested in.

accuracy over speed. Furthermore, participants were informed about physical safety, warned of potential motion sickness, and were assured they could halt the experiment should they experience any discomfort.

After the presentation, participants attached the ECG electrodes to their torso following a lead II configuration. Next, the participants put on the HMD, and the eye-tracker in the device was calibrated. Participants began with a 5-min familiarization phase to allow them to acclimate to the virtual environment and task. A partially filled shelf was presented in front of them for interaction and they were invited to ask questions as needed. They then completed a practice trial where they performed the shopping task without characters, background noise, or arithmetic task. Once they added all items from their shopping list to the cart, they removed the HMD and completed the first NASA-TLX questionnaire through Google Forms. The practice trial was used solely for familiarization purposes and was excluded from the analyses.

Next, each participant underwent the four experimental trials, each trial under a particular condition (i.e., Characters, Noise, Arithmetic, or Combined) in a counterbalanced order, using all 24 permutations of the four conditions (complete counterbalancing). Before each trial, the eye-tracker was recalibrated. Participants were not informed about the condition characteristics before the trial started. After each trial, participants removed the HMD and filled out a NASA-TLX questionnaire related to the just-performed condition. After the final trial, participants completed an oral questionnaire regarding their experiences, challenges, and any discomfort experienced.

The experiment, including intake and questionnaires, lasted approximately 50 min per participant (Fig. 4).

2.7. Statistical analysis

Comparisons between all six combinations of conditions were conducted using paired-sample *t*-tests. All analyses were performed in MATLAB (R2024a). To account for multiple comparisons, we reduced the critical alpha value to $0.05/6 \approx 0.008$.

We used Cohen's *d*, a commonly used measure that describes the difference in means relative to the pooled standard deviation s_p (Cohen, 1969; Goulet-Pelletier and Cousineau, 2018):

$$d = \frac{M_1 - M_2}{s_p} \quad (1)$$

$$s_p = \sqrt{\frac{(n-1)s_1^2 + (n-1)s_2^2}{2n-2}} = \sqrt{\frac{23s_1^2 + 23s_2^2}{46}} \quad (2)$$

, where s_1^2 is the variance for Condition 1, s_2^2 is the variance for Condition 2, and n is the sample size ($n = 24$).

Note that although our experiment used a within-subjects design, it is still possible to interpret Cohen's *d*. This *d*-value indicates the magnitude of the differences between conditions without taking into account how individual participants respond to changes in condition.

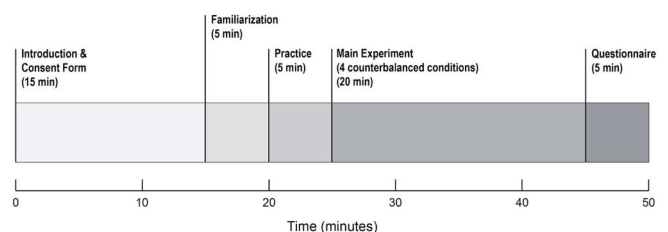


Fig. 4. Timeline of the experimental procedure. After providing informed consent, participants completed a Familiarization phase and a Practice trial (excluded from analysis). This was followed by the main experiment, which consisted of four counterbalanced conditions, with the NASA-TLX being completed after each trial. In the Post-experiment questionnaire phase, participants provided qualitative feedback. Durations are approximate.

We also used Cohen's d_z , a measure that corresponds with the paired-samples t -statistic, and which is defined as the difference in means divided by the standard deviation of the difference scores between the two vectors (Faul et al., 2007; Rosenthal, 1991):

$$d_z = \frac{t}{\sqrt{n}} = \frac{M_{diff}}{S_{diff}} \quad (3)$$

Cohen's d_z represents a statistically more optimal way to describe the magnitude of differences, as it considers how participants' scores change relative to themselves when transitioning from one condition to another. Assuming both scores are positively correlated (e.g., participants with a higher heart rate in Condition 1 will likely also have a higher heart rate in Condition 2), d_z can be expected to be stronger than d .

Thus, reporting both effect sizes provides a more comprehensive view of the results, where d_z indicates the standardized change within participants, while d describes this change with respect to the variability from both conditions. A d value of 0.2 was interpreted as a small effect size, 0.5 as a medium effect, and 0.8 or above as a large effect (Cohen, 1969).

Table 3

Means, standard deviations (SD), and results from the paired-samples t -tests, including effect sizes (Cohen's d and d_z), for the four dependent measures across the experimental conditions.

Task completion time				
	Mean (SD)	Characters	Noise	Arithmetic
Characters	61.2 s (10.9)			
Noise	65.8 s (13.5)	$t = 3.08, p = \mathbf{0.005}$ ($d = 0.38, d_z = 0.63$)		
Arithmetic	86.3 s (16.0)	$t = 9.27, p < \mathbf{0.001}$ ($d = 1.83, d_z = 1.89$)	$t = 7.46, p < \mathbf{0.001}$ ($d = 1.38, d_z = 1.52$)	
Combined	87.6 s (18.3)	$t = 7.42, p < \mathbf{0.001}$ ($d = 1.75, d_z = 1.52$)	$t = 5.81, p < \mathbf{0.001}$ ($d = 1.36, d_z = 1.19$)	$t = 0.45, p = 0.658$ ($d = 0.08, d_z = 0.09$)
Heart rate				
	Mean (SD)	Characters	Noise	Arithmetic
Characters	101.5 bpm (12.0)			
Noise	101.8 bpm (12.1)	$t = 0.42, p = 0.675$ ($d = 0.02, d_z = 0.09$)		
Arithmetic	107.9 bpm (14.8)	$t = 4.62, p < \mathbf{0.001}$ ($d = 0.47, d_z = 0.94$)	$t = 4.33, p < \mathbf{0.001}$ ($d = 0.45, d_z = 0.88$)	
Combined	107.3 bpm (14.4)	$t = 5.12, p < \mathbf{0.001}$ ($d = 0.44, d_z = 1.05$)	$t = 5.18, p < \mathbf{0.001}$ ($d = 0.41, d_z = 1.06$)	$t = -0.44, p = 0.664$ ($d = -0.04, d_z = -0.09$)
Pupil diameter				
	Mean (SD)	Characters	Noise	Arithmetic
Characters	4.79 mm (0.59)			
Noise	4.84 mm (0.62)	$t = 2.19, p = 0.039$ ($d = 0.09, d_z = 0.45$)		
Arithmetic	4.86 mm (0.61)	$t = 1.79, p = 0.086$ ($d = 0.12, d_z = 0.37$)	$t = 0.39, p = 0.699$ ($d = 0.03, d_z = 0.08$)	
Combined	4.91 mm (0.59)	$t = 3.99, p < \mathbf{0.001}$ ($d = 0.20, d_z = 0.81$)	$t = 1.81, p = 0.083$ ($d = 0.11, d_z = 0.37$)	$t = 1.21, p = 0.238$ ($d = 0.08, d_z = 0.25$)
Self-reported mental demand & effort (1 to 10)				
	Mean (SD)	Characters	Noise	Arithmetic
Characters	4.40 (1.78)			
Noise	5.00 (1.97)	$t = 2.05, p = 0.052$ ($d = 0.32, d_z = 0.42$)		
Arithmetic	6.67 (1.31)	$t = 8.33, p < \mathbf{0.001}$ ($d = 1.45, d_z = 1.70$)	$t = 4.77, p < \mathbf{0.001}$ ($d = 1.00, d_z = 0.97$)	
Combined	7.17 (1.38)	$t = 10.23, p < \mathbf{0.001}$ ($d = 1.74, d_z = 2.09$)	$t = 7.52, p < \mathbf{0.001}$ ($d = 1.28, d_z = 1.53$)	$t = 2.50, p = 0.020$ ($d = 0.37, d_z = 0.51$)

