

# VIRTUAL AND AUGMENTED REALITY METHODS IN NEUROSCIENCE AND NEUROPATHOLOGY

EDITED BY: Valerio Rizzo, Thomas D. Parsons and Pietro Cipresso  
PUBLISHED IN: Frontiers in Human Neuroscience



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-299-9

DOI 10.3389/978-2-88966-299-9

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# VIRTUAL AND AUGMENTED REALITY METHODS IN NEUROSCIENCE AND NEUROPATHOLOGY

Topic Editors:

**Valerio Rizzo**, University of Palermo, Italy

**Thomas D. Parsons**, University of North Texas, United States

**Pietro Cipresso**, Istituto Auxologico Italiano (IRCCS), Italy

**Citation:** Rizzo, V., Parsons, T. D., Cipresso, P., eds. (2020). Virtual and Augmented Reality methods in Neuroscience and Neuropathology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-299-9

# Table of Contents

- 05** *Immersive Virtual Reality as an Adjunctive Non-opioid Analgesic for Pre-dominantly Latin American Children With Large Severe Burn Wounds During Burn Wound Cleaning in the Intensive Care Unit: A Pilot Study*  
Hunter G. Hoffman, Robert A. Rodriguez, Miriam Gonzalez, Mary Bernardy, Raquel Peña, Wanda Beck, David R. Patterson and Walter J. Meyer III
- 16** *Technological Competence is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-Analysis*  
Panagiotis Kourtesis, Simona Collina, Leonidas A. A. Doumas and Sarah E. MacPherson
- 33** *The Dynamics of Attention Shifts Among Concurrent Speech in a Naturalistic Multi-speaker Virtual Environment*  
Keren Shavit-Cohen and Elana Zion Golumbic
- 45** *Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology*  
Panagiotis Kourtesis, Simona Collina, Leonidas A. A. Doumas and Sarah E. MacPherson
- 58** *Virtual Reality Analgesia With Interactive Eye Tracking During Brief Thermal Pain Stimuli: A Randomized Controlled Trial (Crossover Design)*  
Najood A. Al-Ghamdi, Walter J. Meyer III, Barbara Atzori, Wadee Alhalabi, Clayton C. Seibel, David Ullman and Hunter G. Hoffman
- 69** *The Adoption of New Treatment Modalities by Health Professionals and the Relative Weight of Empirical Evidence in Favor of Virtual Reality Exposure Versus Mindfulness in the Treatment of Anxiety Disorders*  
Kevin Nolet, Giulia Corno and Stéphane Bouchard
- 80** *Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis*  
Dimitrios Saredakis, Ancret Szpak, Brandon Birkhead, Hannah A. D. Keage, Albert Rizzo and Tobias Loetscher
- 97** *Application of Supervised Machine Learning for Behavioral Biomarkers of Autism Spectrum Disorder Based on Electrodermal Activity and Virtual Reality*  
Mariano Alcañiz Raya, Irene Alice Chicchi Giglioli, Javier Marín-Morales, Juan L. Higuera-Trujillo, Elena Olmos, Maria E. Minissi, Gonzalo Teruel Garcia, Marian Sirera and Luis Abad
- 113** *Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review*  
Cosimo Tuena, Elisa Pedroli, Pietro Davide Trimarchi, Alessia Gallucci, Mattia Chiappini, Karine Goulene, Andrea Gaggioli, Giuseppe Riva, Fabrizia Lattanzio, Fabrizio Giunco and Marco Stramba-Badiale
- 132** *Agency and Performance of Reach-to-Grasp With Modified Control of a Virtual Hand: Implications for Rehabilitation*  
Raviraj Nataraj, Sean Sanford, Aniket Shah and Mingxiao Liu



- 151** *Using More Ecological Paradigms to Investigate Working Memory: Strengths, Limitations and Recommendations*  
Lison Fanuel, Gaën Plancher and Pascale Piolino
- 156** *Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality*  
Silvia Cavedoni, Alice Chirico, Elisa Pedroli, Pietro Cipresso and Giuseppe Riva
- 168** *Being in the Past and Perform the Future in a Virtual World: VR Applications to Assess and Enhance Episodic and Prospective Memory in Normal and Pathological Aging*  
Azzurra Rizzo, Giuditta Gambino, Pierangelo Sardo and Valerio Rizzo



# Immersive Virtual Reality as an Adjunctive Non-opioid Analgesic for Pre-dominantly Latin American Children With Large Severe Burn Wounds During Burn Wound Cleaning in the Intensive Care Unit: A Pilot Study

Hunter G. Hoffman<sup>1\*</sup>, Robert A. Rodriguez<sup>2,3</sup>, Miriam Gonzalez<sup>2,3</sup>, Mary Bernardy<sup>3</sup>, Raquel Peña<sup>2,3</sup>, Wanda Beck<sup>3</sup>, David R. Patterson<sup>4</sup> and Walter J. Meyer III<sup>2,3</sup>

<sup>1</sup> Department of Mechanical Engineering, College of Engineering, University of Washington, Seattle, WA, United States,

<sup>2</sup> Psychiatry and Behavioral Sciences, University of Texas Medical Branch at Galveston, Galveston, TX, United States,

<sup>3</sup> Shriners Hospitals for Children, Galveston, TX, United States, <sup>4</sup> Department of Rehabilitation Medicine, University of Washington, Seattle, WA, United States

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Marco Fyfe Pietro Gillies,  
Goldsmiths University of London,  
United Kingdom  
Daniel Simon Harvie,  
Griffith University, Australia

### \*Correspondence:

Hunter G. Hoffman  
hoontair@gmail.com

**Received:** 09 April 2019

**Accepted:** 11 July 2019

**Published:** 08 August 2019

### Citation:

Hoffman HG, Rodriguez RA, Gonzalez M, Bernardy M, Peña R, Beck W, Patterson DR and Meyer WJ III (2019) Immersive Virtual Reality as an Adjunctive Non-opioid Analgesic for Pre-dominantly Latin American Children With Large Severe Burn Wounds During Burn Wound Cleaning in the Intensive Care Unit: A Pilot Study.  
*Front. Hum. Neurosci.* 13:262.  
doi: 10.3389/fnhum.2019.00262

**Background/Aim:** Using a within-subjects, within-wound care design, this pilot study tested for the first time, whether immersive virtual reality (VR) can serve as an adjunctive non-opioid analgesic for children with large severe burn wounds during burn wound cleaning in the ICU, in a regional burn center in the United States, between 2014–2016.

**Methods:** Participants included 48 children from 6 years old to 17 years of age with >10% TBSA burn injuries reporting moderate or higher worst pain during no VR on Day 1. Forty-four of the 48 children were from developing Latin American countries. Patients played adjunctive SnowWorld, an interactive 3D snowy canyon in virtual reality during some portions of wound care, vs. No VR during comparable portions of the same wound care session (initial treatment condition randomized). Using Graphic Rating scales, children's worst pain ratings during "No VR" (treatment as usual pain medications) vs. their worst pain during "Yes VR" was measured during at least 1 day of wound care, and was measured for up to 10 study days the patient used VR.

**Results:** VR significantly reduced children's "worst pain" ratings during burn wound cleaning procedures in the ICU on Day 1. Worst pain during No VR = 8.52 (SD = 1.75) vs. during Yes VR = 5.10 (SD = 3.27),  $t_{(47)} = 7.11$ ,  $p < 0.001$ , SD = 3.33, CI = 2.45–4.38, Cohen's  $d = 1.03$  (indicating large effect size). Patients continued to report the predicted pattern of lower pain and more fun during VR, during multiple sessions.

**Conclusion:** Immersive virtual reality can help reduce the pain of children with large severe burn wounds during burn wound cleaning in the Intensive Care Unit. Additional research and development is recommended.

**Keywords:** virtual reality, pain, pediatric burn injuries, analgesia, critical care, burn, opioid, developing countries

## INTRODUCTION

Acute pain is a frequent medical problem world wide, but children with large severe burn injuries (e.g., 40% TBSA) experience some of the most painful procedures in medicine. During the course of their weeks in the hospital burn center's intensive care unit, children with large severe burns must have their wounds cleaned/scrubbed frequently to prevent infection and speed up healing. Opioid analgesics are widely regarded as effective and essential tools for acute pain management (Malchow and Black, 2008; Vijayan, 2011; McIntyre et al., 2016; Ballantyne, 2018; Krane, 2019). According to Berterame et al. (2016, p. 1664) *"In developing countries, access to opioids is very limited. In 2009, more than 90% of worldwide use of opioid analgesics occurred in the USA, Canada, Australia, New Zealand, and several European countries. Use in that year was deemed low in 21 countries and very low in more than 100."* Patients in Latin American often have limited access to opioids for pain control (used for both analgesia and anesthesia). Yet even in the U.S.A., there are currently shortages of pharmaceutical medical opioid analgesics needed for acute pain control during medical procedures (Davis et al., 2018). And because of a large increase in opioid related overdose deaths unrelated to burn patients (Chen et al., 2019), there is growing political and legal pressure to further reduce reliance on opioids for pain control in the U.S.A.

For patients treated with opioid pain medications (e.g., patients treated in regional hospital burn centers in the United States), opioid side effects (Dunwoody and Jungquist, 2018) limit dose levels, limiting analgesic effectiveness (Cherny et al., 2001; Malchow and Black, 2008; Clark et al., 2017; Ballantyne, 2018). And opioid tolerance/habituation is a challenge for patients with large severe burns (Bittner et al., 2015), who typically receive the same painful procedures over and over, several times per week, often daily, during several weeks of hospitalization. Excessive pain and/or repeated high opioid doses can pathologically alter the patients pain perception system, disrupting the patient's natural endogenous opioid analgesia system (Schwaller and Fitzgerald, 2014; Ballantyne, 2018; Chambers, 2018), and can increase patient's risk of developing chronic pain, anxiety disorders, and/or Post-Traumatic Stress Disorder (McGhee et al., 2011; Rosenberg et al., 2015, 2018; Padesi and Fuzaylov, 2017; Peña et al., 2017).

Psychological factors such as fear, anxiety, and depression can increase or amplify how much pain patients subjectively experience during painful medical procedures (Hemington et al., 2017; Nitzan et al., 2019), making pain management even more challenging. What people are thinking about during wound care, and where patients direct their attention during medical procedures can influence pain intensity (Melzack and Wall, 1965). For example, if patients predict wound care is going to be painful, that can make their pain worse. According to Fields (2018, p. S8) "...expectation of pain becomes a self-fulfilling prophecy through top down amplification of the pain signal," and memories of previous painful procedures can also increase pain intensity (Noel et al., 2015).

Fortunately, just as psychological factors can make pain worse, psychological treatments can help reduce acute pain during

medical procedures. For example, distraction techniques (e.g., music) are widely used in clinical practice, and can be used in addition to traditional pain medications to help control pain during burn wound care. Some studies show strong benefits of music therapy during burn wound care in patients (e.g., Rohilla et al., 2018). But in other studies the benefits of listening to music during burn wound care had small effect sizes and/or non-significant results (Fratianne et al., 2001; Bellieni et al., 2013; van der Heijden et al., 2018), and/or involved patients with small burn wounds (e.g., 5% TBSA, Hsu et al., 2016).

For the extreme pain levels experienced by children with large severe burn wounds during burn wound debridement in the intensive care unit, creating stronger non-pharmacologic pain control techniques is a national and international priority (Keefe et al., 2018).

Immersive virtual reality is a promising new non-opioid psychological pain distraction technique. There is growing evidence that adjunctive immersive virtual reality distraction can significantly reduce how much pain patients experience during a growing number of different painful medical procedures e.g., during urological endoscopies, physical therapy after surgery for cerebral palsy, venipuncture for onco-therapy, and pediatric dental procedures (Hoffman et al., 2011; Garrett et al., 2014; Scheffler et al., 2017; Atzori et al., 2018a,b; Indovina et al., 2018; Honzel et al., 2019).

Brain scan studies provide converging evidence that VR reduces acute pain. Using neuroimaging assessments, a laboratory functional magnetic resonance imaging study found that in addition to reducing subjective pain ratings, VR reduced pain-related brain activity (Hoffman et al., 2004b). In a second fMRI brain scan study, the amount of pain reduction from VR alone was comparable to the amount of pain reduction from a moderate dose of hydromorphone, and "VR + opioids" combined resulted in the largest pain reductions (Hoffman et al., 2007).

The logic for why VR would reduce pain is based on an attentional mechanism (Hoffman, 1998; Hoffman et al., 2000, 2006). The essence of immersive virtual reality analgesia is the patient's illusion of going to a different place, the subjective experience of "feeling present" in the computer generated world, as if the virtual reality world is a place they are visiting (Slater and Wilbur, 1997). Human brains are limited in how much information they can process (Kahneman, 1973). Pain requires attention. Researchers argue that the illusion of "being there" in virtual reality is unusually attention grabbing, reducing the amount of attentional resources the patient's brain has available for pain perception (Hoffman, 1998; Hoffman et al., 2000, 2003; Hoffman et al., 2004a).

According to a gate control theory explanation of psychological analgesia (Melzack and Wall, 1965, p. 978), "...psychological factors such as past experience, attention, and emotion can influence pain response and perception..." Melzack and Wall proposed that the brain may inhibit nociceptive signals.

Regardless of the mechanism, several small clinical studies have shown encouraging preliminary evidence that adjunctive VR can help reduce pain during burn wound care in adults (Hoffman et al., 2004a, 2011; van Twillert et al., 2007; Maani

et al., 2011a,b; McSherry et al., 2018), and in children with small burns, (Hoffman, 1998; Hoffman et al., 2000; Faber et al., 2013; Jeffs et al., 2014; Khadra et al., 2018). There is also preliminary evidence that VR is more effective than conventional distractions such as video games or movies. In the first study to report using immersive virtual reality for pain control during a medical procedure, two adolescent boys with large burn injuries underwent staple removals from healing burn skin grafts during immersive VR vs. while playing a Nintendo video game (no VR). Both patients reported large reductions in pain during staple removal during immersive virtual reality compared to their pain during staple removal while playing the (no VR) traditional Mario Kart Nintendo video game (Hoffman et al., 2000) during the same wound care session. More recently, in a study by Jeffs et al. (2014) adolescent burn patients with small burns (5% TBSA) treated in an outpatient clinic reported significantly lower pain during virtual reality compared to a group that watched a movie during wound cleaning.

There are a number of barriers to using VR in the ICU tubroom. The patients in the current study had a burn size of 40% Total Body Surface Area (TBSA). As is often the case for patients with such unusually large severe burn injuries, most of the burn patients in our study had head and face burns, preventing them from wearing a conventional commercially available head mounted VR helmet. Furthermore, even when treated with powerful pain medications, pain during burn wound care procedures in the ICU hydrotank is often “severe to excruciating,” which may make it harder for children to concentrate enough to play in VR during wound care. In theory, pain may become so attention grabbing that psychological distraction techniques cannot compete with pain for the patient’s limited attention (Eccleston and Crombez, 1999; Eccleston, 2001). In other words, some patients may not benefit from VR if their acute procedural pain becomes too intense. Similarly, traditional distraction may fail if patients feel threatened during the wound care (McCaul and Malott, 1984; Crombez et al., 1998). High catastrophizers (people who have unusually negative emotions and pessimistic beliefs about their ability to deal with the upcoming pain) may have difficulty disengaging attention from pain information (Verhoeven et al., 2012; Van Loey et al., 2018).

