Applied Ergonomics 112 (2023) 104039

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

A comparison of the psychological effects of robot motion in physical and virtual environments

Nathan Edward Sanders, Ziyang Xie, Karen B. Chen *

Fitts Department of Industrial and Systems Engineering, North Carolina State University, United States of America

ARTICLE INFO

Keywords:
Human-robot interaction
Virtual reality
Workload

ABSTRACT

A mixed-methods approach was used to assess the fidelity of virtual environments as ergonomic assessment tools for human-robot interaction. Participants performed a visual search task in the physical environment while a nearby collaborative robot repeatedly extended its arm toward them. This scenario was reconstructed in two virtual environments with different levels of graphical detail. Measures of presence, task performance, workload, and anxiety were taken to determine the effect of robot motion in all three environments. Task performance decreased in response to robot motion in the physical environment, while workload and anxiety increased. This simple effect of motion was consistent across environments for measures of task performance and anxiety. However, people performed faster overall in virtual reality, and the effect of motion on workload was greatly reduced in virtual reality. Results in the virtual environments suggest that people were distracted by the sound of the robot, and that presence was affected by manipulations of immersion and coherence.

1. Introduction

Immersive virtual environments (VEs) block out sensory cues from the physical environment and replace them with computer-generated stimuli (Jerald, 2016). The power of virtual reality (VR) is that while users recognize that what they perceive is an illusion, they still respond to it as if it were real (Slater, 2018). The ability of VEs to recreate hazardous situations that may be unattainable in the physical environment and, crucially, human responses to them, makes them a boon to researchers, educators, and clinicians. Virtual reality has proved useful for studying driving behavior (Ménétrier et al., 2017), providing surgical training (Harris et al., 2020), and in treating PTSD (Slater and Sanchez-Vives, 2016; Chen et al., 2021). Additionally, as the Fourth Industrial Revolution accelerates, VEs are being used as ergonomic assessment tools for human-robot interaction (HRI) in simulations free from objective hazards (Gualtieri et al., 2021). But when it comes to the use of VEs as research tools, the question of fidelity must be addressed (Harris et al., 2020). That is, to what extent do VEs elicit realistic human behaviors, cognitions, and affective states? As researchers continue to leverage simulations in VEs to study physical and cognitive ergonomics, it is important to understand the extent to which results obtained in VEs relate to those obtained in the physical environment. This study addresses this question by comparing measures of task performance, workload, and anxiety between physical and virtual environments.

In human factors and ergonomics research, a recent topic of interest is HRI. In the past, non-immersive (i.e. 2D screen-based) VEs have been used to study the perception of safe robot speeds (Duffy et al., 2006). More recently, immersive VEs have been used primarily to "[minimize] mental stress and psychological discomfort [...] while sharing the workspace with robots" (Gualtieri et al., 2021). Immersive VEs have been developed for training (Matsas et al., 2013; You et al., 2018; Pérez et al., 2019; Shu et al., 2019; Sievers et al., 2020), assessing human responses to different motion planners (Matsas et al., 2016), evaluating the efficacy of alarms (Matsas and Vosniakos, 2017), and exploring novel interaction strategies (Matsas et al., 2018). VEs have also been used to assess the effect of unpredictable robot motion on human behavior (Oyekan et al., 2019), the effects of robot speed and predictability on task performance, anxiety, and mental workload (Koppenborg et al., 2017), the behavioral and psychological effects of being struck by a robot (Fratczak et al., 2019), and to study how control strategies influence human post-accident behavior (Fratczak et al., 2021).

Since VEs cannot perfectly simulate the sensory qualia of physical scenarios, and since users know that virtual robots cannot hurt them, it seems plausible that human performance, cognition, and affect measured in VR may not be an accurate reflection of reality. To our knowledge, the body of research that has tested for differences between physical and immersive virtual environments in HRI simula-

* Corresponding author at: Fitts Department of Industrial and Systems Engineering, North Carolina State University, 915 Partners Way, Raleigh, NC 27695, USA.
E-mail addresses: nesander@ncsu.edu (N.E. Sanders), zxie8@ncsu.edu (Z. Xie), kbchen2@ncsu.edu (K.B. Chen).

https://doi.org/10.1016/j.apergo.2023.104039
Received 22 November 2022; Received in revised form 15 February 2023; Accepted 19 April 2023
Available online 13 June 2023
0003-6870/

tions is sparse and results have been mixed. For example, Weistroffer et al. (2014) and Paljic (2017) reported that participants preferred being farther away from industrial robots in both physical and virtual environments, but physiological stress increased with proximity only in the physical environment. Kamide et al. (2014) did not find differences in the amount of personal space that people desired between themselves and a robot, but they found differences in how "under control" participants perceived the physical and virtual robots to be. In contrast, Li et al. (2019) found that people preferred closer interaction distances with a physical robot, but participants let the virtual robot come closer when it was accompanied by realistic sounds. The paucity of these comparison studies and their mixed findings underscores the need to assess the fidelity of human factors measures, and to explore how the objective characteristics of VEs might affect fidelity.

One important question in this regard is whether VEs can elicit realistic human responses to nearby robot motion in terms of measures of task performance, mental workload, and anxiety. Mental workload is one of the most widely used constructs in ergonomics (Young et al., 2015), and one of the goals of HRI is the amelioration of workload via the reduction of task demand. Yet it seems plausible that the motion of a nearby robot might actually increase mental workload and decrease task performance via attentional capture. Attentional capture occurs when a salient and unattended stimulus draws attention (Simons, 2000). Resource theory describes attention as a metaphorical pool of mental resources, whose capacity can change depending on physiological arousal and task demand (Kahneman, 1973; Norman and Bobrow, 1975; Young et al., 2002; Yerkes and Dodson, 1908). Mental workload reflects the utilization of that limited pool of attentional resources (Hancock et al., 2019). Auditory distractors (e.g. the sound of robot actuators) can also capture attention (Dalton and Hughes, 2014), especially under conditions of high visual load (Tellinghuisen and Nowak, 2003). Threatening or anxiety-inducing stimuli have also been shown to augment attentional capture (Kim and Anderson, 2020), and industrial robots can induce stress and fear in nearby workers (Arai et al., 2010; Kulić and Croft, 2007; Dehais et al., 2011; Fujita et al., 2010). Both distraction and high mental workload have been associated with inadequate time for information processing and possibly increased likelihood of accidents (Young et al., 2015). The motion and sound of robots, and their effect on task performance, workload, and stress, are therefore important considerations for the well-being of workers who are engaged in HRI. Understanding how the objective characteristics of VEs affect their fidelity is also therefore important for the use of VEs as evaluative tools.

In order to study fidelity, the present study adapted the methodological framework of Harris et al. (2021), which defines *fidelity* as "the extent to which a simulation recreates the real-world system, in terms of its appearance but also the affective states, cognitions or behaviors it elicits from its users" (Harris et al., 2020). Past research has also referred to this issue as ecological validity (Araújo et al., 2007; van der Ham et al., 2015; Paljic, 2017), although the term *validity* is also frequently used in training contexts (Harris et al., 2021), referring to transfer of training (Hoareau et al., 2017) or the ability to distinguish real-world differences between user groups (Weiner and Sanchez, 2020). Earlier research on fidelity has focused on driving simulator experiments, which concerned the level of correspondence between the driving behaviors in a simulator and on real roads (Fisher et al., 2011; Porter, 2011; Ranney, 2011). The present study with its focus on measures of performance, workload, and anxiety is therefore framed by Harris et al. (2021) as an evaluation of the behavioral, psychological, and affective fidelity (see Table 1).

The framework of Harris and colleagues suggests testing physical fidelity through measures of realism or presence. Presence refers to "a user's subjective psychological response to a VR system" (Bowman and McMahan, 2007), and comprises the dimensions of place illusion and plausibility (Slater et al., 2022b). Underlying these constructs are the objective measures of immersion and coherence, respectively, referring to the affordances of the system to deliver sensory information such as graphical detail or sound (Slater et al., 2022b), and to the extent that objects in the VE behave "in a reasonable or predictable way" (Skarbez et al., 2021). Although many researchers and designers believe that high levels of presence are necessary for a VE to be effective, Harris et al. (2020) cast doubt on that assertion, and Slater (2018) suggests that "researchers can study how different levels of immersion might correspond to different levels of place illusion, and the extent to which people respond as if events in the virtual world were really happening." Skarbez et al. (2021) make a similar call for exploratory research involving coherence and plausibility.

Table 1
Four types of fidelity, their descriptions, and proposed measures for testing, adapted from Harris et al. (2021).