Note. Statistically significant differences ($p < 0.05/6$) for the paired-samples t -tests (corresponding to $d_z \leq -0.59$ or $d_z \geq 0.59$) are indicated in boldface. The degrees of freedom (df) for the paired-samples t -tests is 23 in all cases.

3. Results

Table 3 presents descriptive statistics (i.e., means and standard deviations), calculated over the 24 participants, along with paired-sample t -test results for each of the four dependent measures. We included the effect size (d), which describes the difference between means relative to the pooled standard deviation, and the within-subjects effect size (d_z), which is directly related to the t -statistic.

3.1. Effects of visual, auditory, and cognitive demands on the dependent measures

To test our hypotheses, we investigated how visual, auditory, and cognitive demands affected task completion time, self-reported mental demand & effort, heart rate, and pupil diameter across four experimental conditions.

For task completion time, heart rate, and self-reported mental demand & effort, the arithmetic task, both in isolation (Arithmetic condition) and combined with other demands (Combined condition), produced statistically significantly higher scores than the Characters and Noise conditions. Pupil diameter was also significantly larger in the

Combined condition than in the Characters condition.

No measure differed significantly between the Combined and Arithmetic conditions, although self-reported mental demand & effort showed a moderate positive effect size. Furthermore, the Noise condition led to a significant increase in task completion time compared to the Characters condition.

In summary, the results support our hypothesis that cognitive demands result in a higher workload than visual and auditory demands. In addition, our results indicate that auditory demands exert a stronger impact on task completion time than visual demands. While our focus was on mental workload, increases in Temporal demand and Frustration were also observed (see Appendix A), indicating that the arithmetic task in the Arithmetic and Combined conditions also elevated non-cognitive facets of workload compared to the Characters and Noise conditions.

3.2. Effect sizes

Effect sizes in Table 3 indicate that some dependent measures distinguished between conditions in terms of both the standard effect size (high d) and the paired effect size (high d_z), whereas others differentiated only within-subject (high d_z but low d). For example, pupil diameter is a measure that is sensitive to within-subject changes (high d_z), but shows a small effect relative to the overall pooled variance (low d). This is illustrated in the scatter plots in Fig. 5, which show the distribution of the participants' scores for the Combined and Characters conditions. For pupil diameter (Fig. 5D), it can be seen that many data points lie above the diagonal line, indicating that the pupils of most participants are dilated in the Combined condition relative to the Characters condition. However, relative to the size of individual

differences, the effect is small; in the boxplot (Fig. 6), there is little visible distinction between conditions.

3.3. Post-experiment questionnaire

In the post-experiment questionnaire (see Appendix B for details), all participants rated the Arithmetic and/or Combined condition as the most difficult; the Combined condition was most frequently named

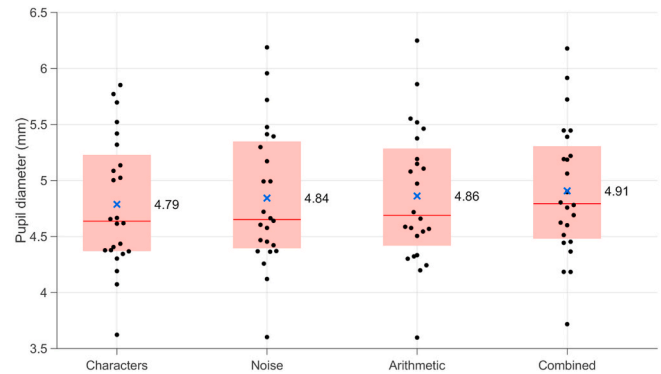


Fig. 6. Boxplots of the pupil diameter for the four experimental conditions. The boxes extend from the 25th to the 75th percentiles, with the horizontal line inside the boxes representing the median. The blue cross and its corresponding numeric value indicate the mean across participants. The markers correspond to individual participants. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