To address these challenges, using a custom water-friendly VR system, the current pilot study tests for the first time, whether adjunctive virtual reality can reduce the acute procedural pain of children with large severe burn injuries during burn wound debridement/cleaning in the pediatric intensive care unit, in an understudied patient population, critically injured pediatric patients.

We hypothesize that compared to standard of care (standard pain medications + No VR), during adjunctive Yes VR, children will report significant reductions in worst pain ratings. Our secondary hypothesis is that during VR, children will report significant reductions in pain unpleasantness, and will spend less time thinking about pain during burn wound debridement in the ICU hydrotank. We further hypothesize that VR will increase how much fun patients have during wound care, and that patients will be more satisfied with their pain management during VR.

## MATERIALS AND METHODS

This research was conducted between Jan 2014 and Dec 2016, in accordance with the Declaration of the World Medical Association ([www.wma.net](http://www.wma.net)). The studies were approved by the IRB from UTMB, and all participants and their parents/legal guardians provided written informed consent/assent in accordance with the Declaration of Helsinki.

Most of the children in the current study were transported from Latin America to Shriners Hospitals for Children in Galveston Texas, U.S.A., where they were hospitalized, treated, and returned to their country of origin, post-discharge.

### Inclusion Criteria

Children were included in the study if they were (1) compliant and able to complete subjective evaluations, (2) had no history of previous psychiatric (DSM-III-R Axis I) disorder(s), (3) were not demonstrating delirium, psychosis, or any form of organic brain disorder, (4) were able to communicate verbally in English or Spanish, and (5) had moderate or higher worst pain during no VR on Day 1, (6) were admitted to Shriners Hospitals for Children: Galveston Texas/University of Texas Medical Branch.

Children were excluded from the study if (1) they had a burn size <10% TBSA, (2) they were not capable of completing the study measures, (3) if no wound cleaning sessions were required, (4) if they had a history of previous psychiatric (DSM-III-R Axis I) disorder(s), (5) if they were demonstrating delirium, psychosis, or organic brain disorder, (6) if the child was unable to communicate verbally in English or Spanish, (7) if they had a history of significant cardiac, endocrine, neurologic, metabolic, respiratory, gastrointestinal, or genitourinary impairment, (8) if they were receiving prophylaxis for alcohol or drug withdrawal, (9) if they had a developmental disability, (10) if they were younger than 6 years old, (11) if they were older than 17 years old, or (12) if they had burns of eyes, eyelids, or face so severe the burns precluded the use of VR equipment, (13) or if patients reported having a previous history of severe motion sickness.

### Equipment

The current study introduced for the first time, a new portable water-friendly VR system customized for the unique needs of pediatric patients with large severe burn injuries during wound care in the intensive care unit hydrotank. As shown in **Figure 1**, a custom robot-like articulated arm goggle holder was used in the current study to hold a pair of VR goggles near the patient’s eyes, so patients did not have to wear a VR helmet on their head. This “Magula arm” robot-like goggle holder minimized or ideally eliminated contact between the patient and the VR goggles. The VR goggles largely blocked the patient’s view of the Intensive Care Unit hydrotank room. The goggles were MX90 VR goggles, from NVIS.com, with 90 degrees field of view diagonal, per eye, and 1,280 × 1,024 pixels resolution per eye. All of the VR equipment in the current study was battery powered. A battery powered laptop and battery powered audio-visual unit were used with the MX90 VR goggles. The 90 degrees diagonal field of view goggles increased the amount of peripheral vision stimulated. During the VR condition, patients were encouraged to interact with

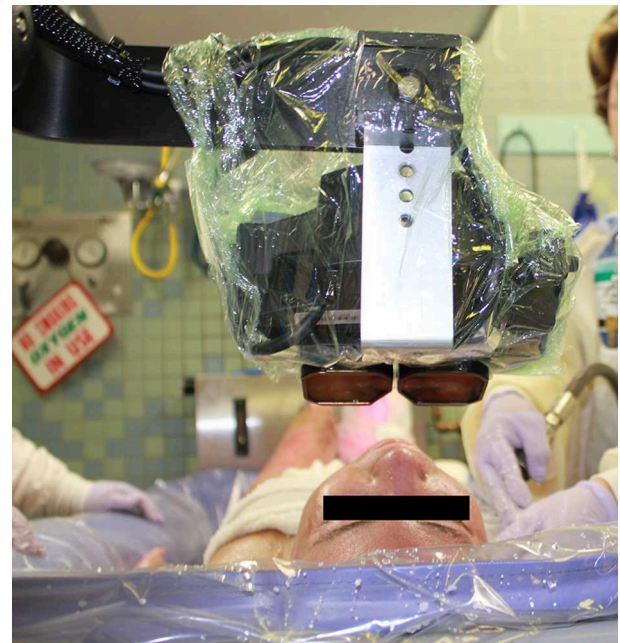




**FIGURE 1** | A patient playing SnowWorld during burn wound debridement in the ICU tankroom. Photo and copyright Hunter Hoffman, www.vrpain.com.

the virtual environment via a wireless computer mouse. Stereo speakers helped isolate patients from hearing hospital sounds. The custom robot-like articulated arm goggle holder was securely mounted to the frame of the Anthro medical cart. The VR goggles orientation could be adjusted and locked into position for a patient who was sitting up during wound care, or the goggles could be rotated and locked into position for a patient who was lying on their backs during wound care (see **Figure 2**). The goggles stayed in one position, and the patient used their wireless mouse to look around, aim and shoot snowballs in SnowWorld (mouse-tracking instead of head tracking).

The portable robot-like arm goggle holder was designed by Hoffman and Magula and built by Jeff Magula, an advanced instrument maker at the University of Washington in Seattle. Once finished, the water-friendly VR system was then safety inspected by Clinical Engineering at the University of Washington, and was inspected again by Clinical Engineering at Shriners Hospitals for Children. The equipment was also approved for use in the Intensive Care Unit and the equipment cleanliness was monitored by infection control at Shriners Hospitals for Children. After each use, the VR cart/portable VR system was returned to the Psychology Department at Shriners Galveston, where it was plugged in to recharge the batteries after each use. As shown in **Figure 2**, the goggles were partially covered with disposable plastic, which was discarded after each use. The equipment was systematically disinfected after each use using chemical disinfectants, and was periodically supercleaned using ultraviolet radiation (using a portable UV lamp wand, UV protective glasses, while wearing latex gloves). For example, the UVC Blade Handheld Germicidal Fixtures by American Ultraviolet. The VR system was periodically tested for pathogens, using swabs that were then analyzed by Shriners infection control, to test for the presence of bacteria. Culture samples (swabs) were sent to the microbiology laboratory at Shriners hospital in Galveston for immediate analysis. The post-cleaning tests all came back as “safe” (no pathogens). There



**FIGURE 2** | A patient looking into VR goggles during burn wound debridement in the ICU tank room. Photo and copyright Hunter Hoffman, www.vrpain.com.

was no significant problem with infection, using the current VR system, which minimized or eliminated physical contact between the patient and the VR goggles.

## MEASURES

After each wound care session, subjects received the following instructions once prior to answering each of five separate questions. “Please indicate how you felt during wound care today by making a mark anywhere on the line. Your response doesn’t have to be a whole number.”

For the primary dependent measure, using Graphic Rating Scales (GRS), after the wound care session, patients answered the following GRS ratings. Pain was measured using Graphic Rating Scales (GRS) (Jensen and Karoly, 2001; Jensen, 2003). In the current study, the GRS tool was used to assess three reports of the pain experience (“worst pain,” “pain unpleasantness,” and “time spent thinking about pain”) that correspond to three separable components of the pain experience; sensory pain, affective pain, and cognitive pain, respectively. The GRS is a 10-unit horizontal line labeled with number and word descriptors. Descriptor labels were associated with each mark to help the respondent rate pain magnitude in each domain. For worst pain, the GRS descriptors were *no pain at all*, *mild pain*, *moderate pain*, *severe pain*, and *excruciating pain*. For pain unpleasantness, the GRS descriptors were *not unpleasant at all*, *mildly unpleasant*, *moderately unpleasant*, *severely unpleasant*, and *excruciatingly unpleasant*. For time spent thinking about pain, the GRS descriptors were *none of the time*, *some of the time*, *half of the time*, *most of the time*, *all of the time*.

The Graphic Rating Scale has previously been used to assess pain intensity in children eight and older and has been documented to be the preferred report method for young children (Tesler et al., 1991). The GRS is more sensitive than simple descriptive pain scales and patients can easily answer these pain ratings despite having no previous experience. Visual Analog Scales have been validated for use in children aged 7 and higher (Bringuier et al., 2009).

A single rating “to what extent did you feel like you ‘went into’ the virtual world,” adapted from Slater et al. (1994) was also used in the present study to assess user presence in the virtual world. Descriptor labels were *I did not feel like I went inside at all, mild sense of going inside, moderate sense of going inside, strong sense of going inside, I went completely inside the computer generated world*. Hendrix and Barfield (1995) showed the reliability of a similar VR presence rating. The measure’s ability to detect treatment effects (Hoffman et al., 2004c) is preliminary evidence of our VR presence measure’s validity. Patients also rated how real the objects seemed in virtual reality, descriptors were *completely fake, somewhat real, moderately real, very real, indistinguishable from a real object*. Patients rated how satisfied they were with their pain management during No VR vs. during VR, with descriptors *completely unsatisfied, mostly unsatisfied, half satisfied, mostly satisfied, completely satisfied*, and patients rated nausea as a result of VR, using a graphic rating scale with descriptors *no nausea at all, mild nausea, moderate nausea, severe nausea, vomit*. All text was translated into Spanish for Spanish speaking participants using an official translator (90% of the participants in this study were Spanish only speaking). To assess whether patients in the upper quartile on catastrophizing showed pain reduction during immersive Virtual Reality, we administered the Pain Catastrophizing Scale for Children (PSC-C) (Sullivan et al., 1995; Crombez et al., 2003). The PCS total score is calculated by summing the 13-item responses, and provides a good index of the catastrophizing construct through the inclusion of highly correlated subscales of helplessness, rumination, and magnification. Higher scores on the PCS-C are indicative of greater pain-related catastrophizing. The PCS-C has been validated for use with children (Crombez et al., 2003).

## Experimental Design

There is high variability in the analgesic effectiveness of any given dose of pharmacologic analgesia from one burn wound care session to the next (Khadra et al., 2018). And furthermore, pain medication dose levels can also vary from day to day. For these reasons, in the current preliminary study, a statistically powerful within-subjects, within-wound care design was used (Maani et al., 2011a). During VR, patients played SnowWorld, an interactive 3D snowy canyon in virtual reality during some portions of wound care, vs. No VR during comparable portions of the same wound care session. Childrens’ worst pain during “No VR” (treatment as usual pain medications) vs. their worst pain during “Yes VR” was measured during at least 1 day of wound care, and was measured for up to 10 study days the patient used VR. Initial treatment order was randomized using blocked randomization, based on random number sequences generated using [www.random.org](http://www.random.org). All patients received their



**FIGURE 3 |** SnowWorld. An icy 3D canyon in virtual reality. Image by Ari Hollander and Howard Rose, copyright Hunter Hoffman, [www.vrpain.com](http://www.vrpain.com).

usual pain medications on all study days, i.e., VR was always used adjunctively, in addition to usual traditional pain medications.

During wound care, the nurses cut off and removed the patient’s gauze bandages, and began cleaning the patients burn wounds, using warm wet washcloths and a hand held warm water shower hose to scrub and rinse away dead tissue and debris out of the burn wound. During wound debridement, patients received No VR and Yes VR during approximately equally painful portions of the same wound care session. The patient began receiving wound care for 5 min with Yes VR vs. 5 min with No VR, Yes VR for five more minutes, etc. repeatedly alternating between No VR and Yes VR every 5 min. Whether patients received Yes VR or No VR during the first 5 min treatment segment was randomized (blocked randomization using a random sequence generated at [random.org](http://random.org)). During the portions of their burn wound care that they received VR, the research staff positioned the VR goggles weightlessly near the patient’s eyes, with little or no physical contact between the VR goggles and the patient, using a robot-like-arm goggle holder (Maani et al., 2008). The patient looked into the VR goggles, and interacted with the virtual reality world.

All patients used SnowWorld (see **Figure 3**) during all VR sessions. SnowWorld is a non-profit VR world specifically designed for pain distraction of immobilized severe burn patients, including children. SnowWorld is designed to give burn patients the illusion of going inside a snowy 3D canyon (Hoffman et al., 2001, 2004b,c; see also Bloemink et al., 2006, p. 104–106). In SnowWorld ([www.vrpain.com](http://www.vrpain.com)), patients interacted with snowmen, igloos, penguins, wooly mammoths, and flying fish by throwing snowballs, using a wireless computer mouse to aim and trigger snowballs while keeping their heads and bodies motionless. During VR, patients heard music (e.g., Paul Simon’s song Graceland, and several Spanish songs), and 3D sound effects e.g., ice breaking when a snowball hits a snowman. Mammoths trumpeted angrily when pelted.

After the wound care session was over, patients briefly rated how much pain they had experienced during No VR



vs. during Yes VR using graphic rating scales. The patient's burns were rebandaged, the patient was wheeled back to their hospital room and returned to their hospital beds, and the research staff thoroughly cleaned and disinfected the VR equipment.

## Statistical Analyses

IBM SPSS (2018) statistical analyses of the primary and secondary hypotheses involved an apriori two-tailed within-subjects paired  $t$ -test, with  $\alpha = 0.05$ .

## RESULTS

Patients participated between January 2014 and December 2016. Out of the 62 patients initially screened, 48 pediatric patients met our apriori inclusion criterion of having a moderate or higher "worst pain" rating during No VR on Day 1 (33 hispanic males children, 11 hispanic female children from developing Latin American countries, and also three non-hispanic female children and one non-hispanic male from the United States). The mean size of the patient's severe burn injuries was 40 percent Total Body Surface Area (TBSA) burned, 28% third degree burns. Patients' ages ranged from 6 to 17 years of age at time of enrollment (Mean age was 12 years old). Seventy-seven percent of the patients had hand burns, 85% had arm burns, 44% had foot burns, 79% had leg burns, 71% had neck/head burns, 79% had trunk/torso burns, and 23% had groin burns. Regarding the (sometimes overlapping) etiology of their burns, 81% had burns involving flame, 6% scalds, 25% electrical, and zero patients had chemical burns.