Type	Description	How to test
Physical	Is there a high degree of detail and realism provided by the physical elements of the simulation?	User reports of presence and experiences of immersion
Behavioral	Does the simulation provide an accurate representation of real task performance?	Ability of the simulation to distinguish real-world performance differences
Affective	Does the simulation elicit emotional responses, such as fear or enjoyment, which are similar to the real task?	Self-reported experiences of users, or online monitoring of psychophysiology
Psychological	Does the simulation replicate the perceptual-cognitive demands of the real task?	Measurement of mental effort, gaze behavior or neural activity

1.1. Research goals

The first research goal (RG1) was to broadly assess the fidelity of VEs in terms of their ability to recreate the real-world effects of robot motion on nearby workers. This was done by selecting a single dependent measure for each of the types of fidelity listed in Table 1 – task performance, workload, and anxiety for behavioral, psychological, and affective fidelity (respectively). The effect of robot motion was measured in the physical environment and two VEs, each with a different level of graphical quality, in a within-subject factorial design. The second research goal (RG2) was to explore how two objective characteristics of VEs – graphical quality and sound – affected measures of task performance, workload, anxiety, and presence, as a way to test the effects of immersion and coherence on behavioral, psychological, affective, and physical fidelity (respectively). A mixed-methods approach was employed (Shorten and Smith, 2017), which used both quantitative and qualitative data to answer the research questions.

2. Methods

2.1. Participants

Twenty-nine (29) participants (M = 24.1 years, S.D. = 6.6 years, 9 female) were recruited in a convenience sample at the university campus. Volunteers were excluded if they were unable to use a keyboard with their hands, had a tendency for motion sickness, or had a history of Lasik eye surgery. The study was approved by the university Institutional Review Board.

2.2. Experimental setup

The experiment took place in a large room with a standing desk, computer monitor, and keyboard at its center. A Sawyer collaborative robot (Rethink Robotics, Boston, MA) was positioned at the opposite side of the desk facing the participants. It had a single arm with seven degrees of freedom controlled by a separate workstation using ROS (www.ros.org) and Python 2.7. Two virtual copies of this setup were

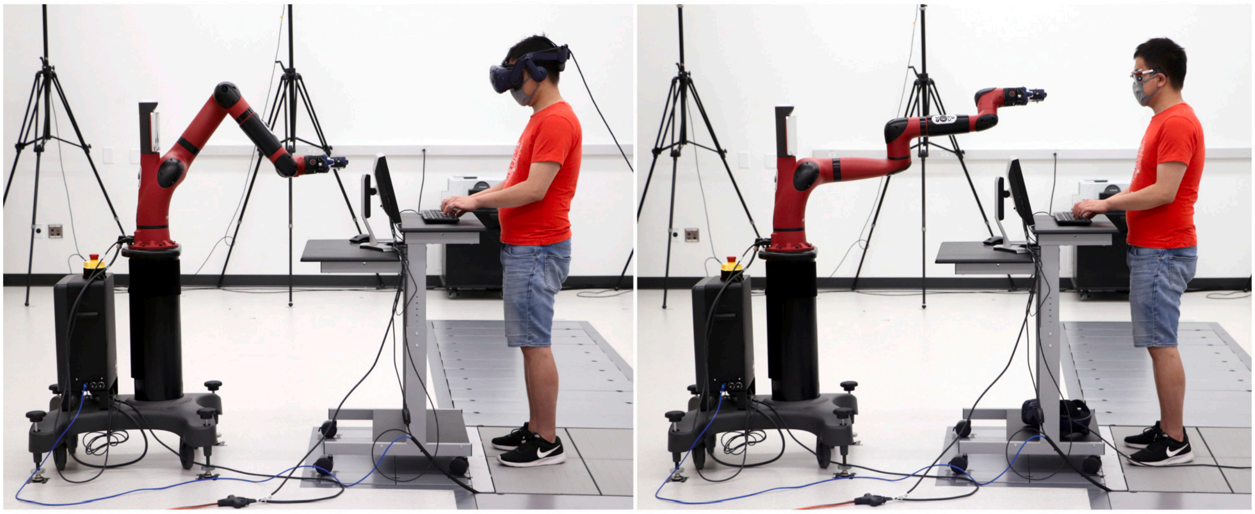

Fig. 1. Starting position with HMD (left), fully extended position without HMD (right). The physical robot did not move when the participants wore the HMD.

created in Unity (2019.4) and rendered on an HTC Vive Pro HMD with a Dell Precision 5810 workstation and an NVIDIA GeForce RTX 3080 Ti. Passive haptics were utilized as participants stood at this desk with their hands on the keyboard in both physical and virtual environments. They performed a visual search task as the robot performed a reach-and-retract motion (see Fig. 1).

2.3. Independent variables

2.3.1. Environment

Environment was a factor with three levels: the Physical environment and two VEs – VR-LiDAR, and VR-CAD – with high and low levels of graphical quality, respectively (Fig. 2). The two levels of graphical quality were chosen to maximize the difference in the Place Illusion component of Presence between the two VEs, with a photorealistic environment at one factor level, and a low-poly environment at another.

The Physical environment (Fig. 2, top) comprised the real-world lab space, with a standing desk, computer monitor, keyboard, and collaborative robot, as described previously.

The VR-LiDAR environment (Fig. 2, middle) was a point-cloud replica of the lab. It was created with a Light Detection and Ranging (LiDAR) scanner (Focus M, FARO, USA) at 43.7 Mpts resolution (6.1 mm at 10 m). It was based on a single scan taken from where the participants would be standing, at the 50th percentile standing eye height (Gordon, 1989). Post-processing of the scan was done in Meshlab (Cignoni et al., 2008) and rendered with the PCX package for Unity (github.com/keijiro/Pcx).

The VR-CAD environment (Fig. 2, bottom) was composed of low-poly parametric models of the lab and its equipment. It was created with Fusion 360 (Audodesk, USA) and Blender (Blender Foundation, Netherlands). The models were based on photographs, measurements, and the aforementioned LiDAR scan.

2.3.2. Motion

Motion was a factor with two levels: still and moving. The motion was a reach-and-retract motion similar to that of Kulić and Croft (2007) and Koppenborg et al. (2017), analogous to what might be used for material handling. The robot began from a starting position, extended its arm toward the participant, paused, and retracted to the starting position where it paused again before repeating (Fig. 1). The extension movement completed in 4.8 seconds; the retraction in 2.5 seconds. The maximum linear speed and acceleration were 600 mm/s and 600 mm/s^2, respectively, below the ISO 10218 standard. The pauses were uniformly and randomly distributed between 1 and 2 seconds. The motion of the real robot was recreated in the VEs using keyframe animation.

2.3.3. Sound

Sound was a factor in the VEs with two levels: off (silent) and on. When on, audio recordings of Sawyer's motors emanated from its arm segments as they moved. Audio recordings of Sawyer were made with an H4N stereo recorder (Zoom Corp., Japan).

2.4. Task

Both workload (Hancock et al., 2019) and attentional capture (Wolfe, 2010; Chan and Hayward, 2013) can be inferred via response time during a visual search task. The search task was adapted from PsyToolkit (Gijsbert Stoet) and was recreated in Unity so that it could be displayed on both the physical and virtual computer screens (Fig. 3). It is an example of conjunctive search, i.e. a search for a specific combination of stimulus features. For example, a cobot might perform material handling while a worker searches the workspace for a component of a particular shape and color, or inspects workpieces for defects of a certain size and orientation. Although the visual search used here is more abstract, it had the advantage of being reproducible in every environment, avoiding a potential confound.

There were 40 trials (stimulus presentations) in each run: 30 trials with the target and 10 catch trials without the target. Each trial contained 20 symbols randomly distributed on the search grid (Fig. 3). Participants stood at the standing desk with their hands on the keyboard and responded via a button press with their right hand ('J') if a target was present, and with their left hand ('F') if the target was absent. Large buttons were affixed to these keys so they could be easily located while wearing the HMD. Participants were told to respond as quickly as possible while maintaining maximum accuracy. Responses were recorded with the Unity Experimental Framework (Brookes et al., 2020). The same keyboard was used in all three environments to avoid another potential confound.

2.5. Dependent variables

2.5.1. Presence

Presence was associated with physical fidelity and was measured via the Slater-Usoh-Steed Questionnaire (SUS) (Usoh et al., 2000). The responses were averaged across all six items to obtain a mean score. The possible score ranged from 0-7.

2.5.2. Task performance

Task performance was associated with both behavioral and psychological fidelity and was defined as target-present mean correct response time for the visual search task. For each treatment condition, response times for correct "yes" responses were averaged across trials.