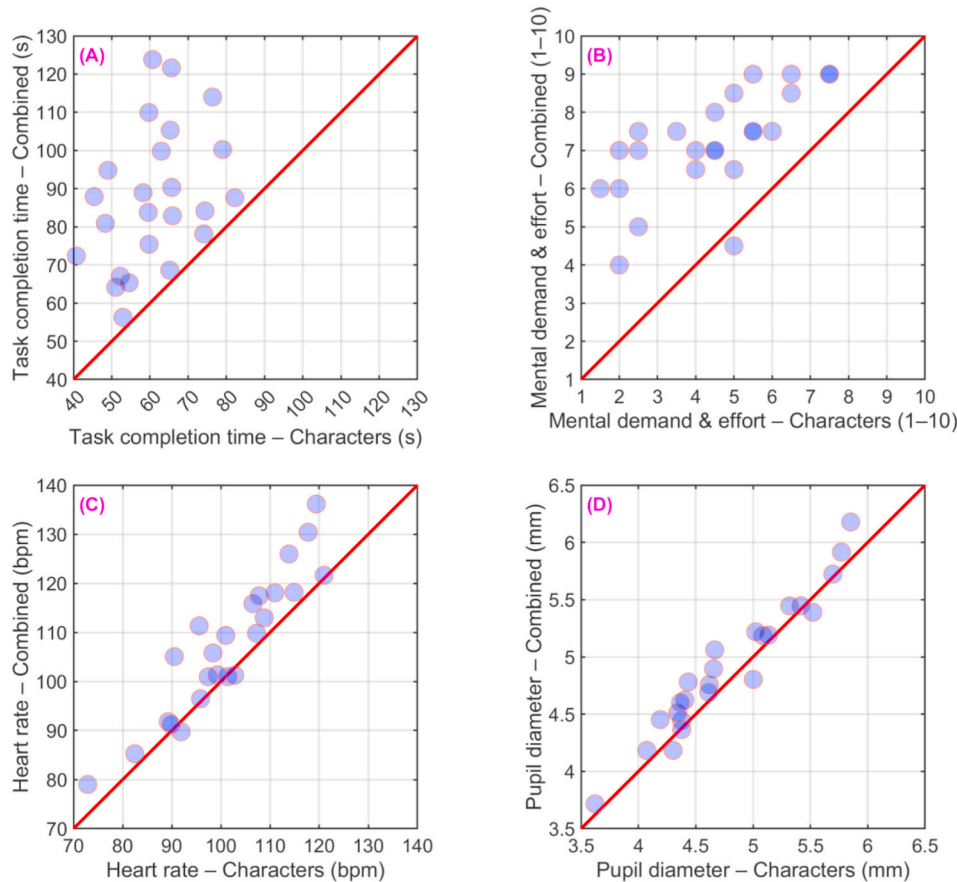


Fig. 5. Scatter plots of the four dependent measures (A: Task completion time, B: Mental demand & effort, C: Heart rate, D: Pupil diameter), comparing scores in the Combined condition to those in the Characters condition. Transparent markers are used to make overlapping data points more visible. The red diagonal lines represent equal values for both conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(16/24), with four naming Arithmetic and four naming both. Twenty-three of 24 participants reported no motion sickness, and one reported experiencing it slightly. Instructions were largely clear (21/24 reported full clarity), and most participants did not notice any pattern in the trials (21/24). Technical issues were rare, limited to occasional object-interaction quirks (e.g., difficulty picking up onions/leeks, one disappearing item, the cart filling too quickly) and minor visual artifacts (e.g., blurry corners). Overall feedback was positive; several participants described the task as fun while suggestions included varying or making products more distinctive and clarifying whether time pressure was intended.

4. Discussion

4.1. Main findings

In this study, 24 participants performed a shopping task in an immersive virtual supermarket under four conditions: visual characters, auditory background noise, a cognitive arithmetic task, and all three elements combined. The results show that the cognitive task had a strong impact on completion time, self-report, and heart rate, whereas pupil diameter showed a significant increase only for the Combined vs. Characters conditions. Auditory demands in the form of background noise also impaired task performance compared to visual demands.

Our findings are consistent with the multiple resource theory (Wickens, 2002), which suggests that interference occurs when tasks draw on the same pools of resources. In our study, the arithmetic task imposed a high load on central processing resources, competing directly with the cognitive requirements of the shopping task, such as reading, remembering, and searching for products. In contrast, participants reported comparatively low mental workload in the Characters and Noise conditions. They likely focused solely on the primary task and filtered out these supplementary stimuli, akin to the phenomenon of inattention blindness (Mack and Rock, 1998). Previous studies have reported similar findings. Redlinger et al. (2022) found that adding visual game-like features (e.g., background, 3D-depth elements) to an immersive VR task did not significantly affect task performance when these demands were irrelevant to task execution. On the other hand, Fratini et al. (2023) found increased perceived mental workload when visual and auditory demands were task-relevant, such as in complex crossing scenarios with higher visual clutter, where participants had to monitor multiple elements to make decisions. There is also evidence suggesting that demands may differ in how easily they can be ignored depending on their sensory modality. In general, irrelevant auditory stimuli can be difficult to ignore and may interfere with attention and task performance even when they are not required for task execution (Grenzebach and Romanus, 2022; Doggett et al., 2021). These findings suggest that the impact of additional demands depends not only on the sensory modality but also on their task relevance.

From a design perspective, our results suggest that reducing enforced cognitive demands may be more effective in decreasing mental workload in immersive VR than reducing non-task-relevant visual or auditory demands. In addition, our results show that auditory noise interferes more with task performance (as reflected in longer completion times) than visual distractions. This is probably because auditory demands are more difficult to filter out, even when they are not required for the task. Generally, while sensory distractions play a role, demands that are integral to the task are more likely to increase mental workload than those that can be readily ignored.

Nevertheless, it is still possible that exposing users to higher levels of visual and auditory demands could negatively impact their performance and result in increased physiological activity. If this is the case, dynamically adjusting these demands could create training environments better suited to users' capacities, thereby reducing the risk of overload and disengagement from the task.

4.2. Comparison of dependent measures

Besides providing insights into visual, auditory, and cognitive demands, our findings also offer insights into the usability of different dependent measures. Task completion time was the easiest measure to obtain, as it was extracted from the simulation. Administering the NASA-TLX is also straightforward. These two simple measures also showed the strongest effects between conditions, evident in both the standard (d) and within-subjects (d_z) effect sizes.

Physiological measures, including heart rate and pupil diameter, required specialized equipment and elaborate procedures. Although statistically significant effects emerged, they were mainly within-subject, making it impractical to set absolute thresholds. For example, a pupil diameter of 4.90 mm or greater cannot, by itself, be taken to indicate 'high mental workload' or used as a basis to simplify task conditions or provide additional user support. Conversely, applying this method to performance measures can be more meaningful. For example, claiming that a task completion time exceeding 100 s indicates slow performance is a justifiable assertion. Of course, such thresholds are still sample-specific, and the current mean performance scores cannot be readily generalized to other groups. Older adults or patients, such as people with brain injury, are likely to perform these visual-motor tasks more slowly than the current student sample, with performance further varying by impairment severity and pathology type.

Another challenge in interpreting physiological measures is the physiological drift that occurs over the course of an experiment. While decreases in task time or self-reported workload across trials can be attributed to learning effects, physiological changes are often more ambiguous. When sorted by presentation order, mean pupil diameter decreased steadily across the four trials (4.93, 4.85, 4.84, to 4.78 mm). While this trend could reflect reduced workload due to task learning (Sibley et al., 2011), it may also be influenced by physiological habituation effects over time. Despite this source of variability, pupil diameter proved sensitive enough to reveal a statistically significant difference between experimental conditions (Table 3).

In summary, while heart rate and pupil diameter were sensitive within participants, substantial inter-individual variability, luminance and order effects, and possible measurement confounds limit their usefulness for setting absolute thresholds or driving adaptation of training conditions. By contrast, the other measures we used, task completion time and self-reported mental workload (NASA-TLX mental demand & effort), are more practical for training: they are simple to collect, showed the strongest effects in our data, and translate more directly into actionable decisions (e.g., flagging slow performance or high perceived load). Accordingly, physiological indices should be applied cautiously to track relative changes against participant-specific baselines, particularly in research settings where such baselines can be established.

4.3. Limitations

This experiment has several limitations.