### Test of Our Primary Hypothesis

The patients GRS pain ratings on Day 1 are shown in **Table 1** and **Figure 4**. On Day 1, on a zero to 10 graphic rating scale, using a paired  $t$ -test, VR significantly reduced children's "worst pain" ratings during burn wound cleaning procedures in the ICU. On Day 1, worst pain during No VR = 8.52 (SD = 1.75) vs. during Yes VR = 5.10 (SD = 3.27),  $t_{(47)} = 7.11$ ,  $p < 0.001$ , SD = 3.33, CI = 2.45–4.38, Cohen's  $d = 1.03$ , indicating a large effect size.

### Descriptive Statistics About "Worst Pain" Ratings

On Day 1, the number of patients reporting excruciating pain (worst pain = 10) during wound care was 22 patients, which dropped to only five patients reporting excruciating pain (worst pain = 10) during Yes VR, and Cohen's  $d$  showed a strong effect size of VR analgesia. However, many of those patients with pain of 10 during No VR only dropped to 8 during VR (i.e., they dropped from excruciating pain during No VR down to severe pain during VR, but still reported severe pain during adjunctive VR).

*On Day 1, 40% of the 48 patients still reported pain of 7 or higher (severe to excruciating) during VR, despite receiving powerful traditional pharmacologic pain medications combined with immersive virtual reality.*

On average, patients spent mean = 16.56 min of wound care during No VR vs. 12.89 min during VR,  $t_{(44)} = 2.47$ ,  $p < 0.05$ , SD = 9.97, CI = 0.67–6.66, e.g., patients could not use VR while having their faces or heads cleaned. On Day 1, 14 of the 48 patients spent exactly the same amount of time during No VR (13.21 min) and during VR (13.21 min). These 14 patients also reported large and statistically significant reductions in pain during VR, worst pain during No VR = 8.50 (SD = 1.83), VR = 4.43 (SD = 3.08),  $t_{(13)} = 4.56$ ,  $p < 0.005$ , SD = 3.34, CI = 2.14–6.00.

The mean number of days that patients rated their pain during Yes VR vs. during No VR was 4 study days. Collapsed across days, VR significantly reduced worst pain: Worst pain during "No VR" (Mean = 7.09, SD = 2.10) vs. worst pain during "Yes virtual reality" (Mean = 4.29, SD = 2.55),  $t_{(47)} = 7.32$ ,  $p < 0.001$ , SD = 2.65, CI = 2.01–3.57, Cohen's  $d = 1.06$ , large effect size.

Consistent with the prediction that VR would continue to reduce pain when used day after day, a one-way within-subjects ANOVA comparing worst pain during "No VR" minus worst pain during "Yes VR" difference scores for days 1–7 showed no significant difference in the size of the VR analgesia effect over days 1–7, Wilks'  $\Lambda_{(4,6)} = 1.50$ ,  $p = 0.36$ , NS.

In exploratory analyses, patients scoring in the upper quartile on the children's pain catastrophizing score (PSC-C) in the current sample, showed significant VR analgesia. For patients scoring in the upper quartile on catastrophizing, mean worst pain during No VR = 7.00 (SD = 3.56), vs. VR = 2.86 (SD = 3.63),  $t_{(6)} = 2.80$ ,  $p < 0.05$ , SD = 2.56, CI = 0.34 vs. 5.09. Patients scoring in the lower quartile in the current sample also showed significant VR analgesia, mean worst pain ratings during No VR = 6.00 (SD = 4.04), and during VR = 2.86 (SD = 2.85),  $t_{(6)} = 4.01$ ,  $p < 0.01$ , SD = 1.03, CI = 1.61–6.67.

### Test of Secondary Hypotheses

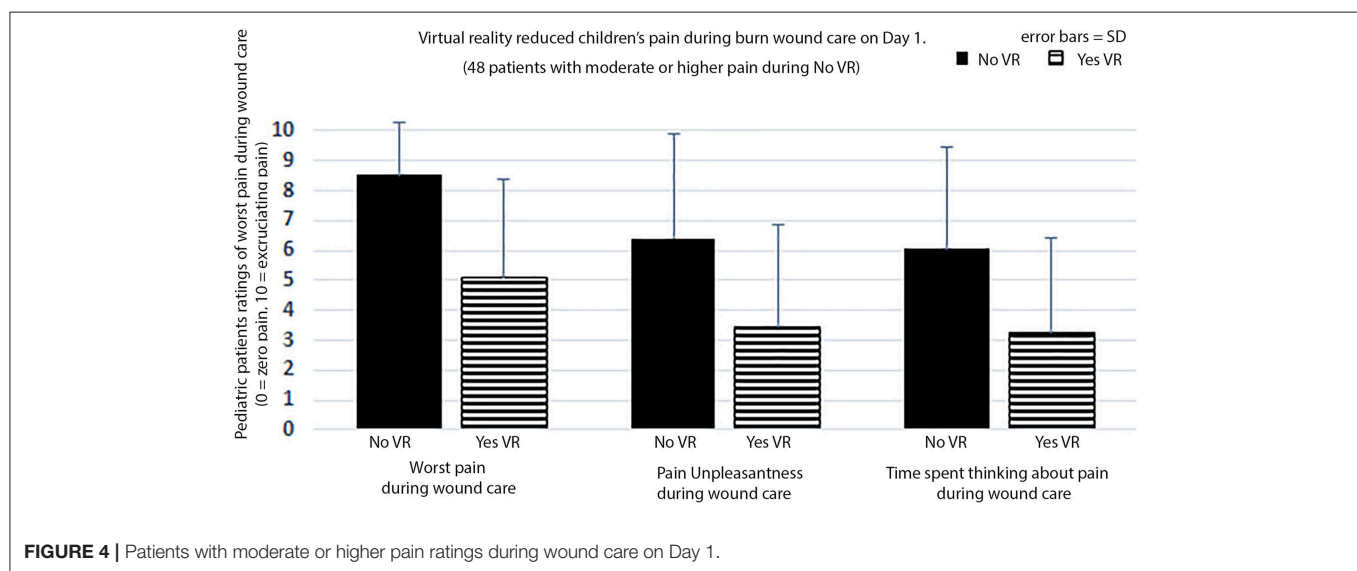
As shown in **Table 1** and **Figure 4**, on secondary GRS measures, on Day 1, pediatric burn patients reported large and significant reductions in pain on secondary measures of "pain unpleasantness" and "time spent thinking about pain during wound care." Although children reported having 27% more fun during VR, the increase in fun on Day 1 was not statistically significant in the paired  $t$ -test. The children were significantly more satisfied with their pain management during VR, on average. Patients reported only a moderate illusion of "being there" inside the 3D computer generated world as if it was a place they visited. VR nausea was nearly zero (<1 on a 10 point scale).

The current study included 48 pediatric patients total. As shown in **Table 2**, in an exploratory analysis, to see if children from developing countries show VR analgesia, the subset (sub-analysis) of 44 patients from developing Latin American countries were analyzed separately from the four patients from the United States. As predicted, children from developing countries showed significant reductions in worst pain during VR, as well as significant reductions in pain unpleasantness (the emotional component of pain) and significant reductions in time spent thinking about pain during wound care (the cognitive component of pain). Encouragingly, analyzed separately, the four

**TABLE 1** | Means (Standard Deviation) in “No-VR” condition vs. “Yes-VR” condition.

	No-VR mean	Yes- VR	t (df)	p-value (Sig 2- tailed)	Confidence interval	Cohen's d effect size	Mean Diff.	SD difference
	Mean (SD)	Mean (SD)						
Worst pain	8.52 (1.75)	5.10 (3.27)	7.11 (47)	<0.001	2.45 to 3.33	1.03 large	3.42	3.33
Time spent thinking about pain	6.04 (3.41)	2.47 (3.37)	5.94 (46)	<0.001	1.86 to 3.76	0.87 large effect size	2.81	3.24
Pain	6.40	3.47 (3.37)	5.49 (44)	<0.001	1.86 to 4.01	0.82 large effect size	2.93	3.58
Unpleasant ness	3.51	(3.37)						
Fun	4.81 (3.93)	6.68 (3.86)	2.01 (46)	0.051 NS	3.75 to 0.004	0.29 small effect size	1.87	6.39
Satisfaction with pain management	5.22 (3.34)	8.04 (2.33)	4.72 (22)	<0.001	1.59 to 4.07	0.99 large effect size	2.83	2.87

All 48 patients (44 children from developing countries and also 4 children from the USA).



participants from the United States also showed the predicted patterns of large reductions of pain during VR.

## DISCUSSION

This pilot study provides preliminary evidence that immersive virtual reality can help reduce the pain of children with large severe burn wounds during burn wound cleaning in the Intensive Care Unit. Although using VR in the ICU hydrotank room was challenging and required creating custom equipment, in the current study, on Day 1, patients reported significant reductions in worst pain (pain intensity), children spent less time thinking about their pain during VR, children reported significant reductions in pain unpleasantness, and the children reported 27% higher ratings of fun during wound care during virtual reality. In addition, these pediatric patients were also significantly more satisfied with their pain management during virtual reality, they reported a moderate illusion of presence in VR (i.e., a moderately strong illusion of “being there” in the

VR computer generated world during wound care), and VR nausea was nearly zero (<1 on a 10 point scale). Patients who received VR during more than 1 day of wound care continued to report the predicted pattern of reductions in worst pain during multiple wound care sessions. And patients with a tendency toward negative emotions and pessimistic beliefs about their ability to deal with the upcoming pain (i.e., patients in the upper quartile on catastrophizing), still benefitted from virtual reality distraction.

The reductions in worst pain ratings in the current study are similar to the pattern of VR analgesia reported in previous studies of 12 U.S. soldiers with combat-related burn injuries (TBSA of 21%) during wound care in their hospital beds. The soldiers spent 6 min in No VR vs. 6 min of wound care during VR (Maani et al., 2011a). In the current study the mean burn size was over 40%, the patients were all children, and the sample size was larger ( $n = 48$  patients). Furthermore, in the current study, on average, patients spent over 12 min in VR and over 12 min in No VR, the wound care was conducted in the ICU instead of in the patients

**TABLE 2 |** Means (Standard Deviation) in “No-VR” condition vs. “Yes-VR” condition.

	No-VR	Yes- VR	t (df)	p-value (Sig 2- tailed)	Confidence interval	Cohen's d effect size	Mean Diff.	SD difference
	Mean (SD)	Mean (SD)						
Worst pain	8.43 (1.80)	5.20 (3.18)	6.65 (43)	<0.001	2.25 to 3.22	1.00 large	3.23	3.22
Time spent thinking about pain	5.86 (3.46)	3.33 (3.28)	5.32 (42)	<0.001	1.57 to 3.50	0.81 large effect size	2.54	3.13
Pain unpleasant ness	6.26 (3.54)	3.57 (3.42)	5.17 (41)	<0.001	1.64 to 3.74	0.80 large effect size	2.69	3.38
Fun	5.00 (3.85)	6.79 (3.76)	1.89 (42)	0.07 NS	3.70 to 0.12	0.29 small effect size	1.79	6.21

*Sub-analysis of only the 44 children from developing Latin American countries (excluding the four children from the USA).*

hospital beds, and the current study is the first to use a portable water-friendly VR system.

## LIMITATIONS

The demographics and characteristics of the participants of this pediatric pain study may limit generalization of findings of this study to other populations. Of interest is that 44 of the 48 patients were Spanish speaking patients from developing Latin American countries. As predicted, in an exploratory sub-analysis, the 44 children from developing Latin American countries showed statistically significant reductions in pain during VR. Encouragingly, analyzed separately, the four participants from the United States also showed the predicted patterns of large reductions of pain during VR. The VR system used in the current study was customized for use in the ICU hydrotank room, for patients with head and facial burns. Future randomized controlled trials research is needed to determine whether the current results replicate, and generalize to other VR systems.

Despite these limitations, the current study makes several important original contributions to the literature, and the results of the current study could have important implications for clinical practice: (a) this is the first study ever to attempt to use virtual reality during burn wound care in the intensive care unit, (b) the patients had unusually severe burn injuries much larger than burn injuries treated in any previous burn debridement VR analgesia study, (c) all of the patients were children, and 44 out of the 48 patients were Spanish speaking children from developing Latin American Countries, (d), the current study shows for the first time that children with large severe burns were generally able to play SnowWorld during severely painful medical procedures, and (e) playing SnowWorld in virtual reality significantly reduced worst pain ratings during wound care.

In the current study, a custom portable water-friendly VR system was used that did not have to physically contact the patient. The equipment was carefully cleaned with sterilizing cloths after each use, and the equipment was periodically swabbed/tested by the hospitals infection control team to test for the presence of any bacterial or viral pathogens. There was no problem with infection in the current study, using the

custom VR system, which minimized or eliminated physical contact between the patient and the VR goggles. For patients with limited ability to wear VR helmets, modified VR systems that reduce contact surfaces (Hoffman et al., 2014) are highly recommended for use of VR during burn wound care for patients with severe unbandaged head and/or face burns. We also recommend discarding disposable foam liners that touch the patients face, after each use. Burn patients are especially vulnerable to infections when unbandaged (during wound care), and VR equipment should be monitored by infection control, especially when used in the Intensive Care Unit.

## CONCLUSION

The results from the current pilot study support our hypothesis that immersive virtual reality can significantly reduce acute pain during burn wound care, even in pediatric patients with large severe burn wounds treated in the hydrotanks in the Intensive Care Unit. And VR continued to reduce pain when used day after day.

## FUTURE DIRECTIONS

Virtual reality (VR) may eventually prove to be “opioid sparing” during hospitalization (Kipping et al., 2012; McSherry et al., 2018). Additional research and development is needed on how to make VR analgesia more powerful (Wender et al., 2009), how to make pharmacologic pain medications more effective (McIntyre et al., 2016), and how to best combine pharmacologic pain medications and VR analgesia, to maximize total pain control. Development of more powerful new non-pharmacologic pain management techniques is a national and international priority (Keefe et al., 2018), and Virtual Reality has strong potential as a new direction for behavioral medicine (Keefe et al., 2012).

Fortunately, VR analgesia is not limited to severe burn patients, but could potentially be used for a wide range of painful medical procedures, and could be especially valuable for highly populated, lower income developing countries (4/5ths of the World's population), where large severe burns and other serious injuries are more common, and powerful pharmacologic

analgesics are more scarce or unavailable. Additional research and development of VR analgesia is recommended.

## DATA AVAILABILITY

The datasets for this study will not be made publicly available because IRB restrictions.

## ETHICS STATEMENT

This research was conducted in accordance with the Declaration of the World Medical Association ([www.wma.net](http://www.wma.net)). The studies were approved by the IRB from UTMB, and all participants provided written informed consent/assent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was financed by Shriners Hospitals for Children, Tampa Florida (award ID #71011-GAL, PI Walter Meyer), with

help from a charitable donation from the MayDay Fund (PI Walter Meyer). The portable water friendly VR system was developed via NIH grant R01GM042725 to DP.