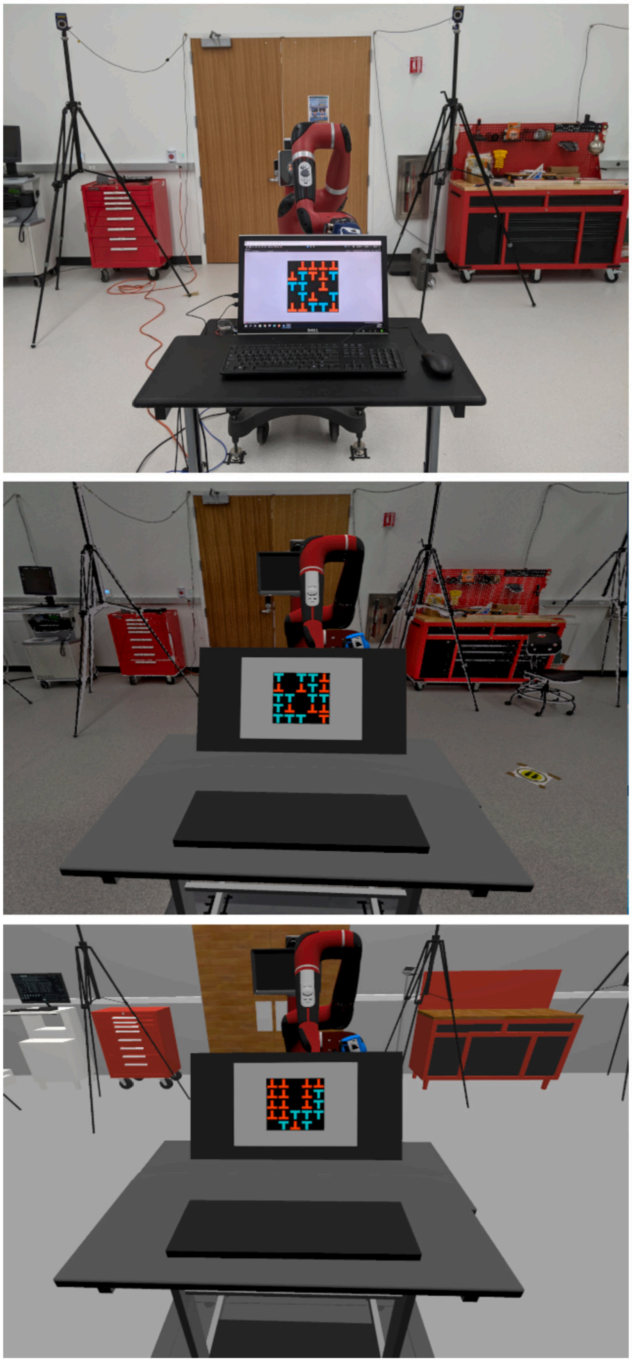

Fig. 2. Participants' first-person view in the physical environment (top), VR-LiDAR (middle), and VR-CAD (bottom).

2.5.3. *Workload*

Workload was associated with psychological fidelity and was measured via the "weighted workload score" (overall workload) of the NASA Task Load Index (TLX) (Hart and Staveland, 1988). The final score was the weighted average of the six subscales and ranged from 0-100.

2.5.4. *Anxiety*

Anxiety was associated with affective fidelity and was measured via the six-item form of the State-Trait Anxiety Inventory Inventory (STAI) (Tluczek et al., 2009). Participants responded to all six items on a scale of 1-4.

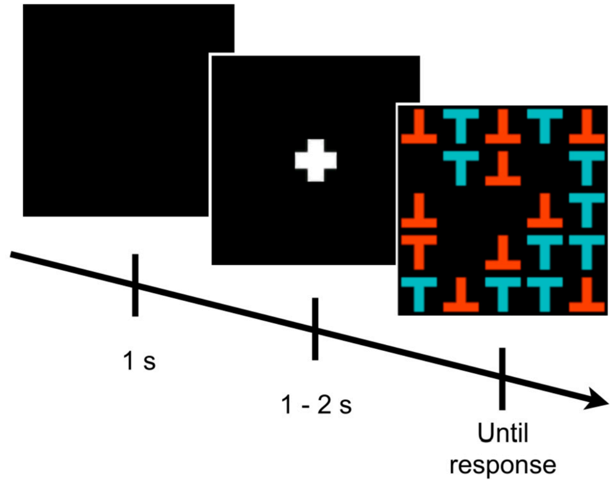

Fig. 3. Stimulus presentation timeline and example of a search grid with target (red 'T'). Each trial comprised a blank screen lasting 1 second followed by a fixation cross lasting 1 - 2 seconds (uniform random) followed by the search grid lasting until response.

Table 2
Initial questions of the semi-structured interview and their relation to quantitative measures, constructs, and types of fidelity in the framework suggested by Harris et al. (2021).

Type of Fidelity	Construct	Measure(s)	Initial Question
Physical	Presence	SUS	How could the VEs be made more realistic?
Behavioral	Performance	Response Time	Do you think you performed the same in all three environments? Why or why not?
Psychological	Attention, Workload	Response Time, TLX	What parts of the environments drew your attention away from the task?
Affective	Anxiety	STAI	How did the robot make you feel?

2.5.5. *Qualitative data*

Four semi-structured interview questions were asked at the end of the session (Table 2). Each open-ended question targeted a specific type of fidelity (e.g., affective fidelity), which was also assessed using a quantitative measure (e.g. the STAI). Since this was a semi-structured interview, follow-up questions were permitted to explore and understand how participants' experiences in VR were qualitatively different from those in the physical environment, and to help explain any observed differences in the quantitative measures. The interviews were recorded using a microphone (Blue Microphones, USA) and transcribed using Nvivo (Version 2020, QSR International, Australia).

2.6. *Procedure*

The experimental procedure is illustrated in Fig. 4. Participants were first briefed, which included obtaining informed consent and explaining the questionnaires and experimental task. Next, they performed one practice run in the physical environment. After the practice run, they were fitted with the HMD and the experimental runs began. Each experimental run lasted 4 minutes, during which response times were recorded. After each trial, participants completed the TLX, SUS, and STAI on a desktop computer, followed by a one-minute break. This process was repeated until all ten experimental runs were complete. A semi-structured interview was conducted at the end of the session. Participants were then debriefed and left the lab.

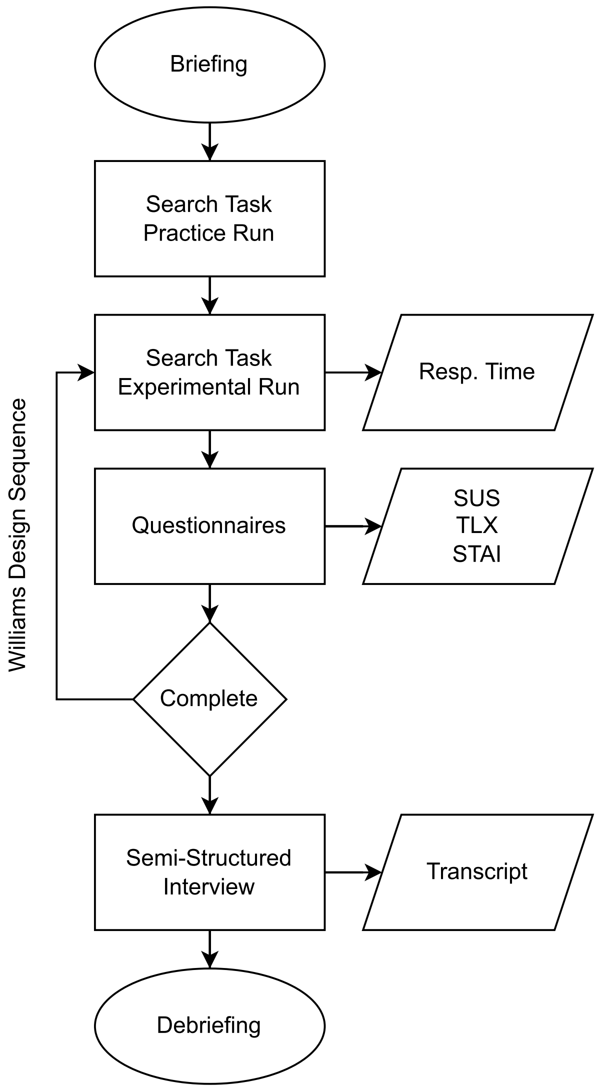

Fig. 4. Flow of each session. Breaks were provided after every run. The run sequence for each participant was determined with a Williams design (Latin square).

2.7. Experimental design and analysis

There were ten within-subject conditions (Fig. 5). In order to control for workload hysteresis (Hancock et al., 2019), the sequence of runs for each participant was determined using a Williams design (Williams, 1949), a Latin square balanced for first-order carryover effects.

Two generalized linear mixed effects models (GLMMs) were structured to address the research goals in Section 1.1. Model 1 was a 2 x 3 (Motion x Environment) which addressed RG1. Model 2 was a 2 x 2 x 2 (Motion x Environment x Sound) which addressed RG2. Because the robot could not move without making sound in the Physical environment, Model 1 included only the six conditions in which robot motion and sound were concurrently on or off (See Fig. 5). Model 2 included all eight virtual conditions but omitted the two physical conditions.