- Our study involved an immersive simulation environment where participants had to carry out search-pick-place actions, a task important for neurorehabilitation purposes, for example. Real shopping tasks involve additional activities such as navigation, while the absence of haptic feedback, known to influence subjective experiences and task performance (Camardella et al., 2023; Fröhlich and Wachsmuth, 2013; Kreimeier et al., 2019; Özen et al., 2022), may further differentiate the experimental setting from real-world shopping. Subtle discrepancies between the virtual and real worlds, such as time delays and impaired depth cues, may also have disrupted the perception-action cycle (e.g., Harris et al., 2019). Future work could replicate this study in a physical mockup to assess similarities and differences between real and virtual environments.

- Participants' in-task actions (e.g., grasp events, gaze targets) were not recorded, which prevented linkage to momentary heart rate or pupil diameter. The pupil diameter is known to be sensitive to changes in workload, as shown in the literature (Ahern, 1978; Klingner, 2010; Marquart and De Winter, 2015). Future research could examine workload variations during tasks using temporally synchronized physiological measures, such as comparing visual search for products on shelves to the motor task of placing them in a cart.
- Although participants were exposed to auditory, visual, and cognitive demands, other types of demands may have partially acted as confounding variables. In particular, the participants' mean heart rate was considerably higher than a typical resting heart rate, suggesting that the physical demands of shopping could have played a role. For example, the heart rate in the Combined condition may have been influenced by participants completing the task at a slower pace, thereby taking longer to finish. In addition, the arithmetic task required verbal responses, which may have introduced some additional physical activity together with the intended cognitive demands. However, verbal (vocal) responses can often be performed simultaneously with spatial tasks with relatively limited interference, because they draw on separate processing codes (Wickens, 2002). Therefore, the longer completion times observed in the arithmetic task conditions are unlikely to be explained by the verbalization itself, but rather by the overall secondary task requirements. Future studies should disentangle the effects of the physical and cognitive components of the secondary arithmetic task. Additionally, we acknowledge that factors such as sustained attention, emotional state, and motivation could have influenced the participants' heart rate and pupil diameter in our study.
- Future research could improve luminance control within the environment. The moving characters likely influenced eye movements and thereby pupil diameter through changes in luminance at the point of regard. Although the Combined and Characters conditions used identical character stimuli, the arithmetic task in Combined may still cause the participant to shift their gaze (e.g., toward the bright shopping list). Future studies could avoid placing bright products on the shelves (see De Winter et al., 2021; Kun et al., 2012, for the importance of controlling for luminance changes in pupillometry) and refrain from using a shopping list displayed on a relatively bright background. Because the HMD was removed between trials, future studies could control for daylight during these intervals by conducting the experiment in a windowless room.
- Task performance might have been affected by the appearance of the products and their locations on the shelf. Ambiguous products could cause confusion, while participants of shorter stature might have struggled with items on higher shelves. Future studies should address these issues by standardizing the virtual camera position in Unity and using more distinct products.
- Our experiment included only four conditions, rather than the eight possible combinations of a fully-crossed $2 \times 2 \times 2$ factorial design. The lack of a baseline condition and certain combinations (e.g., characters + arithmetic) did not allow us to determine with certainty whether a given effect is due, for example, to visual load alone or to the combined effect of visual and cognitive load (see, e.g., load theory of selective attention and cognitive control: Caparos and Linnell, 2009; Lavie et al., 2004). Such interactions should be examined in follow-up research using larger sample sizes. Additionally, a limitation of our study is that we introduced only one level for each of the visual, auditory, and cognitive demands. Future research could examine multiple levels of visual demands by increasing the number and complexity of stimuli, for example, by progressing from stationary figures to interactive digital signage or highly mobile characters. For auditory demands, it could involve transitioning from low background noise to more complex auditory environments with multiple announcements, background music, and customer conversations.

- A final limitation is that the sample consisted of healthy students from a technical university. We did not assess participants' prior immersive VR exposure or grocery-shopping frequency. Such background factors may modulate perceived workload (e.g., via novelty or domain familiarity) and should be recorded and controlled for in future studies. While the virtual environment was novel to all, it could be assumed that participants' overall familiarity with digital technologies was relatively high. Future studies should involve a larger and more diverse sample, including adults of various cognitive abilities and cultural backgrounds, and measure prior experience with immersive VR to assess potential novelty effects, which have been shown to influence user performance and satisfaction in IVR environments (Miguel-Alonso et al., 2023). If this environment is pursued for neurorehabilitation, patients with conditions that can cause sensory hypersensitivity, such as sensory sensitivity following acquired brain injury, should also be included (Thielen et al., 2023).

5. Conclusion

This study, involving 24 participants performing a shopping task in an immersive virtual environment with heart rate and pupil diameter measurements, indicates that a cognitive secondary task contributes most to mental workload and performance costs compared with visual and auditory demands. Additionally, auditory demands had a greater effect on completion time than visual demands.

Adjusting the level of realism may prevent overstimulation and help users focus on the core task at hand. Based on the present results, when simplifying the task environment to reduce mental workload, priority should be given to reducing cognitive demands, followed by auditory demands, and finally visual distractions. The physiological measures (heart rate, pupil diameter) showed smaller standardized effect sizes (d) than task completion time and self-reported mental demand & effort. Therefore, physiological measures should be interpreted against individual baselines.

CRedit authorship contribution statement

Salvatore Luca Cucinella: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Conceptualization. **Joost de Winter:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis, Data curation. **Alex van den Berg:** Supervision, Resources, Conceptualization. **Bibi van den Berg:** Software, Methodology, Investigation, Conceptualization. **Douwe Hoogsteen:** Software, Methodology, Investigation, Conceptualization. **Daan Poutsma:** Software, Methodology, Investigation, Conceptualization. **Floris van Wingerden:** Software, Methodology, Investigation, Conceptualization. **Laura Marchal-Crespo:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Dutch Research Council (NWO Grant no. 18934 – VIDI: “Hyper-Realistic Personalized Multisensory Robotic Neurorehabilitation”) for providing financial support for this project.

The anonymized data are publicly available in the following research data repository: <https://doi.org/10.5281/zenodo.12723709>.

Appendix A. NASA-TLX results per item

Table A1

Mean scores and standard deviations (in parentheses) for the 6 items of the NASA-TLX for the four experimental conditions.