## ACKNOWLEDGMENTS

Thanks to the patients and their parents for volunteering to participate in this study. Thanks to the wound care staff, and to Drs. Laura and Marta Rosenberg at Shriners Hospitals for Children in Galveston TX. Thanks to Jeff Magula (and also Eric J. Seibel and Bill Russell) for the custom VR system, and thanks to Bill Russell and Maribel Ramirez, for early pilot research on water-friendly VR in the ICU at Shriners Galveston. We would also like to thank Kristen Darken and all of the several teams of VR worldbuilders at Multigen-Paradigm, SimWright, Howard Abrams, and Duff Hendrickson, who have created the original SnowWorld 2001 and SnowWorld 2003, and thanks to Ari Hollander and Howard Rose for the latest version of the University of Washington's SnowWorld (2006) used in the current study ([www.vrpain.com](http://www.vrpain.com)). Thanks to the Ladies of the Nile, and Shriners Hospitals for Children Board of Directors, and to Dr. Steve Wolf, and to the Mexican Michou y Mao Foundation for their generous encouragement and support. Special thanks to singer/songwriter Paul Simon for suggesting we use Graceland with SnowWorld.

## REFERENCES

- Atzori, B., Hoffman, H. G., Vagnoli, L., Patterson, D. R., Al-Halabi, W. H. G., Messeri, A., et al. (2018b). Virtual Reality analgesia during venipuncture effectiveness among pediatric patients with onco-hematological diseases. *Front. Psychol.* 9:2508. doi: 10.3389/fpsyg.2018.02508
- Atzori, B., Lauro Grotto, R., Giugni, A., Calabrò, M., Alhalabi, W., and Hoffman, H. G. (2018a). Virtual reality analgesia for pediatric dental patients. *Front. Psychol.* 9:2265. doi: 10.3389/fpsyg.2018.02265
- Ballantyne, J. C. (2018). The brain on opioids. *Pain*. 159(Suppl 1), S24–S30. doi: 10.1097/j.pain.0000000000001270
- Bellieni, C. V., Cioncoloni, D., Mazzanti, S., Bianchi, M. E., Morrone, I., Becattelli, R., et al. (2013). Music provided through a portable media player (iPod) blunts pain during physical therapy. *Pain Manag. Nurs.* 14, e151–e155. doi: 10.1016/j.pmn.2011.09.003
- Berterame, S., Erthal, J., Thomas, J., Fellner, S., Vosse, B., Clare, P., et al. (2016). Use of and barriers to access to opioid analgesics: a worldwide, regional, and national study. *Lancet*. 387, 1644–1656. doi: 10.1016/S0140-6736(16)00161-6
- Bittner, E. A., Shank, E., Woodson, L., and Martyn, J. A. (2015). Acute and perioperative care of the burn-injured patient. *Anesthesiology* 122, 448–464. doi: 10.1097/ALN.0000000000000559
- Bloemink, B., Hodge, B., Lupton, E., and McQuaid, M. (2006). *Design Life Now, National Design Triennial*. New York, NY: Cooper-Hewitt, National Design Museum, Smithsonian Institution. Available online at: [www.cooperhewitt.org](http://www.cooperhewitt.org)
- Bringuier, S., Dadure, C., Raux, O., Dubois, A., Picot, M. C., and Capdevila, X. (2009). The perioperative validity of the visual analog anxiety scale in children: a discriminant and useful instrument in routine clinical practice to optimize postoperative pain management. *Anesth. Analg.* 109, 737–744. doi: 10.1213/ane.0b013e3181af00e4
- Chambers, C. T. (2018). From evidence to influence: dissemination and implementation of scientific knowledge for improved pain research and management. *Pain* 159(Suppl 1), S56–S64. doi: 10.1097/j.pain.0000000000001327
- Chen, Q., Laroche, M. R., Weaver, D. T., Lietz, A. P., Mueller, P. P., Mercaldo, S., et al. (2019). Prevention of prescription opioid misuse and projected overdose deaths in the United States. *JAMA Netw. Open*. 2:e187621. doi: 10.1001/jamanetworkopen.2018.7621
- Cherny, N., Ripamonti, C., Pereira, J., Davis, C., Fallon, M., McQuay, H., et al. (2001). Strategies to manage the adverse effects of oral morphine: an evidence-based report. *J. Clin. Oncol.* 19, 2542–2554. doi: 10.1200/JCO.2001.19.9.2542
- Clark, A., Imran, J., Madni, T., and Wolf, S. E. (2017). Nutrition and metabolism in burn patients. *Burns Trauma*. 17:11. doi: 10.1186/s41038-017-0076-xh
- Crombez, G., Bijttebier, P., Eccleston, C., Mascagni, T., Mertens, G., Goubert, L., et al. (2003). The child version of the pain catastrophizing scale (PCS-C): a preliminary validation. *Pain* 104, 639–646. doi: 10.1016/S0304-3959(03)00121-0
- Crombez, G., Eccleston, C., Baeyens, F., and Eelen, P. (1998). When somatic information threatens, catastrophic thinking enhances attentional interference. *Pain* 75, 187–198. doi: 10.1016/S0304-3959(97)00219-4
- Davis, M. P., McPherson, M. L., Mehta, Z., Behm, B., and Fernandez, C. (2018). What parenteral opioids to use in face of shortages of morphine, hydromorphone, and fentanyl. *Am. J. Hosp. Palliat. Care* 35, 1118–1122. doi: 10.1177/1049909118771374
- Dunwoody, D. R., and Jungquist, C. R. (2018). Sedation scales: do they capture the concept of opioid-induced sedation? *Nurs. Forum*. 53, 399–405. doi: 10.1111/nuf.12266
- Eccleston, C. (2001). Role of psychology in pain management. *Br. J. Anaesth.* 87, 144–152. doi: 10.1093/bja/87.1.144
- Eccleston, C., and Crombez, G. (1999). Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol. Bull.* 125, 356–366. doi: 10.1037/0033-2909.125.3.356
- Faber, A. W., Patterson, D. R., and Bremer, M. (2013). Repeated use of immersive virtual reality therapy to control pain during wound dressing changes in pediatric and adult burn patients. *J. Burn Care Res.* 34, 563–568. doi: 10.1097/BCR.0b013e3182777904
- Fields, H. L. (2018). How expectations influence pain. *Pain* 159(Suppl 1), S3–S10. doi: 10.1097/j.pain.0000000000001272



- Fratianne, R. B., Prensner, J. D., Hutson, M. J., Super, D. M., Yowler, C. J., and Sandley, J. M. (2001). The effect of music-based imagery and musical alternate engagement on the burn debridement process. *J. Burn Care Rehabil.* 22, 47–53. doi: 10.1097/00004630-200101000-00010
- Garrett, B., Taverner, T., Masinde, W., Gromala, D., Shaw, C., and Nagraeff, M. (2014). A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. *Clin. J. Pain* 30, 1089–1098. doi: 10.1097/AJP.0000000000000064
- Hemington, K. S., Cheng, J. C., Bosma, R. L., Rogachov, A., Kim, J. A., and Davis, K. D. (2017). Beyond negative pain-related psychological factors: resilience is related to lower pain affect in healthy adults. *J. Pain* 18, 1117–1128. doi: 10.1016/j.jpain.2017.04.009
- Hendrix, C., and Barfield, W. (1995). "Presence in virtual environments as a function of visual and auditory cues," In *Proceedings of VRAIS'95* (Los Alamitos, CA: Research Triangle Park, NC, IEEE Computer Society Press) 74–82. doi: 10.1109/VRAIS.1995.512482
- Hoffman, H. G. (1998). Virtual reality: a new tool for interdisciplinary psychology research. *CyberPsychol. Behav.* 1, 195–200. doi: 10.1089/cpb.1998.1.195
- Hoffman, H. G., Chambers, G. T., Meyer, W. J., 3rd, Arceneaux, L. L., Russell, W. J., Seibel, E. J., et al. (2011). Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *Ann. Behav. Med.* 41, 183–191. doi: 10.1007/s12160-010-9248-7
- Hoffman, H. G., Doctor, J. N., Patterson, D. R., Carrougner, G. J., and Furness, T. A. III. (2000). Use of virtual reality as an adjunctive treatment of adolescent burn pain during wound care: a case report. *Pain* 85, 305–309. doi: 10.1016/S0304-3959(99)00275-4
- Hoffman, H. G., Garcia-Palacios, A., Kapa, V. A., Beecher, J., and Sharar, S. R. (2003). Immersive virtual reality for reducing experimental ischemic pain. *Int. J. Hum. Comp. Interact.* 15, 469–486. doi: 10.1207/S15327590IJHCI1503\_10
- Hoffman, H. G., Meyer, W. J., 3rd, Ramirez, M., Roberts, L., Seibel, E. J., Atzori, B., et al. (2014). Feasibility of articulated arm mounted Oculus Rift Virtual Reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients. *Cyberpsychol. Behav. Soc. Netw.* 17, 397–401. doi: 10.1089/cyber.2014.0058
- Hoffman, H. G., Patterson, D. R., Carrougner, G. J., and Sharar, S. R. (2001). Effectiveness of virtual reality-based pain control with multiple treatments. *Clin. J. Pain* 17, 229–235. doi: 10.1097/00002508-200109000-00007
- Hoffman, H. G., Patterson, D. R., Magula, J., Carrougner, G. J., Zeltzer, K., Dagadakis, S., et al. (2004a). Water-friendly virtual reality pain control during wound care. *J. Clin. Psychol.* 60, 189–95. doi: 10.1002/jclp.10244
- Hoffman, H. G., Richards, T. L., Coda, B., Bills, A. R., Blough, D., Richards, et al. (2004b). Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. *Neuroreport* 15, 1245–1248. doi: 10.1097/01.wnr.0000127826.73576.91
- Hoffman, H. G., Richards, T. L., Van Oostrom, T., Coda, B. A., Jensen, M. P., Blough, D. K., et al. (2007). The analgesic effects of opioids and immersive virtual reality distraction: evidence from subjective and functional brain imaging assessments. *Anesth. Analg.* 105, 1776–1783. doi: 10.1213/01.ane.0000270205.45146.db
- Hoffman, H. G., Seibel, E. J., Richards, T. L., Furness, T. A., Patterson, D. R., and Sharar, S. R. (2006). Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. *J. Pain* 7, 843–850. doi: 10.1016/j.jpain.2006.04.006
- Hoffman, H. G., Sharar, S. R., Coda, B., Everett, J. J., Ciol, M., Richards, T., et al. (2004c). Manipulating presence influences the magnitude of virtual reality analgesia. *Pain* 111, 162–168. doi: 10.1016/j.pain.2004.06.013
- Honzel, E., Murthi, S., Brawn-Cinani, B., Colloca, G., Kier, C., Varshney, A., et al. (2019). Virtual reality, music and pain: developing the premise for an interdisciplinary approach to pain management. *Pain*. doi: 10.1097/j.pain.0000000000001539. [Epub ahead of print].
- Hsu, K. C., Chen, L. F., and Hsieh, P. H. (2016). Effect of music intervention on burn patients' pain and anxiety during dressing changes. *Burns* 42, 1789–1796. doi: 10.1016/j.burns.2016.05.006
- IBM SPSS (2018). *Statistics 25.0*. Available online at: <https://www.ibm.com/products/spss-statistics>. (accessed July 18, 2019)
- Indovina, P., Barone, D., Gallo, L., Chirico, A., De Pietro, G., and Antonio, G. (2018). Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: a comprehensive literature review. *Clin. J. Pain* 34, 858–877. doi: 10.1097/AJP.0000000000000599
- Jeffs, D., Dorman, D., Brown, S., Files, A., Graves, T., Kirk, E., et al. (2014). Effect of virtual reality on adolescent pain during burn wound care. *J. Burn Care Res.* 35, 395–408. doi: 10.1097/BCR.0000000000000019
- Jensen, M. P. (2003). The validity and reliability of pain measures in adults with cancer. *J. Pain* 4, 2–21. doi: 10.1054/jpai.2003.1
- Jensen, M. P., and Karoly, P. (2001). "Self-report scales and procedures for assessing pain in adults," in *Handbook of Pain Assessment*, 2nd ed., eds Turk, D. C. and Melzack, R. (New York, NY: Guilford Press). 15–34.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Keefe, F. J., Huling, D. A., Coggins, M. J., Keefe, D. F., Zachary Rosenthal, M., Herr, N. R., et al. (2012). Virtual reality for persistent pain: a new direction for behavioral pain management. *Pain* 153, 2163–2166. doi: 10.1016/j.pain.2012.05.030
- Keefe, F. J., Main, C. J., and George, S. Z. (2018). Advancing psychologically informed practice for patients with persistent musculoskeletal pain: Promise, pitfalls, and solutions. *Phys. Ther.* 98, 398–407. doi: 10.1093/ptj/pty024
- Khadra, C., Ballard, A., Déry, J., Paquin, D., Fortin, J. S., Perreault, I., et al. (2018). Projector-based virtual reality dome environment for procedural pain and anxiety in young children with burn injuries: a pilot study. *J. Pain Res.* 14, 343–353. doi: 10.2147/JPR.S151084
- Kipping, B., Rodger, S., Miller, K., and Kimble, R. M. (2012). Virtual reality for acute pain reduction in adolescents undergoing burn wound care: a prospective randomized controlled trial. *Burns* 38, 650–657. doi: 10.1016/j.burns.2011.11.010
- Krane, E. J. (2019). The opioid debate-PRO: opioids have an important role in pain management in children. *Clin. J. Pain* 35, 468–472. doi: 10.1097/AJP.0000000000000700
- Maani, C., Hoffman, H. G., DeSocio, P. A., Morrow, M., Galin, C., Magula, J., et al. (2008). Pain control during wound care for combat-related burn injuries using custom articulated arm mounted virtual reality goggles. *J. CyberTher. Rehabil.* 1, 193–198.
- Maani, C. V., Hoffman, H. G., Fowler, M., Maier, A. J., Gaylord, K. M., and Desocio, P. A. (2011a). Combining ketamine and virtual reality pain control during severe burn wound care: one military and one civilian patient. *Pain Med.* 12, 673–678. doi: 10.1111/j.1526-4637.2011.01091
- Maani, C. V., Hoffman, H. G., Morrow, M., Maier, A., Gaylord, K., McGhee, L. L., et al. (2011b). Virtual reality pain control during burn wound debridement of combat-related burn injuries using robot-like arm mounted VR goggles. *J. Trauma* 71(1 Suppl), S125–S130. doi: 10.1097/TA.0b013e31822192e2
- Malchow, R. J., and Black, I. H. (2008). The evolution of pain management in the critically ill trauma patient: emerging concepts from the global war on terrorism. *Crit. Care Med.* 36, S346–357. doi: 10.1097/CCM.0b013e31817e2fc9
- McCaul, K. D., and Malott, J. M. (1984). Distraction and coping with pain. *Psychol. Bull.* 95, 516–533. doi: 10.1037//0033-2909.95.3.516
- McGhee, L. L., Slater, T. M., Garza, T. H., Fowler, M., DeSocio, P. A., and Maani, C. V. (2011). The relationship of early pain scores and posttraumatic stress disorder in burned soldiers. *J. Burn Care Res.* 32, 46–51. doi: 10.1097/BCR.0b013e318204b359
- McIntyre, M. K., Clifford, J. L., Maani, C. V., and Burmeister, D. M. (2016). Progress of clinical practice on the management of burn-associated pain: Lessons from animal models. *Burns* 42, 1161–1172. doi: 10.1016/j.burns.2016.01.023
- McSherry, T., Atterbury, M., Gartner, S., Helmold, E., Searles, D. M., and Schulman, C. (2018). Randomized, crossover study of immersive virtual reality to decrease opioid use during painful wound care procedures in adults. *J. Burn Care Res.* 39, 278–285. doi: 10.1097/BCR.0000000000000589
- Melzack, R., and Wall, P. D. (1965). Pain mechanisms: a new theory. *Science* 150, 971–979. doi: 10.1126/science.150.3699.971
- Nitzan, U., Hecht, M., Braw, Y., Maoz, H., Levkovitz, Y., Yarnitsky, D., et al. (2019). Initial evaluation of pain intensity among depressed patients as a possible mediator between depression and pain complaints. *Front Psychiatry* 10:48. doi: 10.3389/fpsy.2019.00048
- Noel, M., Rabbitts, J. A., Tai, G. G., and Palermo, T. M. (2015). Remembering pain after surgery: a longitudinal examination of the role of pain catastrophizing in children's and parents' recall. *Pain* 156, 800–808. doi: 10.1097/j.pain.000000000000102