By-participant random intercepts were included to account for individual differences (Barr et al., 2013; Matuschek et al., 2017). Time was included as a covariate to account for fatigue and learning effects. Analysis of the SUS, response time, and TLX data was done using JMP Pro 16 (SAS Institute, Cary, NC) and lme4 (Bates et al., 2015) in R 4.2.0. Effect sizes *d* were calculated according to Brysbaert and Stevens (2018). Because model assumptions were not met for the STAI data, a cumulative logit mixed model (CLMM) was fit to the "I feel tense" item of the STAI using the *ordinal* package (Christensen, 2018). This item was chosen because the word "tense" was often used by participants during interviews to describe their emotional state around the robot, and "tense" was the only one of the three anxiety-present items ("tense", "upset", and "worried") that had a non-sparse contingency table.

A thematic analysis with three hierarchical code levels was performed following the guidelines of Guest et al. (2012); Braun and Clarke (2006); Nowell et al. (2017). The highest-level *structural* codes corresponded to the four types of fidelity – physical, behavioral, affective, and cognitive (Table 2). Three mid-level *conceptual* codes were determined *a priori* and corresponded to the variables of environment, motion, and time. One additional conceptual code (meta-knowledge) was created *a posteriori* during the coding process. Low-level codes, or *themes*, emerged from the data as the analysis progressed.

3. Results

3.1. Task performance

Output from Model 1 showed significant main effects of motion ($F_{1,139} = 11.38$, $p = 0.001$) and environment ($F_{2,139} = 8.26$, $p < 0.001$). The motion x environment interaction was not significant ($F_{2,2} = 0.33$, $p = 0.718$).

The mean correct response time (MCRT) when the robot was still was 1.080 ± 0.032 seconds (Fig. 6b). Participants responded 0.052 ± 0.016 s slower when the robot was moving, regardless of environment ($t_{139} = 3.37$, $p < 0.001$, $d = 1.45$). The MCRT in the physical environment was 1.148 ± 0.033 s. Participants responded 0.052 ± 0.019 seconds faster in the VR-LiDAR environment ($t_{139} = 2.70$, $p = 0.021$, $d = 1.43$) and 0.076 ± 0.019 seconds faster in the VR-CAD environment ($t_{139} = 3.98$, $p = 0.001$, $d = 2.10$) regardless of motion. The difference between the VR-CAD and VR-LiDAR environments was not statistically significant ($t_{139} = 1.28$, $p = 0.411$, $d = 0.67$).

Output from Model 2 showed a statistically significant simple effect of sound in the VR-CAD environment when the robot was still. The MCRT was 0.051 ± 0.022 s slower in response to audio cues ($t_{196} = 2.26$, $p = 0.025$, $d = 1.55$), keeping the other factor levels fixed. Themes relating to behavioral fidelity are shown in Table 4.

3.2. Workload

Output from Model 1 showed a statistically significant interaction effect of motion x environment ($F_{1,133} = 3.86$, $p = 0.024$) and a main effect of motion ($F_{1,133} = 15.57$, $p = 0.001$). The main effect of environment was not significant ($F_{2,133} = 1.801$, $p = 0.169$).

The mean TLX score in the physical environment when the robot was still was 32.99 ± 3.60 (Fig. 6c). When the robot was in motion, scores increased by 7.58 ± 1.92 in the physical environment ($t_{133} = 3.95$, $p = 0.002$, $d = 0.02$), but this effect was reduced by 5.70 ± 2.72 (75%) in the VR-LiDAR environment ($t_{133} = 2.10$, $p = 0.04$, $d = 0.02$) and 7.16 ± 2.73 (95%) in the VR-CAD environment ($t_{133} = 2.62$, $p = 0.010$, $d = 0.02$). Furthermore, the simple effect of motion was not significant in either the VR-LiDAR environment ($p = 0.33$, $d < 0.01$) or the VR-CAD environment ($p = 0.83$, $d < 0.01$).

Output from Model 2 did not reveal any statistically significant effects. Themes relating to psychological fidelity are shown in Tables 5 and 6.

3.3. Anxiety

Output from Model 1 revealed a main effect of motion ($\chi^2_1 = 10.31$, $p = 0.001$). The motion x environment interaction effect was not significant ($\chi^2_2 = 3.62$, $p = 0.163$), nor was the main effect of environment ($\chi^2_2 = 0.06$, $p = 0.973$).

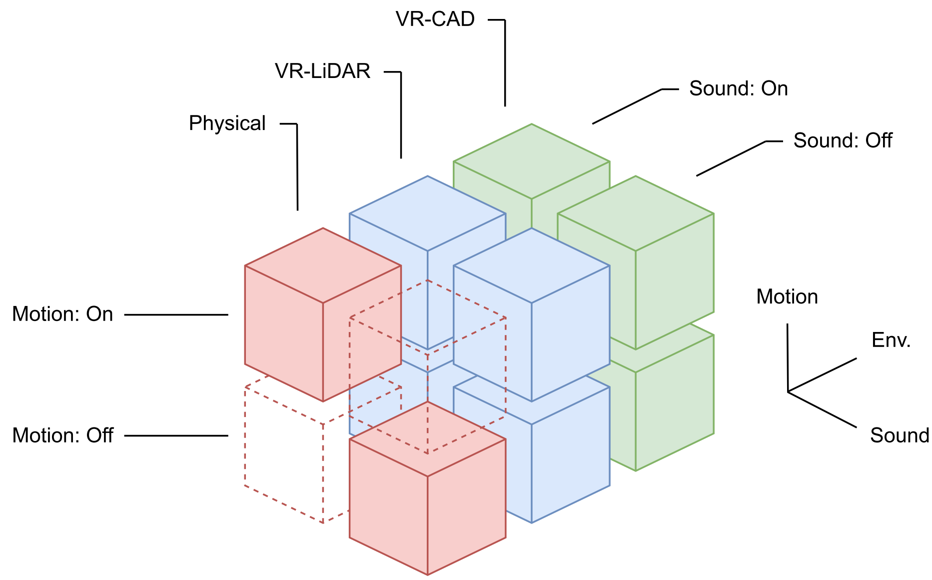

Fig. 5. Design space of the experiment. Note the absence of two impossible conditions in the Physical environment.