	Characters	Noise	Arithmetic	Combined
Mental demand	4.50 (1.79)	5.33 (2.22)	6.83 (1.34)	7.25 (1.45)
Physical demand	4.50 (2.19)	4.50 (2.15)	4.46 (1.84)	4.83 (1.93)
Temporal demand	5.00 (2.27)	5.33 (2.12)	6.38 (1.56)	6.46 (1.72)
Performance	8.25 (0.99)	7.88 (1.51)	6.58 (1.61)	6.62 (1.58)
Effort	4.29 (2.01)	4.67 (1.81)	6.50 (1.41)	7.08 (1.50)
Frustration	2.88 (2.07)	3.08 (2.04)	4.71 (2.12)	5.12 (2.54)

Note. Means and standard deviations were calculated over 24 participants. Items were scored on a scale from 1 (low) to 10 (high).

Table A2

Cohen's *d* (with Cohen's *d_z* in parentheses) for the 6 items of the NASA-TLX for comparisons with the Combined condition.

	Characters vs. Combined	Noise vs. Combined	Arithmetic vs. Combined
Mental demand	-1.69 (-1.82)	-1.02 (-1.23)	-0.30 (-0.39)
Physical demand	-0.16 (-0.24)	-0.16 (-0.30)	-0.20 (-0.25)
Temporal demand	-0.73 (-0.95)	-0.58 (-0.72)	-0.05 (-0.06)
Performance	1.23 (1.18)	0.81 (0.90)	-0.03 (-0.03)
Effort	-1.57 (-1.70)	-1.45 (-1.52)	-0.40 (-0.53)
Frustration	-0.97 (-1.05)	-0.89 (-0.91)	-0.18 (-0.26)

Appendix B. Post-experiment questionnaire results

The post-experiment questionnaire collected qualitative feedback from all 24 participants.

Table B1

Summary of qualitative feedback collected from each participant after the experiment.

Participant	Was everything clear? What do we need to improve?	What did you experience as the hardest part?	Did you see a pattern?	Did you encounter any bugs?	Did you experience motion sickness? What do you think caused it?	Did you experience any other discomfort during or after the experiment?	Do you have any other remarks about the experiment?	The most difficult condition?
1	Fairly clear, better explain the products	Sound with mental math	No	No	No	Slightly blurry vision, more tired later	No	Combined condition
2	Was clear	Math with all sounds and shopping	No	Leek disappeared while grabbing another one	No	No	Was very fun	Combined condition
3	Was clear	Everything at the same time	No	No	No	No	Try different products? Always the same ones	Combined condition
4	Yes, was clear	Mental math, neither was harder	No	No	No	No	No	Arithmetic & Combined conditions
5	Was clear	Mental math with everything at once	No	No	No	No	Well done, label placement good	Combined condition
6	Was clear	Counting down, first time harder (101-7)	No	No	Not really	No	No	Combined condition
7	Was clear	Mental math, 99-7 difficult	No	No	No	No	No	Arithmetic & Combined conditions
8	Was clear	Sound with mental math	No	No	No	No	No, looked good	Combined condition
9	Was clear	Math with sound	Yes	No	No	No	Was quite fun	Combined condition
10	Was clear	Math while grabbing objects, time pressure with sound	No	No	No	No	Was fun	Combined condition
11	Was clear	Mental math 99-7	No	No	No	No	Looks good, fun	Arithmetic condition
12	Was clear	Everything at once	No	No	No	No	Curious about the results	Combined condition

(continued on next page)

Table B1 (continued)

Participant	Was everything clear? What do we need to improve?	What did you experience as the hardest part?	Did you see a pattern?	Did you encounter any bugs?	Did you experience motion sickness? What do you think caused it?	Did you experience any other discomfort during or after the experiment?	Do you have any other remarks about the experiment?	The most difficult condition?
13	Was clear	Math + grabbing products, 99-7 a bit difficult	Similar locations	No	No	No	No	Arithmetic condition
14	Not clear	Mental math with sound	No	Leeks hard to see on top	No	No	No	Combined condition
15	Was clear	Everything at once	No	No	No	No	No	Combined condition
16	Was clear	Math with everything at once	No	Leeks hard to see on top	No	No	Well done, detergent unclear	Combined condition
17	Clear	Remembering during math, Arithmetic condition harder than the Combined condition	Same order	No	No	No, nothing	Clarify if there was time pressure	Arithmetic condition
18	Clear	Combination of everything	Not noticed	Missed grabs, blurry corners	Slightly	No	Was clear enough	Combined condition
19	Clear	Mental math	No	Trouble grabbing onions	No	No	No	Arithmetic & Combined conditions
20	Yes, except only one shelf	Finding products, partial images, multitasking	No	Cart filled too quickly	No	No	No	Combined condition
21	Everything clear	Mental math, second one harder (Arithmetic condition)	No	No	No	No	Totally fine	Arithmetic condition
22	Everything clear	Mental math with everything	No	Onion doesn't look like onion	No	No	Everything fine	Combined condition
23	Everything clear	Math while shopping, all together	No	No	No	No	Everything fine	Combined condition
24	Clear	Math, neither one harder	No	No	No	No	Was quite fun	Arithmetic & Combined conditions

Notes. The responses in the table represent a summary of participants' answers to the questionnaire.