- Pardesi, O., and Fuzaylov, G. (2017). Pain management in pediatric burn patients: review of recent literature and future directions. *J. Burn Care Res.* 38, 335–347. doi: 10.1097/BCR.0000000000000470
- Peña, R., Suman, O. E., Rosenberg, M., Andersen, C. R., Herndon, D. N., and Meyer, W. J. (2017). One-year comparison of a community-based exercise program versus a day hospital-based exercise program on quality of life and mental health in severely burned children. *Arch. Phys. Med. Rehabil.* doi: 10.1016/j.apmr.2017.10.023
- Rohilla, L., Agnihotri, M., Trehan, S. K., Sharma, R. K., and Ghai, S. (2018). Effect of music therapy on pain perception, anxiety, and opioid use during dressing change among patients with burns in India: A quasi-experimental, cross-over pilot study. *Ostomy Wound Manage.* 64, 40–46. doi: 10.25270/owm.2018.10.4046
- Rosenberg, L., Rosenberg, M., Robert, R., Richardson, L., Sharp, S., Holzer, C. E., et al. (2015). Does acute stress disorder predict subsequent posttraumatic stress disorder in pediatric burn survivors? *J. Clin. Psychiatry* 76, 1564–1568. doi: 10.4088/JCP.14m09365
- Rosenberg, L., Rosenberg, M., Sharp, S., Thomas, C. R., Humphries, H. F., and Holzer III, C. E., et al. (2018). Does acute propranolol treatment prevent posttraumatic stress disorder, anxiety, and depression in children with burns? *J. Child Adolesc. Psychopharmacol.* 28, 117–123. doi: 10.1089/cap.2017.0073
- Scheffler, M., Koranyi, S., Meissner, W., Strauß, B., and Rosendahl, J. (2017). Efficacy of non-pharmacological interventions for procedural pain relief in adults undergoing burn wound care: a systematic review and meta-analysis of randomized controlled trials. *Burns* 44, 1709–1720. doi: 10.1016/j.burns.2017.11.019
- Schwaller, F., and Fitzgerald, M. (2014). The consequences of pain in early life: injury-induced plasticity in developing pain pathways. *Eur. J. Neurosci.* 39, 344–352. doi: 10.1111/ejn.12414
- Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in immersive virtual environments. *Presence Teleoper. Virtual Environ.* 3, 130–144. doi: 10.1162/pres.1994.3.2.130
- Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence Teleoper. Virtual Environ.* 6, 603–616. doi: 10.1162/pres.1997.6.6.603
- Sullivan, M. J. L., Bishop, S. R., and Pivik, J. (1995). The pain catastrophizing scale: development and validation. *Psychol. Assess.* 7, 524–532. doi: 10.1037//1040-3590.7.4.524
- Tesler, M. D., Savedra, M. C., Holzemer, W. L., Wilkie, D. J., Ward, J. A., and Paul, S. M. (1991). The word-graphic rating scale as a measure of children's and adolescents' pain intensity. *Res. Nurs. Health* 14, 361–371. doi: 10.1002/nur.4770140507
- van der Heijden, M. J. E., Jeekel, J., Rode, H., Cox, S., van Rosmalen, J., Hunink, M. G. M., et al. (2018). Can live music therapy reduce distress and pain in children with burns after wound care procedures? a randomized controlled trial. *Burns* 44, 823–833. doi: 10.1016/j.burns.2017.12.013
- Van Loey, N. E., Klein-König, I., de Jong, A. E. E., Hofland, H. W. C., Vandermeulen, E., and Engelhard, I. M. (2018). Catastrophizing, pain and traumatic stress symptoms following burns: A prospective study. *Eur. J. Pain* 22, 1151–1159. doi: 10.1002/ejp.1203
- van Twillert, B., Bremer, M., and Faber, A. W. (2007). Computer-generated virtual reality to control pain and anxiety in pediatric and adult burn patients during wound dressing changes. *J. Burn Care Res.* 28, 694–702. doi: 10.1097/BCR.0B013E318148C96F
- Verhoeven, K., Goubert, L., Jaaniste, T., Van Ryckeghem, D. M., and Crombez, G. (2012). Pain catastrophizing influences the use and the effectiveness of distraction in schoolchildren. *Eur. J. Pain* 16, 256–267. doi: 10.1016/j.ejpain.2011.06.015
- Vijayan, R. (2011). Managing acute pain the developing world. *Pain Clin Updates.* 19, 1–7.
- Wender, R., Hoffman, H. G., Hunner, H. H., Seibel, E. J., Patterson, D. R., and Sharar, S. R. (2009). Interactivity influences the magnitude of virtual reality analgesia. *J. Cyber Ther. Rehabil.* 2, 27–33.

**Conflict of Interest Statement:** All authors have completed the ICMJE uniform disclosure and declare support for the submitted work.

Copyright © 2019 Hoffman, Rodriguez, Gonzalez, Bernardy, Peña, Beck, Patterson and Meyer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Technological Competence Is a Pre-condition for Effective Implementation of Virtual Reality Head Mounted Displays in Human Neuroscience: A Technological Review and Meta-Analysis

Panagiotis Kourtesis<sup>1,2,3,4\*</sup>, Simona Collina<sup>3,4</sup>, Leonidas A. A. Doulmas<sup>2</sup> and Sarah E. MacPherson<sup>1,2</sup>

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Camila Rosa De Oliveira,  
Faculdade Meridional (IMED), Brazil  
Uri Maoz,  
Chapman University, United States

### \*Correspondence:

Panagiotis Kourtesis  
pkourtes@exseed.ed.ac.uk

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 27 June 2019

**Accepted:** 18 September 2019

**Published:** 02 October 2019

### Citation:

Kourtesis P, Collina S, Doulmas LAA  
and MacPherson SE (2019)  
Technological Competence Is a  
Pre-condition for Effective  
Implementation of Virtual Reality Head  
Mounted Displays in Human  
Neuroscience: A Technological  
Review and Meta-Analysis.  
Front. Hum. Neurosci. 13:342.  
doi: 10.3389/fnhum.2019.00342

<sup>1</sup> Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom,  
<sup>2</sup> Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, <sup>3</sup> Lab of Experimental Psychology,  
Suor Orsola Benincasa University of Naples, Naples, Italy, <sup>4</sup> Interdepartmental Centre for Planning and Research  
"Scienza Nuova", Suor Orsola Benincasa University of Naples, Naples, Italy

Immersive virtual reality (VR) emerges as a promising research and clinical tool. However, several studies suggest that VR induced adverse symptoms and effects (VRISE) may undermine the health and safety standards, and the reliability of the scientific results. In the current literature review, the technical reasons for the adverse symptomatology are investigated to provide suggestions and technological knowledge for the implementation of VR head-mounted display (HMD) systems in cognitive neuroscience. The technological systematic literature indicated features pertinent to display, sound, motion tracking, navigation, ergonomic interactions, user experience, and computer hardware that should be considered by the researchers. Subsequently, a meta-analysis of 44 neuroscientific or neuropsychological studies involving VR HMD systems was performed. The meta-analysis of the VR studies demonstrated that new generation HMDs induced significantly less VRISE and marginally fewer dropouts. Importantly, the commercial versions of the new generation HMDs with ergonomic interactions had zero incidents of adverse symptomatology and dropouts. HMDs equivalent to or greater than the commercial versions of contemporary HMDs accompanied with ergonomic interactions are suitable for implementation in cognitive neuroscience. In conclusion, researchers' technological competency, along with meticulous methods and reports pertinent to software, hardware, and VRISE, are paramount to ensure the health and safety standards and the reliability of neuroscientific results.

**Keywords:** virtual reality, VRISE, HMD, cybersickness, neuroscience, neuropsychology, psychology, VR

## INTRODUCTION

In recent years, virtual reality (VR) technology has attracted attention, demonstrating its utility and potency in the field of neuroscience and neuropsychology (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). Traditional approaches in human neuroscience involve the utilization of static and simple stimuli which arguably lack ecological validity (Parsons, 2015). VR offers the usage of dynamic stimuli and interactions with a high degree of control within an ecologically valid environment which enables the collection of advanced cognitive and behavioral data (Rizzo et al., 2004; Bohil et al., 2011; Parsons, 2015). VR can be combined with non-invasive imaging techniques (Bohil et al., 2011; Parsons, 2015) and has been effective in the assessment of cognitive and affective functions and clinical conditions (e.g., social stress disorders) which require ecological validity (Rizzo et al., 2004; Parsons, 2015) for their assessment, rehabilitation and treatment (e.g., post-traumatic stress disorder) (Rizzo et al., 2004; Bohil et al., 2011).

However, researchers and clinicians have reported caveats with the implementation of immersive VR interventions and assessments, particularly when head mounted display (HMD) systems are utilized (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017). A predominant concern is the presence of adverse physiological symptoms (i.e., cyber/simulation-sickness which includes nausea, disorientation, instability, dizziness, and fatigue). These undesirable effects are categorized as VR Induced Symptoms and Effects (VRISE) (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017), and are evaluated by using questionnaires such as the Simulator Sickness Questionnaire (Kennedy et al., 1993) and the Virtual Reality Sickness Questionnaire (Kim et al., 2018).

VRISE may risk the health and safety of participants or patients (Kane and Parsons, 2017; Parsons et al., 2018), which raises ethical considerations for the adoption of VR HMDs as research and clinical tools. Additionally, the presence of VRISE has modulated substantial decline in reaction times and overall cognitive performance (Plant and Turner, 2009; Nalivaiko et al., 2015; Plant, 2016; Nesbitt et al., 2017; Mittelstaedt et al., 2018), as well as increasing body temperature and heart rates (Nalivaiko et al., 2015). Also, the presence of VRISE robustly increases cerebral blood flow and oxyhemoglobin concentration (Gavani et al., 2018), the power of brain signals (Arafat et al., 2018), and the connectivity between stimulus response brain regions and nausea-processing brain regions (Toschi et al., 2017). Thus, VRISE could be considered confounding variables, which significantly undermine the reliability of neuropsychological, physiological, and neuroimaging data.

VRISE are predominantly mediated by an oculomotor discrepancy between what is being perceived through the oculomotor (optic nerve) sensor and what is being sensed via the rest of the afferent nerves in the human body (Sharples et al., 2008; Davis et al., 2015; de França and Soares, 2017; Palmisano et al., 2017). Nevertheless, technologically speaking, VRISE are derivatives of hardware and software inadequacies, i.e., the type of display screen, resolution, and refresh rate of the image,

the size of the field of view (FOV) as well as non-ergonomic movements within an interaction in the virtual environment (VE; de França and Soares, 2017; Palmisano et al., 2017). Notably, VR HMDs have substantially evolved during the last two decades. Important differences may be seen between the HMDs released before 2013 (old generation) and those released from 2013 onwards (new generation). While the last old generation HMD was released in 2001 (i.e., nVisor SX111), the year 2013 is used to distinguish between old and new generation HMDs, since it is the year that the first new generation HMD prototype (i.e., Oculus Development Kit 1) was released. This systematic review attempts to clarify the technological etiologies of VRISE and provide pertinent suggestions for the implementation of VR HMDs in cognitive neuroscience and neuropsychology. In addition, a meta-analysis of the neuroscience studies that have implemented VR HMDs will be conducted to elucidate the frequency of VRISE and dropout rates as per the VR HMD generation.

## TECHNOLOGICAL SYSTEMATIC REVIEW

In **Table 1**, a glossary of the key terms and concepts is provided to assist with comprehension of the utilized terminology. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines using a decremental stepwise method to perform the literature review (see **Figure 1**). The selected papers and book chapters included an explicit explanation and discussion of VRISE and users' experiences pertinent with the specified technological features of the VR hardware and software. Digital databases specialized in technologies were used: (1) IEEE Xplore Digital Library; (2) ACM Digital Library; (3) ScienceDirect; (4) MIT CogNet; and (5) Scopus. Two categories of keywords were used, where each category had three or more keywords and each paper had to include at least one keyword from each category in the main body of the text. The categories were: (1) "virtual reality" OR "immersive virtual reality" OR "head-mounted display;" AND (2) "VRISE" OR "motion sickness" OR "cyber sickness" OR "simulation sickness." Finally, the extracted information from the identified papers was clustered together under common features (i.e., display, sound, motion tracking, navigation, ergonomic interactions, user experience, and computer hardware).

### Technological Etiologies of VRISE Display

VR HMDs use the following three types of screens: Cathode Ray Tubes (CRT); Liquid Crystal Display (LCD); and organic light emitting diode (OLED). LCD screens replaced CRT ones due to VRISE (Costello, 1997). LCD, in comparison to CRT, alleviated the probability of visual complications and physical burdens (e.g., fatigue) (Costello, 1997). However, the suitability of LCD was challenged by the emergence of OLED screens. While old generation VR HMDs mainly utilize LCD screens (Costello, 1997), the commercial versions of new generation VR HMDs predominantly use OLED screens (Kim J. W. et al., 2017). The OLED screens have been found to be

**TABLE 1 |** Glossary of key terms and concepts.

	Terms and concepts	Explanation/definition
Headsets	Head mounted display (HMD)	A display device which is worn on the head and provides an immersive virtual reality for the wearer
	Development kit (DK) HMD	A prototype device, which is utilized by the VR Software developers to develop VR software before the commercial version of an HMD. The DKs are not provided for general use
	Commercial version (CV) HMD	The final version of an HMD, which is dispersed to the market for general use
Display	Liquid crystal display (LCD)	A type of display/screen that uses the light-modulating properties of liquid crystals. Liquid crystals emit light indirectly, instead of using a reflector to produce images
	Organic light emitting diode (OLED)	A type of display/screen that uses an organic compound film that emits light in response to an electric current. OLEDs are used as displays in devices such as television screens, computer monitors, and smartphones
	Field of view (FOV)	The area captured by the display device. The size of the FOV and the size of the display device directly affect the quality of the image
	Refresh rate and frame rate	The refresh rate is the number of times that the hardware updates its display per second. It involves the repeated display of identical frames. The frame rate indicates the frequency that software can add new data to a display
	Resolution	The number of distinct pixels in each dimension displayed in a frame
Interactions	Motion tracking	The process of tracking the movement of objects or people. It is facilitated by motion sensors which detect the position of motion trackers embedded in devices (e.g., HMDs and 6DoF controllers)
	Controllers/Wands with 6 degrees of freedom (DoF)	Controllers which have 6DoF of movement in 3-D space on three directional axes (i.e., Forward-Back, Left-Right, Up-Down) and three rotational axes (i.e., Roll, Pitch, Yaw)
	Direct hand interaction	A motion tracking device (i.e., a motion sensor) which directly tracks hand movements
	Teleportation	A navigation system, which allows the user to be transferred to a new location in the virtual environment without physically moving in the real environment
	Ergonomic interactions	These resemble real-life interactions, which optimize user experience and overall VR system performance (see also Definition of Ergonomic Interactions). Ergonomic interactions are facilitated by and restricted to the capabilities of the VR hardware and software
	Virtual environment (VE)	A three-dimensional artificial environment which is displayed on a display device and allows the users to interact with it

better than LCD screens for general implementation in VR, because of their faster response times, lighter weight, and better color quality (Kim J. W. et al., 2017). OLED screens decrease the likelihood of VRSE and offer an improved VR display (Kim J. W. et al., 2017).