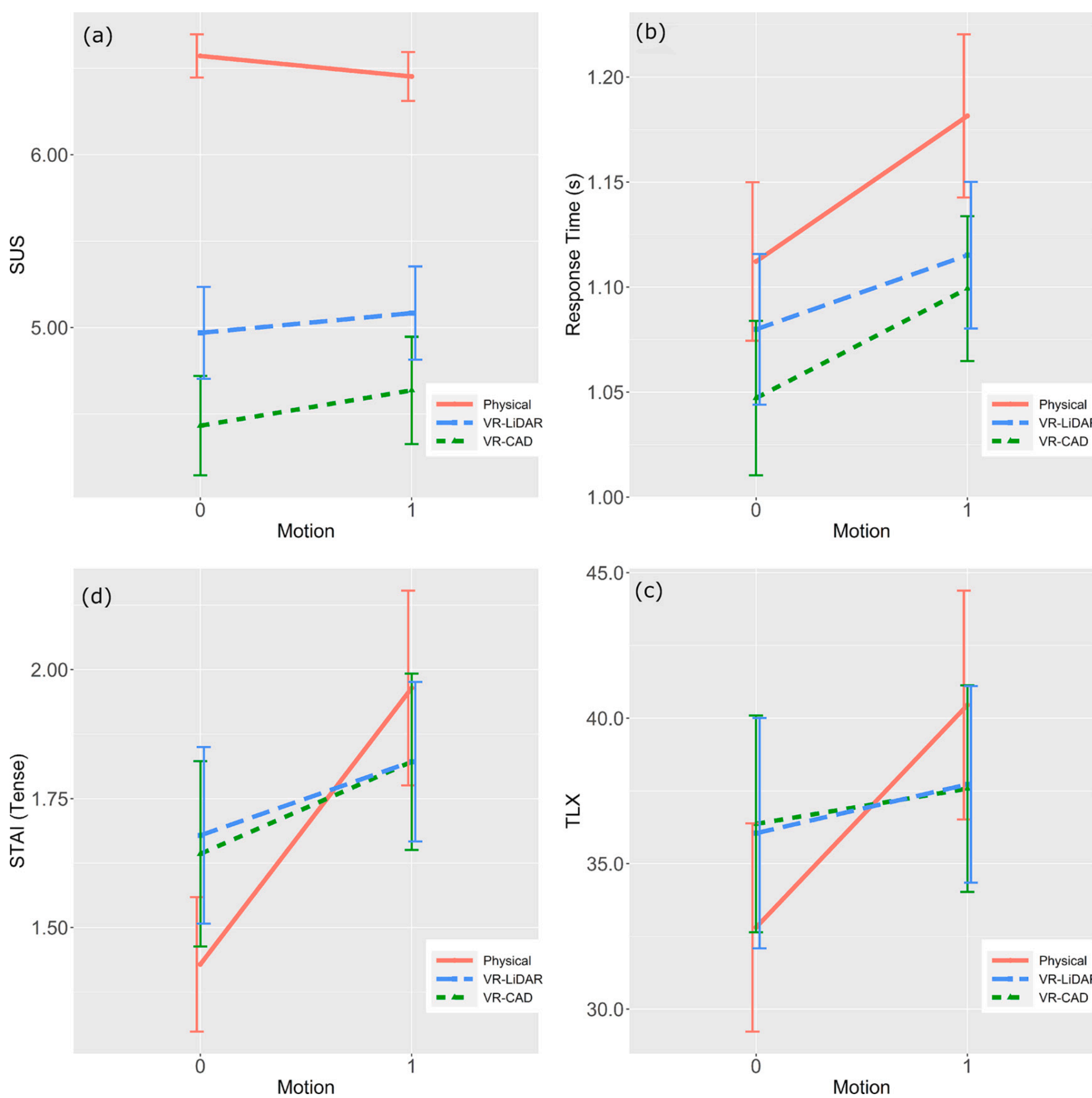

Fig. 6. The effect of robot-motion in different environments on: (a) presence, (b) task performance, (c) workload, and (d) anxiety. Error bars show ± 1 standard error.

The latent mean tense score in the Physical environment when the robot was still was 1.43 ± 0.13 (Fig. 6d). When the robot was in motion, the log-odds of obtaining a greater tense score increased by a factor of $\hat{\beta}_{Motion} = 2.14 \pm 0.66$ (Wald $z = 3.25$, $p = 0.001$), an 89.5% increase in probability. Despite the lack of a statistically significant motion x environment interaction, *post hoc* tests of latent mean scores showed that the simple effect of motion was not statistically significant in the VR-LiDAR environment (Wald $z = 1.15$, $p = 0.250$) or the VR-CAD environment (Wald $z = 1.07$, $p = 0.288$).

Output from Model 2 did not reveal any statistically significant effects. Themes relating to affective fidelity are shown in Table 7.

3.4. Presence

Output from Model 1 revealed a statistically significant main effect of environment ($F_{2,131.8} = 61.30$, $p < 0.0001$). The motion x environment interaction was not significant ($F_{2,131.8} = 0.47$, $p = 0.63$), nor was the main effect of motion ($F_{1,132.1} = 0.29$, $p = 0.59$).

The estimated mean SUS score in the physical environment was 6.51 ± 0.20 (Fig. 6a). Mean scores were 1.50 ± 0.19 points lower in the VR-LiDAR environment ($t_{131.8} = 8.00$, $p < 0.0001$, $d = 0.91$) and 1.99 ± 0.19 points lower in the VR-CAD environment ($t_{131.8} = 10.62$, $p < 0.0001$, $d = 1.20$) regardless of motion. The difference between the VR-LiDAR and VR-CAD environments was also statistically significant ($t_{131.8} = 2.61$, $p = 0.03$, $d = 0.30$)

Output from Model 2 showed a significant main effect of graphical detail ($t_{186} = 2.86$, $p = 0.005$, $d = 0.25$). The mean score in the VR-CAD environment was 4.36 ± 0.25. Scores were 0.59 ± 0.21 greater in VR-LiDAR, averaged across levels of motion and sound. The motion x sound interaction was similar in magnitude ($\hat{\beta} = 0.44 \pm 0.24$, $t_{186} = 1.87$, $p = 0.063$, $d = 0.19$), but was not statistically significant. However, *post hoc* contrasts did show a statistically significant simple effect of motion when sound was on ($t_{186} = 2.89$, $p = 0.022$, $d = 0.21$), with presence being 0.48 ± 0.17 greater when motion and sound co-occurred compared to when sounds were played but the robot was still.

Themes relating to physical fidelity are shown in Table 3.

3.5. Themes

Table 3
Count of participants (N) who expressed themes associated with physical fidelity "How could the VEs be made more realistic?"

Theme	N
Environment	18
Shadows in VR-LiDAR *"The white shadows threw me off [. . .] like this is not really real."*	6
Lack of detail in VR-CAD *"It feels kind of cartoonish."*	6
Embodiment and interaction *"Actually being able to see [my hands] would have made it feel more realistic for me."*	5
HMD tracking issues *"When I moved my head it just all kind of like distorted a little."*	5
Depth compression *"It felt like the space was more compacted than it was in real life."*	4
Sound localization *"I didn't like how the sounds didn't come from the same direction as the robot in the VR."*	4
Sound quality *"There was a pretty obvious difference between the headphones and hearing it in person."*	3
Color differences *"The colors were really bright in VR. Kind of like a fever dream or something."*	3

Table 4
Count of participants (N) who expressed themes associated with behavioral fidelity: "Do you think you performed the same in all three environments? Why or why not?"

Theme	N
Environment	**14**
Game-like quality of VR *"It was easier to do it virtually because it just feels like a game."*	5
Fewer distractions in VR *"[In VR] it's like all the distractions are kind of cut out."*	4
Sound *"I think [the sound] made it harder to do the task."*	3
Novelty *"Because of the novelty, I was taking every chance I could get to look up at [the real robot]."*	2
Motion	**3**
Motion per se *"I think I performed worse in the actual world when the robot was moving."*	3
Time	**2**
Practice *"I think over the course of the tests I got more used to the doing the task."*	2

Table 5
Count of participants (N) who expressed themes associated with psychological fidelity: "What parts of the environments drew your attention away from the task?"

Theme	N
Environment	**16**
Sound *"I think the sound drew my attention more than the visual aspects."*	11
Bright colors in VR *"I was just like really fascinated by the colors, just like pops."*	7
Details in VR-LiDAR *"In the more realistic one, just the amount of detail."*	5
Fewer distractors in VR *It felt clean and sharp, that's it. Less distracting.*	3
Motion	**14**
Proximity *It just felt like it was getting in my face and I wasn't able to concentrate.*	5
Motion toward *It moving towards me [. . .] definitely drew my attention.*	4
Novelty *That was very distracting because it is cool to look at.*	3
Kinematic qualities *The way [the robot] was moving probably drew most of my attention.*	2

4. Discussion

The results of this study suggest that the simple effect of robot motion was consistent across environments for measures of task performance and anxiety, meaning that the VEs were fidelitous in that sense. However, there were several statistically significant results which have important implications for the psychological, behavioral, and physical fidelity of VEs for HRI.

4.1. Effect of robot motion in the physical environment

In the physical environment, search task performance decreased in response to robot motion, while workload and anxiety increased. This suggests a reallocation of attentional resources away from the task and toward the moving robot via attentional capture, consistent with re-

Table 6
Count of participants (N) who expressed themes associated with psychological fidelity (cont'd): "What parts of the environments drew your attention away from the task?"

Theme	N
Meta knowledge	**13**
Looking for differences between environments *Once I started to notice the differences between the actual lab and VR, that would catch my attention.*	13
Time	**3**
Practice *As we kept going, I wasn't thinking about it as hard.*	3

Table 7
Count of participants (N) who expressed themes associated with affective fidelity: "How did the robot make you feel?"

Theme	N
Motion	**18**
Proximity *It felt like it was invading my space.*	13
Motion toward *I think it's just that it came straight towards me.*	6
Kinematic qualitites *It moves like a snake about to strike.*	5
Time	**15**
Predictability *[Knowing that it had one set pattern] helped reduce the feeling of risk and stress.*	11
Habituation *[It] just kind of phased out of my experience.*	4
Meta Knowledge	**11**
The real robot will not really hurt me *I wasn't worried because I remembered the [consent form] said it could not hurt you.*	4
VR is not real *The virtual one didn't really bother me because I knew it was virtual.*	4
Novelty *I haven't seen any robot arms before, so it was fun to look at.*	3
Environment	**8**
Sound *[It] sounded a little more threatening [in reality]. So [it] made me a little more tense.*	5
Pleasing aesthetics *I liked the designed environment because it felt aesthetically more pleasing.*	3
Warning coloration on robot *This may sound weird, but the color red obviously is more associated with danger.*	2

source theory (Hancock et al., 2019) and previous research using visual search tasks (Wolfe, 2010; Chan and Hayward, 2013). The motion of the robot, its sound, proximity, direction straight toward the body, and unsettling animal-like kinematic qualities were identified as themes associated with both anxiety and workload (Tables 5, 6, and 7). These are consistent with previous findings that physiological stress increases with robot speed (Arai et al., 2010; Fujita et al., 2010) and that straight-line reaching trajectories are deemed more unsafe (Kulić and Croft, 2007) and stressful than more circuitous motion plans (Fujita et al., 2010; Dehais et al., 2011).