References

- Ahern, S.K., 1978. *Activation and Intelligence: Pupillometric Correlates of Individual Differences in Cognitive Abilities* (Doctoral Dissertation). University of California.
- Artillery Intelligence, 2024. VR usage & consumer attitudes, wave 9. <https://artilleryiq.com/reports/vr-usage-consumer-attitudes-wave-9/>.
- Benoit, C.-E., Solopchuk, O., Borraán, G., Carbone, A., Van Durme, S., Zénon, A., 2019. Cognitive task avoidance correlates with fatigue-induced performance decrement but not with subjective fatigue. *Neuropsychologia* 123, 30–40. <https://doi.org/10.1016/j.neuropsychologia.2018.06.017>.
- Bertrand, A., Maxwell, W., Eagan, J.R., 2023. Questioning the ability of feature-based explanations to empower non-experts in robo-advised financial decision-making. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, pp. 943–958. <https://doi.org/10.1145/3593013.3594053>.
- Bradley, M.M., Miccoli, L., Escrig, M.A., Lang, P.J., 2008. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 602–607. <https://doi.org/10.1111/j.1469-8986.2008.00654.x>.
- Bristow, T., Jih, C.S., Slabich, A., Gunn, J., 2016. Standardization and adult norms for the sequential subtracting tasks of serial 3's and 7's. *Appl. Neuropsychol. Adult* 23, 372–378. <https://doi.org/10.1080/23279095.2016.1179504>.
- Camardella, C., Chiaradia, D., Bortone, I., Frisoli, A., Leonardis, D., 2023. Introducing wearable haptics for rendering velocity feedback in VR serious games for neuro-rehabilitation of children. *Front Virtual Real.* 3, 1019302. <https://doi.org/10.3389/frvir.2022.1019302>.
- Caparos, S., Linnell, K.J., 2009. The interacting effect of load and space on visual selective attention. *Vis. Cogn.* 17, 1218–1227. <https://doi.org/10.1080/13506280902924083>.
- Charles, R.L., Nixon, J., 2019. Measuring mental workload using physiological measures: a systematic review. *Appl. Ergon.* 74, 221–232. <https://doi.org/10.1016/j.apergo.2018.08.028>.
- Chen, I.C., Chuang, I.C., Chang, K.C., Chang, C.H., Wu, C.Y., 2023. Dual task measures in older adults with and without cognitive impairment: response to simultaneous cognitive-exercise training and minimal clinically important difference estimates. *BMC Geriatr.* 23, 663. <https://doi.org/10.1186/s12877-023-04390-3>.
- Cohen, J., 1969. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, New York.
- Cucinella, S.L., De Winter, J.C.F., Grauwmeijer, E., Evers, M., Marchal-Crespo, L., 2025. Towards personalized immersive virtual reality neurorehabilitation: a human-centered design. *J. NeuroEng. Rehabil.* 22, 7. <https://doi.org/10.1186/s12984-024-01489-5>.
- Dahlstrom, N., Dekker, S., Van Winsen, R., Nyce, J., 2009. Fidelity and validity of simulator training. *Theor. Issues Ergon. Sci.* 10, 305–314. <https://doi.org/10.1080/14639220802368864>.
- De Pasquale, P., Bonanno, M., Mojdehdehbaheer, S., Quartarone, A., Calabrò, R.S., 2024. The use of head-mounted display systems for upper limb kinematic analysis in post-stroke patients: a perspective review on benefits, challenges and other solutions. *Bioengineering* 11, 538. <https://doi.org/10.3390/bioengineering11060538>.
- De Winter, J.C.F., Petermeijer, S.M., Kooijman, L., Dodou, D., 2021. Replicating five pupillometry studies of Eckhard Hess. *Int. J. Psychophysiol.* 165, 145–205. <https://doi.org/10.1016/j.ijpsycho.2021.03.003>.
- Doggett, R., Sander, E.J., Birt, J., Ottley, M., Baumann, O., 2021. Using virtual reality to evaluate the impact of room acoustics on cognitive performance and well-being. *Front Virtual Real.* 2, 620503. <https://doi.org/10.3389/frvir.2021.620503>.
- Elghoul, Y., Bahri, F., Trabelsi, K., Chtourou, H., Frikha, M., Clark, C.C.T., Glenn, J.M., Bragazzi, N., Souissi, N., 2022. Optimizing motor learning: difficulty manipulation combined with feedback frequency enhance under-time-pressure fine-motor-coordination skill acquisition and retention. *J. Mot. Behav.* 54, 490–502. <https://doi.org/10.1080/00222895.2021.2016573>.
- Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. <https://doi.org/10.3758/BF03193146>.
- Faure, C., Limballe, A., Bideau, B., Kulpa, R., 2020. Virtual reality to assess and train team ball sports performance: a scoping review. *J. Sports Sci.* 38, 192–205. <https://doi.org/10.1080/02640414.2019.1689807>.
- Fisher, A.V., Godwin, K.E., Seltman, H., 2014. Visual environment, attention allocation, and learning in young children: when too much of a good thing may be bad. *Psychol. Sci.* 25, 1362–1370. <https://doi.org/10.1177/0956797614533801>.
- Fratini, E., Welsh, R., Thomas, P., 2023. Ranking crossing scenario complexity for eHMs testing: a virtual reality study. *Multimodal. Technol. Interact.* 7, 16. <https://doi.org/10.3390/mti7020016>.
- Fröhlich, J., Wachsmuth, I., 2013. The visual, the auditory and the haptic – a user study on combining modalities in virtual worlds. In: Shumaker, R. (Ed.), *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*. Springer, Berlin, Heidelberg, pp. 240–249. https://doi.org/10.1007/978-3-642-39405-8_19. VAMR 2013.
- Garde, A., Laursen, B., Jørgensen, A., Jensen, B., 2002. Effects of mental and physical demands on heart rate variability during computer work. *Eur. J. Appl. Physiol.* 87, 456–461. <https://doi.org/10.1007/s00421-002-0656-7>.