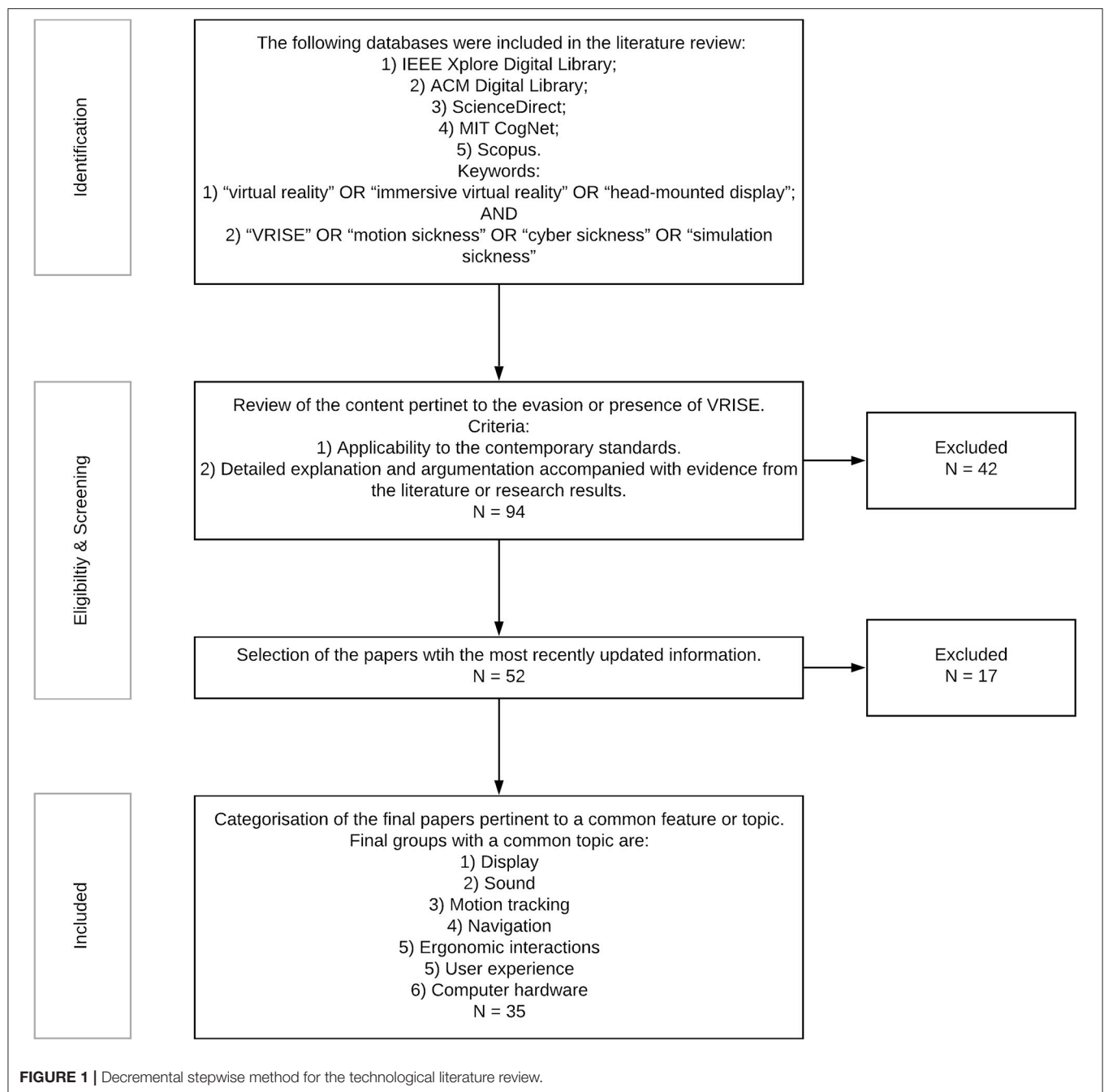
Three more factors related to display type are crucial for the avoidance of VRSE: the width of the FOV (Rakkolainen et al., 2016; Kim J. W. et al., 2017); the resolution of the image per eye (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018); and the latency of the images (frames per second) (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). A wider FOV significantly decreases the chance of VRSE and increases the level of immersion (Rakkolainen et al., 2016; Kim J. W. et al., 2017). The canonical guidelines suggest a lowest threshold of 110° FOV (diagonal) (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). In addition, an increased refresh rate and resolution alleviates the danger of discomfort or VRSE (Hecht, 2016; Rakkolainen et al., 2016; Kim J. W. et al., 2017; Brennesholtz, 2018). The refresh rate should be  $\geq 75$  Hz (i.e.,  $\geq 75$  frames per s) (Goradia et al., 2014; Hecht, 2016; Brennesholtz, 2018), while the resolution is required to be higher than  $960 \times 1,080$  sub-pixels per eye (Goradia et al., 2014).

## Sound

A second important consideration for a user's experience in VR is the sound quality. The integration of spatialized sounds (e.g., ambient and feedback sounds) in the VE may increase the level of immersion, pleasantness of the experience, and successful navigation (Vorländer and Shinn-Cunningham, 2014), while they significantly decrease the likelihood of VRSE (Viirre et al., 2014). However, the volume and localization of sounds need to be optimized in terms of audio spatialization to ensure a user's experience is pleasant without adverse VRSE (Viirre et al., 2014; Vorländer and Shinn-Cunningham, 2014).

## Motion Tracking

Motion tracking in VR is a pre-condition for naturalistic movement within an immersive VE (Slater and Wilbur, 1997; Stanney and Hale, 2014). Motion tracking allows the precise tracking of the user's physical body within the VE (i.e., it allows the computer to provide accurate environmental feedback, which modulates and consolidates the awareness of the position and movement of the user's body). This phenomenon is called proprioception or kinesthesia (Slater and Wilbur, 1997) and is linked with vestibular and oculomotor mediated VRSE (Slater and Wilbur, 1997; Plouzeau et al., 2015; Caputo et al., 2017). Hence, motion tracking should be adequately



**FIGURE 1 |** Incremental stepwise method for the technological literature review.

rapid and accurate to facilitate ergonomic interactions in the VE (Caputo et al., 2017).

### Navigation

A highly important factor in the quality of VR software and to avoid VRISE is the movement of the user in the VE (Porcino et al., 2017). New generation HMDs deliver an adequate play area for interactions to facilitate ecologically valid scenarios (Porcino et al., 2017; Borrego et al., 2018). However, there are restrictions in the size of the play area, which does not permit navigation solely by physical walking

(Porcino et al., 2017; Borrego et al., 2018). Teleportation allows movement beyond the play area size and elicits a high-level of immersion and pleasant user experience, whilst alleviating VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). In contrast, movement dependent on a touchpad, keyboard, or joystick results to high occurrences of VRISE (Bozgeyikli et al., 2016; Frommel et al., 2017; Porcino et al., 2017). Therefore, teleportation in conjunction with physical movement (i.e., free movement of the upper limbs and walking in a small-restricted area) is the most suitable method for movement in VR (Bozgeyikli et al., 2016;

**TABLE 2 |** Minimum hardware criteria: old and new generation VR HMDs.

Product	Generation	Resolution (per eye)	Display screen	Refresh rate	FOV (Diagonal)	Motion trackers and sensors (Type and quantity)
VFX 3D	Old	480 × 240	LCD	45 Hz	45°	–
VUZIX Wrap 1200	Old	852 × 480	LCD	60 Hz	35°	Unknown type (1), 3 magnetometers, 3 accelerometers, and 3 gyroscopes
eMagin Z800 3DVisor	Old	800 × 600	OLED	60 Hz	40°	–
nVisor SX111	Old	1,280 × 1,024	LCD	60 Hz	110°	–
Oculus rift development kit 1	New	640 × 800	LCD	60 Hz	110°	–
Oculus rift development kit 2	New	960 × 1,080	OLED	75 Hz	110°	–
Minimum hardware criteria for the avoidance of VRSE	NA	>960 × 1,080	OLED or LCD	≥75Hz	≥110°	Tracking should be adequately rapid and accurate to facilitate ergonomic interactions
Oculus rift commercial version	New	1,080 × 1,200	OLED	90 Hz	110°	Accelerometer, gyroscope, magnetometer, 360° constellation tracking camera
HTC VIVE commercial version	New	1,080 × 1,200	OLED	90 Hz	110°	Sensors (>70) including MEMS, magnetometer, gyroscope, accelerometer, and laser position sensors, lighthouse laser tracking system (2 base stations emitting pulsed InfraRed lasers), front-facing camera

MEMS, Microelectromechanical systems.

**TABLE 3 |** Criteria for suitable VR software in cognitive neuroscience and neuropsychology.

Domains	User experience	Game mechanics	In-game assistance	VRSE
Criteria	An adequate level of immersion	A suitable navigation system (e.g., Teleportation)	Digestible tutorials	Absence or insignificant presence of nausea
	Pleasant VR experience	Availability of physical movement	Helpful tutorials	Absence or insignificant presence of disorientation
	High quality graphics	Naturalistic picking/placing of items	Adequate duration of tutorials	Absence or insignificant presence of dizziness
	High quality sounds	Naturalistic use of items	Helpful in-game instructions	Absence or insignificant presence of fatigue
	Suitable hardware (HMD and computer)	Naturalistic 2-handed interaction	Helpful in-game prompts	Absence or insignificant presence of instability

Frommel et al., 2017; Porcino et al., 2017). Yet, there are additional factors such as external hardware (i.e., controllers and wands), which are needed to facilitate optimal ergonomic interactions in VR.

## Ergonomic Interactions

Ergonomic and naturalistic interactions are essential to minimize the risk of VRSE, while non-ergonomic and non-naturalistic interactions increase the occurrence of them (Slater and Wilbur, 1997; Stanney and Hale, 2014; Plouzeau et al., 2015; Caputo et al., 2017; Porcino et al., 2017). Importantly, controllers, joysticks, and keyboards do not support ergonomic and naturalistic interactions in VR (Plouzeau et al., 2015; Bozgeyikli et al., 2016; Caputo et al., 2017; Frommel et al., 2017; Porcino et al., 2017; Sportillo et al., 2017; Figueiredo et al., 2018). Instead, wands with 6 degrees of freedom (DoF) of movement (e.g., Oculus Rift and HTC Vive wands), and realistic interfaces with direct hand interactions (e.g., Microsoft's Kinect) facilitate

naturalistic and ergonomic interactions (Sportillo et al., 2017; Figueiredo et al., 2018). Both hardware systems facilitate easy familiarization with their controls and their utilization (Sportillo et al., 2017; Figueiredo et al., 2018). However, direct hand interactions are easier than 6DoF controllers-wands in terms of familiarization with their controls and efficiency (Sportillo et al., 2017; Figueiredo et al., 2018). Direct hand interactions were also found to offer more pleasant user experiences (Sportillo et al., 2017; Figueiredo et al., 2018), although, they are substantially less accurate than 6DoF controller-wands (Sportillo et al., 2017; Figueiredo et al., 2018).

## User Experience

Notably, ergonomic interactions might be available to the user; however, the user is required to learn the necessary interactions and how the VE functions to facilitate a pleasant user experience (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). The inclusion of comprehensible tutorials where the user may



spend an adequate amount of time acquiring the necessary skills (i.e., navigation, use and grab of items, two-handed interactions) and knowledge of the VE (i.e., how it reacts to your controls) is crucial (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018). Additionally, in-game instructions and prompts should be offered to the user through interactions in the VE (e.g., directional arrows, non-player characters, signs, labels, ambient sounds, audio, and videos) (Gromala et al., 2016; Jerald et al., 2017; Brade et al., 2018).

### Computer Hardware

The computer hardware (i.e., the processor, graphics card, sound card) should at least meet the minimum requirements of the VR software and HMD (Anthes et al., 2016). The performance of VR HMDs is analogous to the computing power and the quality of the hardware (Stanney and Hale, 2014; Anthes et al., 2016; Borrego et al., 2018). The processor, graphics card, sound card, and operating system (e.g., Windows) need to be considered and reported because they modulate the performance of the software (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018). Research software developers and researchers are required to be technologically competent in order to opt for the appropriate hardware and software to achieve their research and/or clinical aims (Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018).

### Conclusions

Based on the outcomes of the above technological review, VR HMDs should have a good quality display-screen (i.e., OLED or upgraded LCD), an adequate FOV (i.e., diagonal FOV  $\geq 110^\circ$ ), adequate resolution per eye (i.e., resolution  $> 960 \times 1,080$  sub-pixels per eye), and an adequate image refresh rate (i.e., refresh rate  $\geq 75$  Hz) to safeguard the health and safety of the participants and the reliability of the neuroscientific results (see **Table 2**). Also, the VR HMD should have external hardware which offers an adequate VR area, fast and accurate motion tracking, spatialized audio, and ergonomic interactions. The computer's processor, graphics card, and sound card should meet the minimum requirements of the VR software and HMD too. New generation VR HMDs appear to have all the necessary hardware characteristics (i.e., graphics, level of immersion, and sound) to be used in ecological valid research and clinical paradigms (Borrego et al., 2018; see **Table 2** for a comparison between old and new generation HMDs). New generation VR HMDs have the required hardware to support and produce high-quality spatialized sounds in VEs (Borrego et al., 2018). Additionally, new generation VR HMDs have integrated rapid and precise motion tracking which facilitates naturalistic and ergonomic interactions within the VE (Borrego et al., 2018).