4.2. Behavioral fidelity

Task performance was better overall in the VR-CAD environment compared to the physical environment. The themes in Table 4 suggest three possible explanations: the relative lack of distractions in VR, increased physiological arousal, or an HMD-specific neurological phenomenon.

Participants cited the lack of environmental distractions as a reason for better performance in VR: "it's like all the distractions are kind of cut out" (Table 4). The VR-CAD environment in particular was less detailed and contained fewer visual cues of any kind: "it feels kind of cartoonish" (Table 3). Performance enhancement due to the lack of distractions is consistent with work in diminished reality (Murph et al., 2021), a form of augmented reality which eliminates aspects of the physical environment to promote focus on specific tasks (Richardson et al., 2021). It has been used to augment the attention of individuals with autism (Yantaç et al., 2015), and to improve attention and memory during video conferencing (Yao et al., 2013). Future work in VR might investigate the role of visual clutter on the fidelity of performance-based measures of attention allocation.

The themes of novelty and game-like qualities of VR (Tables 4 and 5) suggest that participants may have been more excited or aroused when they were wearing the HMD. Increased arousal can increase task performance up to a point, as described by the Yerkes-Dodson law (Yerkes and Dodson, 1908). Although the present study did not measure arousal, HMD-based VR has been shown to increase self-reported anger and sexual arousal compared to computer monitors (Magdin et al., 2021; Elsey et al., 2019; Milani et al., 2022; Simon and Greitemeyer, 2019). However, a recent study Li et al. (2020) found that both neurological and performance-based measures of selective attention were enhanced in VR, but there was no difference in neurological measures of arousal. The results of that study, as well as the themes of color in Tables 3 and 5 ("the colors were really bright in VR. Kind of like a fever dream") may suggest that faster response times were the result of the optical characteristics of the HMD displays and their effect on the human nervous system, rather than increased physiological arousal due to a novel and game-like VR experience. Future work might be done to investigate this possibility.

4.2.1. The role of sound

The effect of sound on task performance from Model 2 in Section 3.1 suggests that motion-related performance decrements were associated with the sound of the robot's motors. This is consistent with previous research showing that auditory distractors can capture attention during visual search tasks (Dalton and Hughes, 2014; Tellinghuisen and Nowak, 2003), and with themes in Tables 4, 5, and 7 showing that the robot's sound made the task more difficult, captured attention, and felt threatening. This implies that care should be taken to replicate the psychometric properties of the auditory aspects of the physical environment even if the primary task is visual, and that the sound design of robotic systems may play an important role in the performance of human-robot teams.

4.3. Psychological fidelity

The effect of motion on TLX scores was significantly lower in both the VEs compared to the physical environment. One possible interpretation is that workload did not increase with robot motion in VR. Another is that workload did increase in VR, but that the TLX dissociated from task performance in an example of *workload insensitivity* (Hancock et al., 2019). This latter interpretation is supported by the response time results and the qualitative results: The lack of a significant motion x environment interaction effect on response time suggests that the deleterious effect of motion was consistent across environments, implying that attentional capture occurred in all environments (Olk et al., 2018; Parsons et al., 2013). The themes in Tables 4, 5, and 6 also suggest that explicit attentional capture occurred in the virtual environments. Thus, based on the expectations of resource theory, one would expect subjective workload to have increased with motion in the VEs as it did in the physical environment, but it did not. Because the TLX is the gold standard for measuring mental workload in VR (Reinhardt et al., 2019), the possibility that it can become insensitive under some circumstances has serious implications for the use of virtual environments as testbeds for HRI research and perhaps human factors research in general.

The themes of “game-like quality” (Table 5) and “pleasing aesthetics” (Table 7) may point to a connection to hedonomics (Hancock et al., 2005), or the pleasurable aspects of the VR experience. Participants stated that “In the real world it felt like a task, but in VR it was like a game”, that the robot was “fun to look at”, and that VR was “aesthetically more pleasing”. This may have made performing the task in the face of distraction by the moving robot less a source of mental demand and effort, and instead a source of fascination and fun, rendering the TLX insensitive while maintaining the sensitivity of the response time measure.

Future work could confirm the possible workload insensitivity observed here and determine its cause. Individual differences in VR experience and the “wow factor” may be worth investigating using a mixed-methods approach including additional physiological and neurological indicators of mental workload.

4.4. Affective fidelity

The results from Model 1 in Section 3.3 suggest that the effect of motion on anxiety was not different across environments. Participants said that the proximity of the moving robot made it feel like it was invading their space (Table 7), but that the virtual robot did not bother them because they knew it was not real (Table 7). However, at least two participants thought the real robot was moving when they were wearing the HMD, which may have attenuated the motion x environment interaction effect. Future experimenters should be sure to tell participants what will occur in the physical environment while they are immersed in VR.

4.5. Physical fidelity

The significant effect of graphical quality on presence from Model 2 in Section 3.4 suggests that the SUS was sensitive to changes in place illusion via manipulations of immersion. This seems to contradict Skarbez et al. (2021), who found that the SUS did not respond to increased immersion as a main effect on place illusion. This discrepancy may have resulted from how immersion was manipulated in the two studies: via the graphical detail in this study, and via field-of-view, passive haptics, and sound cues in Skarbez et al. (2021).

The *post hoc* contrasts from Model 2 in Section 3.4 also showed that, in VR, greater presence was elicited when sound effects accompanied robot motion compared to when sound effects were played but the robot was clearly still (Section 3.4). One possible interpretation is that in the first case motion and sound were coherent but in the second case they were not, because in the physical environment motion always occurs with sound and *vice versa*. Past research has also found that presence increased when visual cues were accompanied by coherent audio cues (Poeschl et al., 2013; Hoppe et al., 2019), but decreased when audio cues were heard despite clear visual evidence to the contrary (Slater et al., 2022a,b). An ergonomic application of this might be a virtual factory, programmed such that ambient noises are heard but the motions of machinery that cause those sounds are not animated, perhaps due to time or budgetary constraints. Such a condition might actually reduce presence. So if presence is important, designers might consider omitting such sounds, or including some sort of easy-to-animate motion to indicate their physical origin.

Again, the results of this study seem to contradict Skarbez et al. (2021), who found that the SUS did not respond to manipulations of coherence. This may be because coherence was manipulated by the congruence of motion and sound cues in the present study, but by physics and narrative coherence in Skarbez et al. (2021). Taken together, this supports the idea of coherence as represented by a mathematical vector whose important elements are still largely undiscovered (Skarbez et al., 2021). It may be that the congruence of motion and sound is one such important element, whereas physics and narrative coherence are not.

4.6. Limitations and future work

There several limitations to this study. First, the conjunctive search task (Section 2.4) had the advantage of being a well studied paradigm that could be reproduced in every environment, but it was rather abstract. Future work might utilize more ecologically valid stimuli or tasks based on more realistic HRI applications.

Second, our sample comprised mostly university students, so it is probably not representative of the demographic who would be working with collaborative robots in real manufacturing environments, and certainly not representative of the working population as a whole. Future work might use a more representative sample.

Third, creating a virtual environment based on multiple co-aligned LiDAR scans was challenging, and therefore a single scan taken at eye level was ultimately employed (see Section 2). It resulted in “shadows” behind every object due to occlusion. These made the VR-LiDAR environment less realistic (Table 3) and served as distractions (Table 6). A more viable approach might have been to apply photorealistic textures to the models created for the VR-CAD environment. In addition, the point-cloud renderings in the VR-LiDAR environment were only believable for background objects. Because of this, the desk, monitor, keyboard, and robot were rendered in the VR-LiDAR environment just as they were in the VR-CAD environment. This may have reduced the SUS scores for the VR-LiDAR environment.

Lastly, the measures required to fully address fidelity are more numerous and complex than those used here. Future work might explore a single type of fidelity more thoroughly by triangulating constructs via subjective, physiological, and performance-based measures.

4.7. Conclusions

The results demonstrate a possible example of workload insensitivity in VR, with potentially serious implications for the ability of VEs to recreate real-world changes in subjective workload associated with the motion of nearby robots. The results also corroborate the ability of HMD-based VR to enhance attention, the practical implications of which are worth exploring. Additionally, presence responded to manipulations of both immersion and coherence, with implications for the theory of presence. Finally, the performance decrements associated with robot motion were likely due to the sound of the robot’s motors, with potential implications for the performance of human-robot teams.

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.

Acknowledgements

This research was supported in part by the National Science Foundation (award IIS-1850055) and the National Institute of Occupational Safety and Health (award T42OH008673).

References

Arai, T., Kato, R., Fujita, M., 2010. Assessment of operator stress induced by robot collaboration in assembly. CIRP Ann. 59 (1), 5–8. https://doi.org/10.1016/j.cirp.2010.03.043.