- Gonçalves, G., Coelho, H., Monteiro, P., Melo, M., Bessa, M., 2022. Systematic review of comparative studies of the impact of realism in immersive virtual experiences. *ACM Comput. Surv.* 55, 115. <https://doi.org/10.1145/3533377>.
- Goulet-Pelletier, J.-C., Cousineau, D., 2018. A review of effect sizes and their confidence intervals, part I: the Cohen's *d* family. *Quantitat. Methods Psychol.* 14, 242–265. <https://doi.org/10.20982/tqmp.14.4.p242>.
- Grenzebach, J., Romanus, E., 2022. Quantifying the effect of noise on cognitive processes: a review of psychophysiological correlates of workload. *Noise Health* 24, 199–214. <https://doi.org/10.4103/nah.nah.34.22>.
- Guadagnoli, M.A., Lee, T.D., 2004. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J. Mot. Behav.* 36, 212–224. <https://doi.org/10.3200/JMBR.36.2.212-224>.
- Harris, D.J., Buckingham, G., Wilson, M.R., Vine, S.J., 2019. Virtually the same? How impaired sensory information in virtual reality may disrupt vision for action. *Exp. Brain Res.* 237, 2761–2766. <https://doi.org/10.1007/s00221-019-05642-8>.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Hertzum, M., 2022. Associations among workload dimensions, performance, and situational characteristics: a meta-analytic review of the Task load index. *Behav. Inf. Technol.* 41, 3506–3518. <https://doi.org/10.1080/0144929X.2021.2000642>.
- Hess, T.M., Smith, B.T., Sharifian, N., 2016. Aging and effort expenditure: the impact of subjective perceptions of task demands. *Psychol. Aging* 31, 653–660. <https://doi.org/10.1037/pag0000127>.
- Hoover, A., Singh, A., Fishel-Brown, S., Muth, E., 2012. Real-time detection of workload changes using heart rate variability. *Biomed. Signal Process Control* 7, 333–341. <https://doi.org/10.1016/j.bspc.2011.07.004>.
- Hou, N., Nishina, D., Sugita, S., Jiang, R., Kindaichi, S., Oishi, H., Shimizu, A., 2024. Virtual reality space in architectural design education: learning effect of scale feeling. *Build. Environ.* 248, 111060. <https://doi.org/10.1016/j.buildenv.2023.111060>.
- Imbo, I., Vandierenonck, A., Vergauwe, E., 2007. The role of working memory in carrying and borrowing. *Psychol. Res.* 71, 467–483. <https://doi.org/10.1007/s00426-006-0044-8>.
- Johnston, J.H., Driskell, J.E., Salas, E., 1997. Vigilant and hypervigilant decision making. *J. Appl. Psychol.* 82, 614–622. <https://doi.org/10.1037/0021-9010.82.4.614>.
- Kahneman, D., Beatty, J., 1966. Pupil diameter and load on memory. *Science* 154, 1583–1585. <https://doi.org/10.1126/science.154.3756.1583>.
- Klingner, J., 2010. *Measuring Cognitive Load During Visual Tasks by Combining Pupillometry and Eye Tracking* (Doctoral Dissertation). Stanford University.
- Kreimeier, J., Hammer, S., Friedmann, D., Karg, P., Bühner, C., Bankel, L., Götzmann, T., 2019. Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality. In: *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 289–298. <https://doi.org/10.1145/3316782.3321536>.
- Krijn, M., Emmelkamp, P.M.G., Olafsson, R.P., Biemond, R., 2004. Virtual reality exposure therapy of anxiety disorders: a review. *Clin. Psychol. Rev.* 24, 259–281. <https://doi.org/10.1016/j.cpr.2004.04.001>.
- Kritikos, J., Alevizopoulos, G., Koutsouris, D., 2021. Personalized virtual reality human-computer interaction for psychiatric and neurological illnesses: a dynamically adaptive virtual reality environment that changes according to real-time feedback from electrophysiological signal responses. *Front. Hum. Neurosci.* 15, 596980. <https://doi.org/10.3389/fnhum.2021.596980>.
- Kun, A.L., Palinko, O., Razumenić, I., 2012. Exploring the effects of size and luminance of visual targets on the pupillary light reflex. In: *Automotive '12: Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Portsmouth, UK, pp. 183–186. <https://doi.org/10.1145/2390256.2390287>.
- Ladeira, W., De Oliveira Santini, F., Jardim, W.C., 2020. Gaze behaviour in front-of-shelf orientation. *Int. J. Retail Distrib. Manag.* 48, 186–206. <https://doi.org/10.1108/IJRDM-04-2019-0128>.
- Langhorne, P., Bernhardt, J., Kwakkel, G., 2011. Stroke rehabilitation. *Lancet* 377, 1693–1702. [https://doi.org/10.1016/S0140-6736\(11\)60325-5](https://doi.org/10.1016/S0140-6736(11)60325-5).
- Lavie, N., Hirst, A., De Fockert, J.W., Viding, E., 2004. Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* 133, 339–354. <https://psycnet.apa.org/doi/10.1037/0096-3445.133.3.339>.
- Levac, D.E., Huber, M.E., Sternad, D., 2019. Learning and transfer of complex motor skills in virtual reality: a perspective review. *J. NeuroEng. Rehabil.* 16, 121. <https://doi.org/10.1186/s12984-019-0587-8>.
- Levin, M.F., Demers, M., 2021. Motor learning in neurological rehabilitation. *Disabil. Rehabil.* 43, 3445–3453. <https://doi.org/10.1080/09638288.2020.1752317>.
- Lewis, M.M., Waltz, C., Scelina, K., Scelina, L., Owen, K., Hastilow, K., Miller Koop, M., Rosenfeldt, A.B., Alberts, J.L., 2023. Older adults exhibit declines in instrumental activities of daily living during a virtual grocery shopping task. *Front. Virtual Real.* 4. <https://doi.org/10.3389/frvir.2023.1261096>.
- Luong, T., Martin, N., Raison, A., Argelaguet, F., Diverrez, J.-M., Lécuyer, A., 2020. Towards real-time recognition of users' mental workload using integrated physiological sensors into a VR HMD. In: *Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality*, Porto de Galinhas, pp. 425–437. <https://doi.org/10.1109/ISMAR50242.2020.00068>. Brazil.
- Mack, A., Rock, I., 1998. Inattention blindness: perception without attention. In: *Wright, R.D. (Ed.), Visual Attention*. Oxford University Press, pp. 55–76. <https://doi.org/10.1093/oso/9780195126938.003.0003>.
- Mao, R.Q., Lan, L., Kay, J., Lohre, R., Ayeni, O.R., Goel, D.P., de Sá, D., 2021. Immersive virtual reality for surgical training: a systematic review. *J. Surg. Res.* 268, 40–58. <https://doi.org/10.1016/j.jss.2021.06.045>.
- Marquart, G., De Winter, J., 2015. Workload assessment for mental arithmetic tasks using the task-evoked pupillary response. *PeerJ Comput. Sci.* 1, e16. <https://doi.org/10.7717/peerj-cs.16>.
- Medathati, N.V.K., Desai, R., Hillis, J., 2020. Towards inferring cognitive state changes from pupil size variations in real world conditions. In: *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*. <https://doi.org/10.1145/3379155.3391319>. Stuttgart, Germany.
- Mehler, B., Reimer, B., Coughlin, J.F., 2012. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Hum. Factors* 54, 396–412. <https://doi.org/10.1177/0018720812442086>.
- Melo, M., Gonçalves, G., Monteiro, P., Coelho, H., Vasconcelos-Raposo, J., Bessa, M., 2022. Do multisensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Trans. Visual. Comput. Graph.* 28, 1428–1442. <https://doi.org/10.1109/TVCG.2020.3010088>.
- Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., Bustillo, A., 2023. Countering the novelty effect: a tutorial for immersive virtual reality learning environments. *Appl. Sci.* 13, 593. <https://doi.org/10.3390/app13010593>.
- Munawar, A., Li, Z., Nagururu, N., Trakimas, D., Kazanzides, P., Taylor, R.H., Creighton, F.X., 2024. Fully immersive virtual reality for skull-base surgery: surgical training and beyond. *Int. J. Comput. Assist. Radiol. Surg.* 19, 51–59. <https://doi.org/10.1007/s11548-023-02956-5>.
- Nath, D., Singh, N., Saini, M., Srivastava, M.V.P., Mehndiratta, A., 2022. Design and validation of virtual reality task for neuro-rehabilitation of distal upper extremities. *Int. J. Environ. Res. Publ. Health* 19, 1442. <https://doi.org/10.3390/ijerph19031442>.
- Newman, M., Gatersleben, B., Wyles, K.J., Ratcliffe, E., 2022. The use of virtual reality in environment experiences and the importance of realism. *J. Environ. Psychol.* 79, 101733. <https://doi.org/10.1016/j.jenvp.2021.101733>.
- Novak, D., Zihnerl, J., Olensek, A., Milavec, M., Podobnik, J., Mihelj, M., Munih, M., 2010. Psychophysiological responses to robotic rehabilitation tasks in stroke. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 351–361. <https://doi.org/10.1109/TNSRE.2010.2047656>.
- Ouellet, É., Boller, B., Corriveau-Lecavalier, N., Cloutier, S., Belleville, S., 2018. The virtual shop: a new immersive virtual reality environment and scenario for the assessment of everyday memory. *J. Neurosci. Methods* 303, 126–135. <https://doi.org/10.1016/j.jneumeth.2018.03.010>.
- Özen, Ö., Buetler, K.A., Marchal-Crespo, L., 2022. Towards functional robotic training: motor learning of dynamic tasks is enhanced by haptic rendering but hampered by arm weight support. *J. NeuroEng. Rehabil.* 19, 19. <https://doi.org/10.1186/s12984-022-00993-w>.
- Palacios-Navarro, G., Hogan, N., 2021. Head-mounted display-based therapies for adults post-stroke: a systematic review and meta-analysis. *Sensors* 21, 1111. <https://doi.org/10.3390/s21041111>.
- Pan, J., Tompkins, W.J., 1985. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* BME 32, 230–236. <https://doi.org/10.1109/TBME.1985.325532>.
- Partala, T., Surakka, V., 2003. Pupil size variation as an indication of affective processing. *Int. J. Hum. Comput. Stud.* 59, 185–198. [https://doi.org/10.1016/S1071-5819\(03\)00017-X](https://doi.org/10.1016/S1071-5819(03)00017-X).
- Pastel, S., Petri, K., Chen, C.H., Wiegand Cáceres, A.M., Stirnatis, M., Nübel, C., Schlotter, L., Witte, K., 2023. Training in virtual reality enables learning of a complex sports movement. *Virtual Real.* 27, 523–540. <https://doi.org/10.1007/s10055-022-00679-7>.
- Ragan, E.D., Bowman, D.A., Kopper, R., Stinson, C., Scerbo, S., McMahan, R.P., 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Trans. Visual. Comput. Graph.* 21, 794–807. <https://doi.org/10.1109/TVCG.2015.2403312>.
- Raghubar, K.P., Barnes, M.A., Hecht, S.A., 2010. Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. <https://doi.org/10.1016/j.lindif.2009.10.005>.
- Redlinger, E., Glas, B., Rong, Y., 2022. Impact of visual game-like features on cognitive performance in a virtual reality working memory task: within-subjects experiment. *JMIR Serious Games* 10, e35295. <https://doi.org/10.2196/35295>.
- Rosenthal, R., 1991. *Meta-Analytic Procedures for Social Research*. SAGE Publications, Newbury Park, CA. <https://doi.org/10.4135/9781412984997>.
- Runswick, O.R., Roca, A., Mark Williams, A., Bezodis, N.E., McRobert, A.P., North, J.S., 2018. The impact of contextual information and a secondary task on anticipation performance: an interpretation using cognitive load theory. *Appl. Cogn. Psychol.* 32, 141–149. <https://doi.org/10.1002/acp.3386>.
- Saldana, D., Neureither, M., Schmiesing, A., Jahng, E., Kysh, L., Roll, S.C., Liew, S.L., 2020. Applications of head-mounted displays for virtual reality in adult physical rehabilitation: a scoping review. *Am. J. Occup. Ther.* 74, 7405205060. <https://doi.org/10.5014/ajot.2020.041442>.
- Sedghamiz, H., 2014. Complete pan tomlpkins implementation ECG QRS detector. <https://nl.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector>.
- Sibley, C., Coyne, J., Baldwin, C., 2011. Pupil dilation as an index of learning. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 55, 237–241. <https://doi.org/10.1177/1071181311551049>.
- Specht, J., Stegmann, B., Gross, H., Krakow, K., 2023. Cognitive training with head-mounted display virtual reality in neurorehabilitation: pilot randomized controlled trial. *JMIR Serious Games* 11, e45816. <https://doi.org/10.2196/45816>.
- Spyridonis, F., Daylamani-Zad, D., Nightingale, J., 2024. PublicVR: a virtual reality exposure therapy intervention for adults with speech anxiety. *Virtual Real.* 28, 105. <https://doi.org/10.1007/s10055-024-00998-x>.