Both the Oculus development kit (DK) 1 and DK2 do not meet the minimum hardware features highlighted by the technological review, despite being new generation VR HMDs (see **Table 2**). The DK1 has substantially lower resolution per eye and image refresh rates, while the DK2 has marginally acceptable refresh rates, yet a slightly lower resolution per eye. These DKs are not available for general use but are used by professional

developers to produce beta (early) versions of their games or apps (Goradia et al., 2014; Suznjevic et al., 2017). Moreover, they were removed from the market after the release of the Oculus Rift CV. VR HMDs should have hardware characteristics equal to or better than the commercial versions (CV) of the Oculus Rift and HTC Vive in order to ensure the health and safety of the participants, as well as the reliability of the neuroscientific results (i.e., physiological, neuropsychological, and neuroimaging data). The researchers and clinicians should have the technological competence to choose an HMD which is equal to or greater than the CVs of the Oculus Rift and HTC Vive (e.g., Valve Index, HTC Vive Pro, Oculus Quest, Pimax VR, and StarVR).

However, the VR software's features are equally important. The VR software should include an ergonomic interaction and navigation system, as well as tutorials, in-game instructions, and prompts. A suitable navigation system should combine teleportation and physical movement, while ergonomic interactions should include those that simulate real-life interactions by using a direct hands system or 6DoF controllers. Also, the tutorials, in-game instructions, and prompts should be informative and easy to follow, especially for experimental or clinical purposes where users should be equally able to interact with the VE (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018). The criteria for effective VR software are displayed in **Table 3**. These criteria should be met before implementing VR software for research and/or clinical purposes. Otherwise, researchers or clinicians may compromise the reliability of their study's results (Plant and Turner, 2009; Plant, 2016; Kane and Parsons, 2017; Parsons et al., 2018), and/or jeopardize the health and safety of their participants/patients (Kane and Parsons, 2017; Parsons et al., 2018).

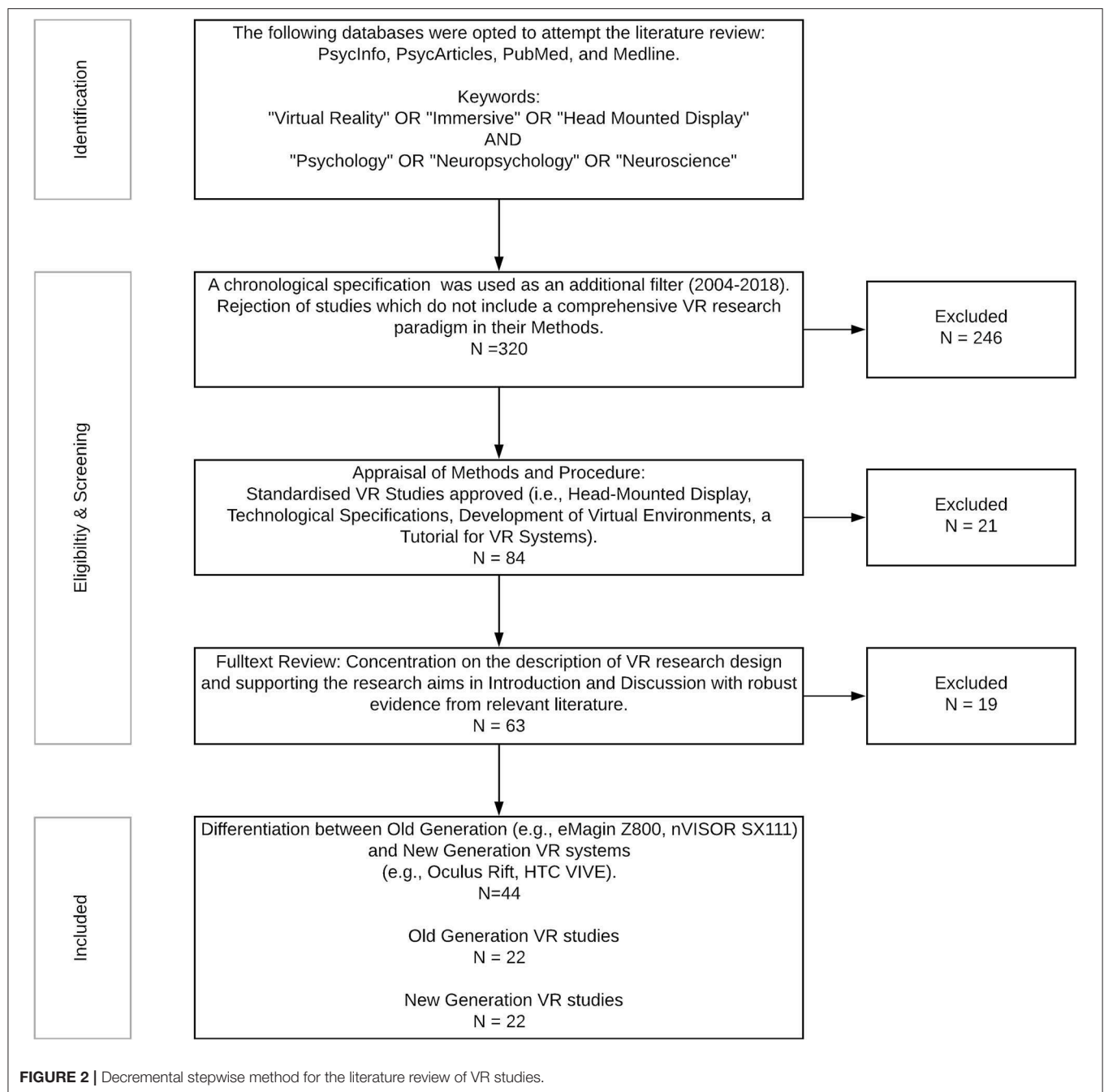
The above features enable researchers or clinicians to administer a sophisticated and pleasant VR experience, which substantially alleviates or eradicates adverse VRSE. Therefore, the technological competency of neuroscientists and neuropsychologists is a pre-condition for the efficient adoption and implementation of innovative technologies like VR HMDs in cognitive neuroscience or neuropsychology.

## META-ANALYSIS OF VR STUDIES IN COGNITIVE NEUROSCIENCE

### Literature Research and Inclusion Criteria

We followed the PRISMA guidelines to conduct the literature research using a decremental approach, where the selection commenced with a relatively vast accumulation of abstracts and concluded with a diminished list of full papers that comprise standardized and detailed VR research paradigms. The procedure is described in **Figure 2**. The following databases were used for the literature research: (1) PsycInfo; (2) PsycArticles; (3) PubMed; and (4) Medline. Two categories of keywords were used, with three keywords in each category. The minimum threshold for each study was the inclusion of at least one keyword from each category in the main body of text. The keywords





for each category were: (1) "virtual reality;" OR "Immersive;" OR "Head Mounted Display;" AND (2) "Psychology;" OR "Neuropsychology" OR "Neuroscience." Additional filters and criteria were: (1) chronological specification (2004 and later); and (2) a comprehensive description of the VR research methods in conjunction with the research aims and results. Finally, the selected studies were allocated into two groups according to the generation of the implemented VR HMD. Two tables display the studies that utilize old generation (**Table 4**) and new generation (**Table 5**) HMDs.

## Data Collection and Coding Target Variables

The principal aim of the meta-analysis was to measure the frequency of VRIFE in neuroscience or psychology studies using a VR HMD. However, only six studies reported VRIFE quantitatively (i.e., using a questionnaire). For this reason, we considered only the presence or absence of VRIFE. The dichotomous VRIFE variable (i.e., presence or absence of VRIFE) was quantified (i.e., absent VRIFE = 0; present VRIFE = 1) to facilitate a comparison (i.e., Bayesian *t*-tests) between the studies

**TABLE 4 |** Neuroscience studies employing old generation VR HMDs.

References	Topic	HMD	Ergonomic interactions	Clinical condition	Age group	N	VRISE	Dropouts
			YES = 12 NO = 10				YES = 14 NO = 8	YES = 9 NO = 13
Kim et al. (2004)	Visuospatial functions	Eye-trek FMD-250W	YES	Brain injury	MA	52	YES	NO - 0
Moreau et al. (2006)	Executive functions	Eye-trek FMD-250W	YES	ADHD and autism	YA	22	YES	YES - 1
Botella et al. (2007)	Therapy (VRET)	V6 VR	NO	Panic disorder	YA and MA	46	YES	YES - 9
Matheis et al. (2007)	Memory	eMagin z800	YES	Brain injury	MA	40	NO	NO - 0
Parsons et al. (2007)	Executive functions	eMagin z800	NO	ADHD	C	20	YES	YES - 1
Banville et al. (2010)	Memory	eMagin z800	NO	Brain injury	YA	62	NO	NO - 0
Rizzo et al. (2010)	Therapy (VRET)	eMagin z800	NO	PTSD	YA	20	YES	YES - 5
Reger et al. (2011)	Therapy (VRET)	eMagin z800	NO	PTSD	YA	24	YES	YES - 6
Bioulac et al. (2012)	Executive functions	eMagin z800	YES	ADHD	YA	36	NO	NO - 0
Carlozzi et al. (2013)	Rehabilitation	eMagin z800	YES	Spinal cord injury	MA	54	YES	YES - 10
Meyerbroeker et al. (2013)	Therapy (VRET)	nVISOR SX111	YES	Agoraphobia	MA	55	YES	YES - 17
Parsons et al. (2013)	Attention assessment	eMagin Z800	YES	Healthy	YA	50	YES	NO - 0
Peck et al. (2013)	Racial biases	nVISOR SX111	YES	Healthy	YA	60	NO	NO - 0
Freeman et al. (2014)	Social cognition	nVISOR SX111	YES	Paranoia	YA	60	YES	NO - 0
Rothbaum et al. (2014)	Therapy (VRET)	eMagin Z800	NO	PTSD	YA and MA	156	NO	NO - 0
Veling et al. (2014)	Paranoid thoughts	eMagin Z800	NO	Psychosis	YA	41	YES	NO - 0
Hartanto et al. (2014)	Social stress	eMagin Z800	NO	Healthy	YA	54	NO	NO - 0
Gaggioli et al. (2014)	Stress levels	Vuzix Wrap 1200VR	NO	Healthy	MA	121	YES	NO - 0
Shiban et al. (2015)	Therapy (VRET)	eMagin Z800	NO	Arachnophobia	YA	58	YES	YES - 8
Freeman et al. (2016)	Therapy (VRET)	nVISOR SX111	YES	Persecutory delusions	MA	30	YES	YES - 1
Parsons and Carlew (2016)	Attention assessment	eMagin Z800	YES	Healthy	YA	50	NO	NO - 0
Parsons and Barnett (2017)	Attention assessment	eMagin Z800	YES	Healthy	YA and OA	89	NO	NO - 0

HMD, Head-Mounted Display; VRISE, VR induced adverse symptoms and effects; YA, Young Adults; MA, Middle-Aged Adults; OA, Older Adults; C, Children; VRET, VR Exposure Therapy; PTSD, Post-Traumatic Stress Disorder; ADHD, Attention Deficit Hyperactivity Disorder.

that used old generation HMDs, new generation DK HMDs, and new generation CV HMDs, as well as the examination of potential correlations with other variables (i.e., Bayesian Pearson's correlation analysis).

secondary aim of the meta-analysis was to inspect the dropout rates in neuroscience or psychology studies that used VR HMDs. However, as the vast majority of studies had no dropouts, studies with some dropouts (e.g., 3, 5,

6) were statistically considered as outliers. For this reason, we considered the existence of dropouts in each study. The dropout variable was dichotomized as presence = 1 and absence = 0. This dichotomous dropout variable was used to investigate whether using a certain generation HMD (i.e., old generation HMDs, new generation DK HMDs, or new generation CV HMDs) could increase/decrease the dropout size. We compared (i.e., Bayesian *t*-tests)

**TABLE 5 |** Neuroscience studies employing new generation VR HMDs.

References	Topic	HMD	Ergonomic interactions	Clinical condition	Age group	N	VRISE	Dropouts
			YES = 18 NO = 4				YES = 4 NO = 18	YES = 4 NO = 18
Foerster et al. (2016)	Attention assessment	Oculus DK2	NO	Healthy	YA	44	NO	YES - 2
Quinlivan et al. (2016)	Attention assessment	Oculus DK2	YES	Healthy	YA	40	NO	NO - 0
Kim A. et al. (2017)	VR presence	Oculus DK2	YES	PD	OA	33	NO	NO - 0
Montenegro and Argyriou (2017)	Memory, attention, executive functions	Oculus DK2	YES	AD (early stages)	OA	20	NO	NO - 0
Parsons and McMahan (2017)	Memory assessment	HTC vive	YES	Healthy	YA	103	NO	NO - 0
Kelly et al. (2017)	Spatial perception	HTC vive	YES	Healthy	YA	76	NO	NO - 0
Bourdin et al. (2017)	Fear of death	Oculus DK2	YES	Healthy	YA	36	NO	NO - 0
Hasler et al. (2017)	Racial bias	Oculus DK2	YES	Healthy	YA	36	NO	NO - 0
Mottelson and Hornbæk (2017)	Navigation, attention, B-P	HTC vive	YES	Healthy	YA and MA	31	NO	NO - 0
Rooney et al. (2017)	Social cognition	Oculus rift CV	YES	Healthy	YA and MA	103	NO	NO - 0
Zimmer et al. (2018)	Social stress	Oculus DK2	NO	Healthy	YA and MA	93	YES	YES - 5
Hsieh et al. (2018)	Spatial perception and navigation	HTC vive	YES	Healthy	YA	70	NO	NO - 0
Yeh et al. (2018)	Anxiety	HTC vive	YES	Healthy	YA	34	NO	NO - 0
Collins et al. (2018)	Psychoeducation on DBS	Oculus rift CV	YES	Movement disorder	OA	30	NO	NO - 0
Barberia et al. (2018)	Fear of death	Oculus DK2	YES	Healthy	YA	31	YES	YES - 1
Banakou et al. (2018)	Embodiment, cognition—IQ	HTC vive	YES	Healthy	YA	30	NO	NO - 0
Christou et al. (2018)	Motor-rehabilitation	HTC vive	YES	Stroke patients	YA and MA	29	NO	NO - 0
Gómez-Jordana et al. (2018)	Balance and walking rehabilitation	Oculus DK2	YES	PD	OA	22	NO	NO - 0
Lubetzky et al. (2018)	Sensory integration and balance	Oculus DK2	NO	Healthy	YA and MA	21	YES	NO - 0
Oagaz et al. (2018)	Memory assessment	HTC vive	YES	Healthy	YA	20	NO	NO - 0
George et al. (2018)	Working memory and attention assessment	HTC vive	YES	Healthy	YA	20	NO	NO - 0
Detez et al. (2019)	Gambling	HTC vive	NO	Healthy	YA and MA	60	YES	YES - 3

*HMD, Head-Mounted Display; VRISE, VR induced adverse symptoms and effects; YA, Young Adults; MA, Middle-Aged Adults; OA, Older Adults; PD, Parkinson's disease; AD, Alzheimer's disease; DK, Development Kit; CV, Commercial Version; B-P, Body Perception; DBS, Deep Brain Stimulation.*

the dropout rate across studies that used old generation HMDs, new generation DK HMDs, and new generation CV HMDs. We also inspected whether the dropout rates correlated with other variables by using Bayesian Pearson's correlation analysis.