Araújo, D., Davids, K., Passos, P., 2007. Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: comment on Rogers, Kadar, and Costall (2005). Ecol. Psychol. 9 (1), 69–78. https://doi.org/10.1080/10407410709336951.

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68 (3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1). https://doi.org/10.18637/jss.v067.i01. arXiv:1406.5823.

Bowman, D.A., McMahan, R.P., 2007. Virtual reality: how much immersion is enough? Computer 40 (7), 36–43. https://doi.org/10.1109/MC.2007.257.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., Mushtaq, F., 2020. Studying human behavior with virtual reality: the unity experiment framework. Behav. Res. Methods 52, 455–463. https://doi.org/10.1101/459339.

Brysbaert, M., Stevens, M., 2018. Power analysis and effect size in mixed effects models: a tutorial. J. Cogn. 1 (1), 1–20. https://doi.org/10.5334/joc.10.

Chan, L.K.H., Hayward, W.G., 2013. Visual search. Cogn. Sci. 4 (4), 415–429. https://doi.org/10.1002/wcs.1235.

Chen, J., Xie, Z., Or, C., 2021. Effectiveness of immersive virtual reality-supported interventions for patients with disorders or impairments: a systematic review and meta-analysis. Health Technol. 11 (4), 811–833. https://doi.org/10.1007/s12553-021-00561-7.

Christensen, R., 2018. Cumulative link models for ordinal regression with the R package ordinal. http://cran.uni-muenster.de/web/packages/ordinal/vignettes/clm_article.pdf.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., 2008. MeshLab: an open-source mesh processing tool. In: Scarano, V., Chiara, R.D., Erra, U. (Eds.), Eurographics Italian Chapter Conference.

Dalton, P., Hughes, R.W., 2014. Auditory attentional capture: implicit and explicit approaches. Psychol. Res. 78 (3), 313–320. https://doi.org/10.1007/s00426-014-0557-5.

Dehais, F., Sisbot, E.A., Alami, R., Causse, M., 2011. Physiological and subjective evaluation of a human-robot object hand-over task. Appl. Ergon. 42 (6), 785–791. https://doi.org/10.1016/j.apergo.2010.12.005.

Duffy, V.G., Or, C.K., Lau, V.W., 2006. Perception of safe robot speed in virtual and real industrial environments. Hum. Factors Ergon. Manuf. 16 (4), 369–383. https://doi.org/10.1002/hfm.20058.

Elsey, J.W., van Andel, K., Kater, R.B., Reints, I.M., Spiering, M., 2019. The impact of virtual reality versus 2D pornography on sexual arousal and presence. Comput. Hum. Behav. 97, 35–43. https://doi.org/10.1016/j.chb.2019.02.031.

Fisher, D.L., Rizzo, M., Caird, J., Lee, J.D. (Eds.), 2011. Handbook of Driving Simulation for Engineering, Medicine, and Psychology.

Fratczak, P., Goh, Y.M., Kinnell, P., Justham, L., Soltoggio, A., 2021. Robot apology as a post-accident trust-recovery control strategy in industrial human-robot interaction. Int. J. Ind. Ergon. 82, 103078. https://doi.org/10.1016/j.ergon.2020.103078.

Fratczak, P., Goh, Y.M., Kinnell, P., Soltoggio, A., Justham, L., 2019. Understanding human behaviour in industrial human-robot interaction by means of virtual reality. In: ACM International Conference Proceeding Series.

Fujita, M., Kato, R., Arai, T., 2010. Motion of Industrial Robot Manipulator, pp. 361–366.

Gordon, C.C., 1989. 1988 anthropometric survey of U.S. personnel: summary statistics interim report. Retrieved from https://multisite.eos.ncsu.edu/www-ergocenter-ncsu-edu/wp-content/uploads/sites/18/2016/06/Anthropometric-Detailed-Data-Tables.pdf.

Gualtieri, L., Rauch, E., Vidoni, R., 2021. Emerging research fields in safety and ergonomics in industrial collaborative robotics: a systematic literature review. Robot. Comput.-Integr. Manuf. 67, 101998. https://doi.org/10.1016/j.rcim.2020.101998.

Guest, G., MacQueen, K., Namey, E., 2012. Writing up thematic analyses. In: Applied Thematic Analysis, pp. 241–278.

Hancock, P.A., Matthews, G., 2019. Workload and performance: associations, insensitivities, and dissociations. Hum. Factors 61 (3), 374–392. https://doi.org/10.1177/0018720818809590.

Hancock, P.A., Pepe, A.A., Murphy, L.L., 2005. Hedonomics: the power of positive and pleasurable ergonomics. Ergon. Des. 13 (1), 8–14. https://doi.org/10.1177/106480460501300104.

Harris, D.J., Bird, J.M., Smart, P.A., Wilson, M.R., Vine, S.J., 2020. A framework for the testing and validation of simulated environments in experimentation and training. Front. Psychol. 11, 605.

Harris, D.J., Hardcastle, K.J., Wilson, M.R., Vine, S.J., 2021. Assessing the learning and transfer of gaze behaviours in immersive virtual reality. Virtual Real. 25, 961–973. https://doi.org/10.1007/s10055-021-00501-w.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52 (100), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hoareau, C., Querrec, R., Buche, C., Ganier, F., 2017. Evaluation of internal and external validity of a virtual environment for learning a long procedure. Int. J. Hum.-Comput. Interact. 33 (10), 786–798. https://doi.org/10.1080/10447318.2017.1286768.

Hoppe, M., Karolus, J., Dietz, F., Woźniak, P.W., Schmidt, A., Machulla, T., 2019. VRsneaky: increasing presence in VR through gait-aware auditory feedback. In: Conference on Human Factors in Computing Systems - Proceedings, pp. 1–9.

Jerald, J., 2016. The VR Book: Human Centered Design for Virtual Reality. Association for Computing Machinery.

Kahneman, D., 1973. Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ.

Kamide, H., Mae, Y., Takubo, T., Ohara, K., Arai, T., 2014. Direct comparison of psychological evaluation between virtual and real humanoids: personal space and subjective impressions. Int. J. Hum.-Comput. Stud. 72 (5), 451–459. https://doi.org/10.1016/j.ijhcs.2014.01.004.

Kim, A.J., Anderson, B.A., 2020. Threat reduces value-driven but not salience-driven attentional capture. Emotion 20 (5), 874–889. https://doi.org/10.1037/emo0000599.supp.

Koppenborg, M., Nickel, P., Naber, B., Lungfiel, A., Huelke, M., 2017. Effects of movement speed and predictability in human–robot collaboration. Hum. Factors Ergon. Manuf. 27 (4), 197–209. https://doi.org/10.1002/hfm.20703.

Kulić, D., Croft, E., 2007. Physiological and subjective responses to articulated robot motion. Robotica 25 (1), 13–27. https://doi.org/10.1017/S0263574706002955.

Li, G., Anguera, J.A., Javed, S.V., Khan, M.A., Wang, G., Gazzaley, A., 2020. Enhanced attention using head-mounted virtual reality. J. Cogn. Neurosci. 32 (8), 1438–1454. https://doi.org/10.1162/jocn_a_01560.

Li, R., Van Almkerk, M., Van Waveren, S., Carter, E., Leite, I., 2019. Comparing human-robot proxemics between virtual reality and the real world. In: ACM/IEEE International Conference on Human-Robot Interaction. 2019-March, pp. 431–439.

Magdin, M., Balogh, Z., Reichel, J., Francisti, J., Koprda, Š., György, M., 2021. Automatic detection and classification of emotional states in virtual reality and standard environments (LCD): comparing valence and arousal of induced emotions. Virtual Real. 25 (4), 1029–1041. https://doi.org/10.1007/s10055-021-00506-5.

Matsas, E., Batras, D., Vosniakos, G.-C., 2013. Beware of the robot: a highly interactive and immersive virtual reality training application in robotic manufacturing systems. In: Emmanouilidis, C., Taisch, M., Kiritsis, D. (Eds.), Advances in Production Management Systems: Competitive Manufacturing for Innovative Products and Services, AMPS 2012, pt I. In: IFIP Advances in Information and Communication Technology, vol. 397, pp. 606–613. IFIP WG 5 7 Advances Prod Management Syst; ATHENA Res & Innovat Ctr, Informat, Commun & Knowledge Technologies; Hellen Maintenance Soc.

Matsas, E., Vosniakos, G.C., Batras, D., 2018. Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality. Robot. Comput.-Integr. Manuf. 50, 168–180. https://doi.org/10.1016/j.rcim.2017.09.005.

Matsas, E., Vosniakos, G.-C., Batras, D., 2016. Modelling simple human-robot collaborative manufacturing tasks in interactive virtual environments. In: VRIC'16: Proceedings of the 2016 Virtual Reality International Conference.