- Stanica, I.-C., Moldoveanu, F., Portelli, G.-P., Dascalu, M.-I., Moldoveanu, A., Ristea, M. G., 2020. Flexible virtual reality system for neurorehabilitation and quality of life improvement. *Sensors* 20, 6045. <https://doi.org/10.3390/s20216045>.
- Statista, 2023. Number of virtual reality (VR) and augmented reality (AR) users in the United States from 2017 to 2023. <https://www.statista.com/statistics/1017008/united-states-vr-ar-users>.
- Sweller, J., 2020. Cognitive load theory and educational technology. *Educ. Technol. Res. Dev.* 68, 1–16. <https://doi.org/10.1007/s11423-019-09701-3>.
- Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., Zhang, T., 2019. A systematic review of physiological measures of mental workload. *Int. J. Environ. Res. Publ. Health* 16, 2716. <https://doi.org/10.3390/ijerph16152716>.
- Teodoro, T., Koreki, A., Chen, J., Coebergh, J., Poole, N., Ferreira, J.J., Edwards, M.J., Isaacs, J.D., 2023. Functional cognitive disorder affects reaction time, subjective mental effort and global metacognition. *Brain* 146, 1615–1623. <https://doi.org/10.1093/brain/awac363>.
- Thielen, H., Tuts, N., Welkenhuyzen, L., Huenges Wajer, I.M.C., Lafosse, C., Gillebert, C. R., 2023. Sensory sensitivity after acquired brain injury: a systematic review. *J. Neuropsychol.* 17, 1–31. <https://doi.org/10.1111/jnp.12284>.
- Van Merriënboer, J.J.G., Sweller, J., 2005. Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. <https://doi.org/10.1007/s10648-005-3951-0>.
- Wenk, N., Buetler, K.A., Penalver-Andres, J., Müri, R.M., Marchal-Crespo, L., 2022. Naturalistic visualization of reaching movements using head-mounted displays improves movement quality compared to conventional computer screens and proves high usability. *J. NeuroEng. Rehabil.* 19, 137. <https://doi.org/10.1186/s12984-022-01101-8>.
- Wickens, C.D., 2002. Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* 3, 159–177. <https://doi.org/10.1080/14639220210123806>.
- Wickens, C.D., 2021. Attention: theory, principles, models and applications. *Int. J. Hum. Comput. Interact.* 37, 403–417. <https://doi.org/10.1080/10447318.2021.1874741>.
- Wiskerke, E., Kool, J., Hilfiker, R., Sattelmayer, K.M., Verheyden, G., 2022. Determining the optimal virtual reality exergame approach for balance therapy in persons with neurological disorders using a rasch analysis: longitudinal observational study. *JMIR Serious Games* 10, e30366. <https://doi.org/10.2196/30366>.