Matsas, E., Vosniakos, G.-C., 2017. Design of a virtual reality training system for human-robot collaboration in manufacturing tasks. Int. J. Interact. Des. Manuf. 11 (2), 139–153. https://doi.org/10.1007/s12008-015-0259-2.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D., 2017. Balancing type I error and power in linear mixed models. J. Mem. Lang. 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001. arXiv:1511.01864.

Ménétrier, E., Richard, P., Boucher, V., Boujon, C., 2017. Using virtual reality techniques to study cognitive processes in car driving activity. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017) - HUCAPP, pp. 138–145.

Milani, S., Jabs, F., Brown, N.B., Zdaniuk, B., Kingstone, A., Brotto, L.A., 2022. Virtual reality erotica: exploring general presence, sexual presence, sexual arousal, and sexual desire in women. Arch. Sex. Behav. 51 (1), 565–576. https://doi.org/10.1007/s10508-021-02111-8.

Murph, I., McDonald, M., Richardson, K., Wilkinson, M., Robertson, S., Karunakaran, A., et al., 2021. Diminishing reality: potential benefits and risks. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 65 (1), 164–168. https://doi.org/10.1177/1071181321651103.

Norman, D.A., Bobrow, D.G., 1975. On data-limited and resource-limited processes. Cogn. Psychol. 7 (1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3.

Nowell, L.S., Norris, J.M., White, D.E., Moules, N.J., 2017. Thematic analysis: striving to meet the trustworthiness criteria. Int. J. Qual. Methods 16 (1), 1–13. https://doi.org/10.1177/1609406917733847.

Olk, B., Dinu, A., Zielinski, D.J., Kopper, R., 2018. Measuring visual search and distraction in immersive virtual reality. R. Soc. Open Sci. 5 (5). https://doi.org/10.1098/rsos.172331.

Oyekan, J.O., Hutabarat, W., Tiwari, A., Grech, R., Aung, M.H., Mariani, M.P., et al., 2019. The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans. Robot. Comput.-Integr. Manuf. 55, 41–54. https://doi.org/10.1016/j.rcim.2018.07.006.

Paljic, A., 2017. Ecological Validity of Virtual Reality: Three Use Cases. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10590, pp. 301–310.

Parsons, T.D., Courtney, C.G., Dawson, M.E., 2013. Virtual reality Stroop task for assessment of supervisory attentional processing. J. Clin. Exp. Neuropsychol. 35 (8), 812–826. https://doi.org/10.1080/13803395.2013.824556.

Pérez, L., Diez, E., Usamentiaga, R., García, D.F., 2019. Industrial robot control and operator training using virtual reality interfaces. Comput. Ind. 109, 114–120. https://doi.org/10.1016/j.compind.2019.05.001.

Poeschl, S., Wall, K., Doering, N., 2013. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In: Proceedings - IEEE Virtual Reality, pp. 129–130.

Porter, B., 2011. Handbook of Traffic Psychology. Academic Press, London, United Kingdom.

Ranney, T.A., 2011. Psychological fidelity: perception of risk. In: Fisher, D.L., Rizzo, M., Caird, J., Lee, J.D. (Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. https://doi.org/10.1201/b10836.

Reinhardt, D., Haesler, S., Hurtienne, J., Wienrich, C., 2019. Entropy of controller movements reflects mental workload in virtual reality. In: 2019 26th IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE; IEEE Comp Soc; IEEE Comp Soc Visualizat & Graph Tech Comm, pp. 802–808.

Richardson, K., Collins McLaughlin, A., McDonald, M., Crowson, A., 2021. The effects of diminished reality on the detection of and response to notifications. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 65 (1), 159–163. https://doi.org/10.1177/1071181321651236.

Shorten, A., Smith, J., 2017. Mixed methods research: expanding the evidence base. Evid.-Based Nurs. 20 (3), 74–75. https://doi.org/10.1136/eb-2017-102699.

Shu, B., Sziebig, G., Pieters, R., 2019. Architecture for safe human-robot collaboration: multi-modal communication in virtual reality for efficient task execution. In: IEEE International Symposium on Industrial Electronics. 2019-June, pp. 2297–2302.

Sievers, T.S., Schmitt, B., Rückert, P., Petersen, M., Tracht, K., 2020. Concept of a mixed-reality learning environment for collaborative robotics. Proc. Manuf. 45 (2019), 19–24. https://doi.org/10.1016/j.promfg.2020.04.034.

Simon, S.C., Greitemeyer, T., 2019. The impact of immersion on the perception of pornography: a virtual reality study. Comput. Hum. Behav. 93, 141–148. https://doi.org/10.1016/j.chb.2018.12.018.

Simons, D.J., 2000. Attentional capture and inattentional blindness. Trends Cogn. Sci. 4 (4), 147–155. https://doi.org/10.1016/S1364-6613(00)01455-8.

Skarbez, R., Brooks, F.P., Whitton, M.C., 2021. Immersion and coherence: research agenda and early results. IEEE Trans. Vis. Comput. Graph. 27 (10), 3839–3850. https://doi.org/10.1109/TVCG.2020.2983701.

Slater, M., 2018. Immersion and the illusion of presence in virtual reality. Br. J. Psychol. 109 (3), 431–433. https://doi.org/10.1111/bjop.12305.

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., Oliva, R., 2022a. A separate reality: an update on place illusion and plausibility in virtual reality. Front. Virtual Real. 3, 1–16. https://doi.org/10.3389/frvir.2022.914392.

Slater, M., Cabriera, C., Senel, G., Banakou, D., Beacco, A., Oliva, R., Gallego, J., 2022b. The sentiment of a virtual rock concert. Virtual Real. https://doi.org/10.1007/s10055-022-00685-9.

Slater, M., Sanchez-Vives, M.V., 2016. Enhancing our lives with immersive virtual reality. Front. Robot. AI, 3. https://doi.org/10.3389/frobt.2016.00074.

Tellinghuisen, D.J., Nowak, E.J., 2003. The inability to ignore auditory distractors as a function of visual task perceptual load. Percept. Psychophys. 65 (5), 817–828. https://doi.org/10.3758/BF03194817.

Tluczek, A., Henriques, J.B., Brown, R.L., 2009. Support for the reliability and validity of a six-item state anxiety scale derived from the state-trait anxiety inventory. J. Nurs. Meas. 17 (1), 19–28. https://doi.org/10.1891/1061-3749.17.1.19.

Usoh, M., Catena, E., Arman, S., Slater, M., 2000. Using presence questionnaires in reality. Presence 9 (5), 497–503. https://doi.org/10.1145/3290605.3300590.

van der Ham, I.J.M., Faber, A.M.E., Venselaar, M., van Kreveld, M.J., Löffler, M., 2015. Ecological validity of virtual environments to assess human navigation ability. Front. Psychol. 6, 1–6. https://doi.org/10.3389/fpsyg.2015.00637.

Weiner, E.J., Sanchez, D.R., 2020. Cognitive ability in virtual reality: validity evidence for VR game-based assessments. Int. J. Sel. Assess. 28 (3), 215–235. https://doi.org/10.1111/ijsa.12295.

Weistroffer, V., Paljic, A., Fuchs, P., Hugues, O., Chodacki, J.P., Ligot, P., Morais, A., 2014. Assessing the acceptability of human-robot co-presence on assembly lines: a comparison between actual situations and their virtual reality counterparts. In: IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions, pp. 377–384.

Williams, E.J., 1949. Experimental designs balanced for the estimation of residual effects of treatments. Aust. J. Chem. 2 (2), 149–168. https://doi.org/10.1071/CH9490149.

Wolfe, J.M., 2010. Visual search. Curr. Biol. 20 (8), 346–349. https://doi.org/10.1016/j.cub.2010.02.016.

Yantaç, A.E., Çorlu, D., Fjeld, M., Kunz, A., 2015. Exploring diminished reality (DR) spaces to augment the attention of individuals with autism. In: Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality Workshops. ISMARW 2015, pp. 68–73.

Yao, L., Devincenzi, A., Pereira, A., Ishii, H., 2013. FocalSpace: multimodal activity tracking, synthetic blur and adaptive presentation for video conferencing. In: SUI 2013 - Proceedings of the ACM Symposium on Spatial User Interaction, pp. 73–76.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol. 18, 459–482. http://psychclassics.yorku.ca/Yerkes/Law/.

You, S., Kim, J.H., Lee, S.H., Kamat, V., Robert, L.P., 2018. Enhancing perceived safety in human–robot collaborative construction using immersive virtual environments. Autom. Constr. 96, 161–170. https://doi.org/10.1016/j.autcon.2018.09.008.

Young, M.S., Stanton, N.A., 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Hum. Factors 44 (3), 365–375. https://doi.org/10.1518/0018720024497709.

Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. https://doi.org/10.1080/00140139.2014.956151.