### Grouping Variables

We subdivided studies into groups based on the HMD generation they used. Hence, two groups of studies were created and compared by using Bayesian *t*-tests; the first group included studies that utilized old generation HMDs, while the second

group included studies which utilized new generation HMDs (i.e., both DKs and CVs).

The new generation studies were further distinguished and compared by using Bayesian *t*-tests based on the type of new generation HMDs adopted (i.e., DK or CV). Two sub-groups were formed; the first group included studies that utilized DK HMDs, and the second group included studies that utilized CV HMDs.

Furthermore, the recency of the HMD technology was compared by using an ordinal variable where 1 indicated old generation HMDs, 2 indicated new generation DKs, and 3 indicated new generation CVs. This ordinal variable allowed us to inspect whether the HMD generation correlated with other variables by using Bayesian Pearson's correlation analysis.

Lastly, we considered the type of interactions, where the type of interactions were expressed in a binary form (i.e., non-ergonomic interactions = 0 and ergonomic interactions = 1). This allowed a comparison between the VR studies which had ergonomic interactions and the VR studies which had non-ergonomic interactions by using a Bayesian *t*-test. It also allowed us to inspect whether the interaction type correlated with other variables (i.e., Bayesian Pearson's correlation analysis).

### Definition of Ergonomic Interactions

In line with the definition of ergonomic interactions in our technological review, we considered interactions to be ergonomic or non-ergonomic based on their proximity to real-life interactions. We provide some examples below to clarify our criteria:

Example 1—Ergonomic Interaction: if the VR software required the participant to look around moving his or her head.

Example 2—Non-Ergonomic Interaction: if the VR software required the participant to look around by using a joystick or mouse.

Example 3—Ergonomic Interaction: if the VR software required the user to interact with objects (e.g., pushing a button, holding an item) in the VE or to navigate within the VE by using either 6DoF controllers or direct-hand interactions.

Example 4—Non-Ergonomic Interaction: if the VR software required the user to interact with objects (e.g., pushing a button, holding an item) in the VE or to navigate within the VE by using a keyboard or joystick (e.g., Xbox controller).

### Statistical Analyses

Bayesian statistics were preferred over null hypothesis significance testing (NHST). The Bayesian factor ( $BF_{10}$ ) was therefore used instead of *p*-values for statistical inference, although we do report both  $BF_{10}$  and *p*-values. *P*-values measure the difference between the data and the null hypothesis ( $H_0$ ) (e.g., the assumption of no difference or no effect), while the  $BF_{10}$  calibrates *p*-values by converting them into evidence in favor of the alternative hypothesis ( $H_1$ ) over the  $H_0$  (Cox

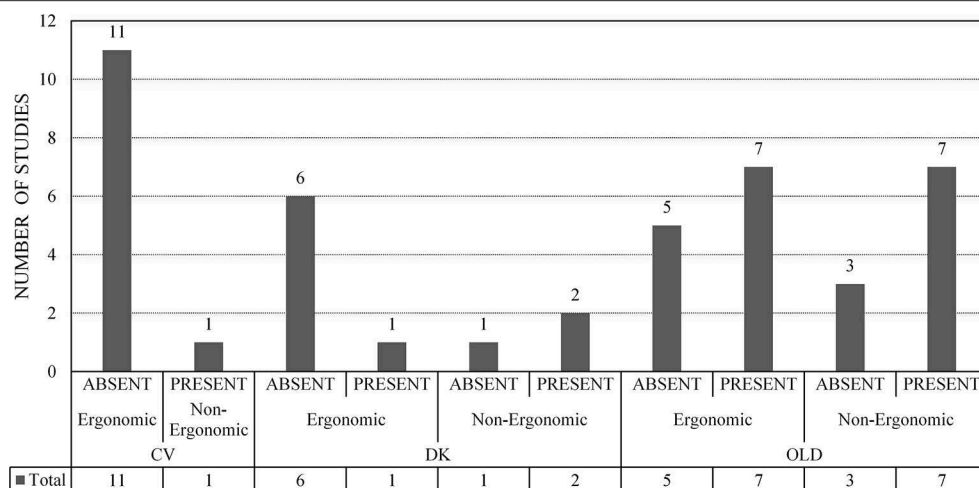
and Donnelly, 2011; Bland, 2015; Held and Ott, 2018).  $BF_{10}$  is considered substantially more parsimonious than the *p*-value in evaluating the evidence against the  $H_0$  (Cox and Donnelly, 2011; Bland, 2015; Held and Ott, 2018). Also, the difference between  $BF_{10}$  and the *p*-value in evaluating the evidence against  $H_0$  is even greater in small sample sizes (Held and Ott, 2018). Bayesian Factor ( $BF_{10}$ ) threshold  $\geq 10$  was set for statistical inference in all analyses, which indicates strong evidence in favor of the  $H_1$  (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017), and corresponds to a *p*-value  $< 0.01$  (e.g.,  $BF_{10} = 10$ ) (Cox and Donnelly, 2011; Bland, 2015; Held and Ott, 2018). JASP software was used to perform the statistical analyses (JASP Team, 2018). Bayesian independent samples *t*-tests were conducted to investigate the difference in VRISE frequency and dropout occurrence between old and new generation HMDs, as well as between new generation DKs and CVs. A Bayesian Pearson's correlations analysis examined the possible statistical relationships amongst the HMD generations, VRISE presence, the type of interactions, and dropout occurrences.

## Results

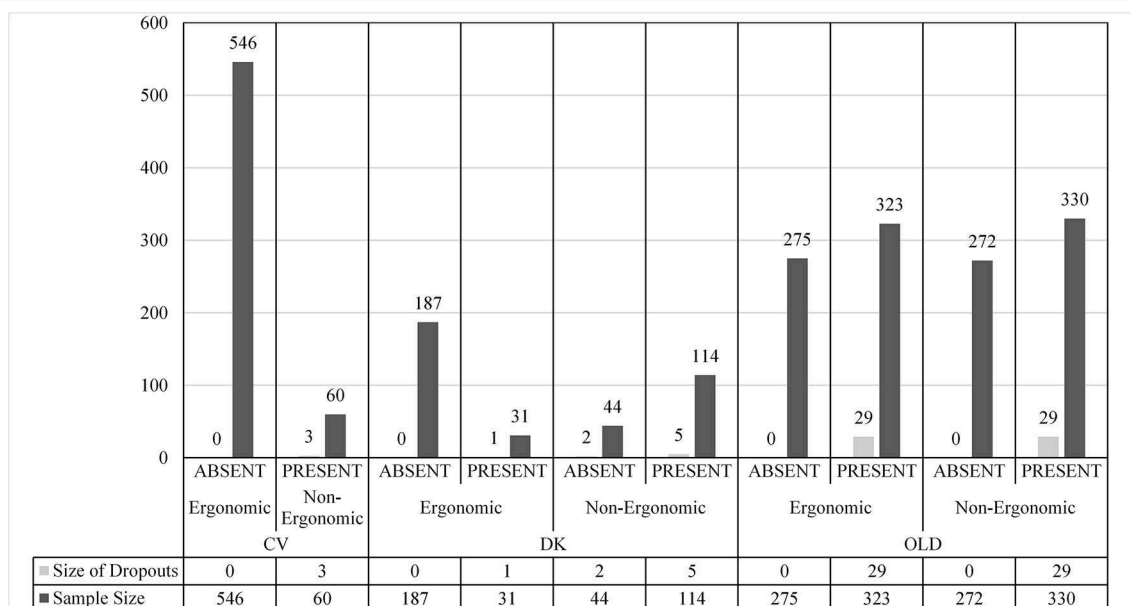
### The Implementation of Old and New Generation HMDs in Cognitive Neuroscience

The studies that utilized old generation HMDs are displayed in **Table 4** and recruited 1,200 participants in total. Nine out of 22 studies examined stress disorders, 7 of these were VR exposure therapy (VRET) studies either for phobias or post-traumatic stress disorder (PTSD), while 2 studies attempted to assess stress levels in context (e.g., assessment of social stress during a job interview). In 9 studies, there were VR assessments of cognitive functions, 2 studies assessed memory, 3 studied attention, 3 examined executive functions, and 1 examined visuospatial ability. Two of the studies involved social cognition while only one involved paranoid thinking. Lastly, only one study provided rehabilitation sessions in VR for patients with spinal injuries. The targeted age groups were young adults in 18 studies, middle-aged adults in 8 studies, older adults in one study, and children in one study.

The studies that utilized new generation HMDs are displayed in **Table 5** and recruited 982 individuals in total. Specifically, 376 individuals were recruited in 10 studies where new generation DKs were used, while 606 individuals were recruited in 12 studies where new generation CVs were used. Nine out of the 22 studies attempted to assess cognitive functions (i.e., memory, attention, visuospatial ability, executive functions), 4 investigated anxiety disorders (i.e., fear of death, social stress, general anxiety disorder), 3 provided sensorimotor rehabilitation interventions, 3 studies examined the effects of presence in specific VEs, 2 assessed social cognition and 1 study offered a psychoeducational session to patients with motor-related disorders. Lastly, the targeted age groups were young adults in 18 studies, middle-aged adults in 6 studies, and older adults in 4 studies.



**FIGURE 3 |** VRISE per HMD generation and ergonomic interactions. ABSENT, Absence of VRISE; PRESENT, Presence of VRISE; OLD, Old Generation HMD; DK, New Generation HMD—Development Kit; CV, New Generation HMD—Commercial Version; Ergonomic, Ergonomic Interactions; Non-Ergonomic, Non-Ergonomic Interactions.



**FIGURE 4 |** Dropouts and sample size per HMD, VRISE presence, and ergonomic interactions. ABSENT, Absence of VRISE; PRESENT, Presence of VRISE; OLD, Old Generation HMD; DK, New Generation HMD—Development Kit; CV, New Generation HMD—Commercial Version; Ergonomic, Ergonomic Interactions; Non-Ergonomic, Non-Ergonomic Interactions.

## Meta-Analysis

The descriptive statistics are presented in **Figures 3, 4**. In **Figure 3**, the number of studies with VRISE are displayed according to their HMD generation and interaction type. In **Figure 4**, the dropouts and sample sizes are presented according to their HMD generation, VRISE presence, and interaction type. The presence of VRISE substantially becomes less frequent when new generation HMDs are implemented (**Figure 3**). In new

generation HMDs, VRISE are present in only 4 out of 22 studies, while across 982 participants, there are only 11 dropouts. In contrast, in old generation HMDs, VRISE are present in 14 out of 22 studies, while in a total sample size of 1,200 participants, there are 58 dropouts.

In the 14 old generation HMDs studies where VRISE are present, half of them involved ergonomic interactions and the other half involved non-ergonomic interactions. Similarly, there



is an equal distribution of dropouts (29 in each) between the old generation HMDs studies that had ergonomic and non-ergonomic interactions. When only old HMDs with ergonomic interactions are considered, VRSE are present in 7 out of 12 studies, while in an entire sample size of 598, the dropouts are 29. In the studies with new generation DKs, non-ergonomic interactions had an increased presence of VRSE than the ones with ergonomic interactions. Also, in the studies which used DKs with non-ergonomic interactions, 7 participants out of 158 dropped out, while in studies with ergonomic interactions, only one participant out of 218 dropped out. Importantly, when new generation CVs with ergonomic interactions are exclusively considered, there are no VRSE or dropouts in any of the 11 studies with 546 participants. Finally, VRSE were only present in one study using a new generation CV HMD, where 3 participants dropped out. This study was the only one with a new generation CV HMD that did not involve ergonomic interactions.

The Bayesian independent samples *t*-test highlighted that studies involving new generation VR HMDs have significantly less frequent VRSE ( $BF_{10} = 144.68$ ;  $p < 0.001$ ). The difference in the existence of dropouts was not substantial, yet the studies with new generation HMDs have less frequent dropouts ( $BF_{10} = 4.69$ ;  $p < 0.05$ ) than studies with old HMDs. Notably, the studies which used a new generation CV HMD have significantly less frequent VRSE ( $BF_{10} = 46.39$ ;  $p < 0.001$ ) but not less frequent dropouts ( $BF_{10} = 1.66$ ;  $p = 0.16$ ) than the studies which used a new generation DK HMD. Finally, the studies which implemented VR software with ergonomic interactions had substantially less frequent VRSE ( $BF_{10} = 19.54$ ;  $p < 0.001$ ) and dropouts ( $BF_{10} = 16.01$ ;  $p < 0.001$ ) than studies which used VR software with non-ergonomic interactions.

The Bayesian Pearson's correlations demonstrated a substantial negative correlation between the presence of VRSE and the HMD generation [ $BF_{10} = 328.03$ ;  $r_{(44)} = -0.56$ ,  $p < 0.001$ ], while the presence of VRSE robustly demonstrated a positive correlation with the existence of dropouts [ $BF_{10} = 83510.53$ ;  $r_{(44)} = 0.68$ ,  $p < 0.001$ ]. Also, the utilization of ergonomic interactions was significantly negatively correlated with VRSE [ $BF_{10} = 20.11$ ;  $r_{(44)} = -0.42$ ,  $p < 0.001$ ] and the existence of dropouts [ $BF_{10} = 16.11$ ;  $r_{(44)} = -0.41$ ,  $p < 0.001$ ].

## Discussion

The results of the meta-analysis indicated that VR HMDs have been implemented in diverse clinical conditions and age groups, as well as the unquestionable difference between old generation and new generation HMDs. There were significantly more frequent VRSE in the studies involving old generation VR HMDs compared to studies with new generation HMDs. Additionally, the frequency of VRSE correlated negatively with the HMD generation. Hence, the older the utilized HMD, the higher the VRSE frequency. Moreover, the existence of dropouts significantly and positively correlated with the presence of VRSE.

Nevertheless, one potential reason for the higher dropouts in old generation studies is that several studies included follow-up sessions (e.g., VRET) and participants may have opted not to return to complete the remaining sessions for reasons other

than the presence of VRSE. However, in the old generation studies, the dropout rates were low in relation to the size of the population, albeit there were VRSE present. The low dropout rates in the old generation HMD studies may be due to the fixed intervals between the VR sessions where the participants were able to rest and obtain relief from any adverse effects they were experiencing.

Furthermore, the incidence of VRSE in old generation HMD studies may be due to anxiety levels (Bouchard et al., 2011) or be self-induced (Almeida et al., 2017) as several of the studies had either stress-related aims or included participants with stress disorders. However, several of the new generation HMD studies also had comparable aims and/or populations and included patients with clinical conditions (e.g., Alzheimer's disease, Parkinson's disease, stroke, and movement disorders) which have high comorbidity with stress and anxiety (Factor et al., 1995; Smith et al., 2000; Jenner, 2003; Allen and Bayraktutan, 2009). Also, the rates of self-induced VRSE are expected to be equal in both new and old generation HMD studies. In addition, the reporting of VRSE may be for reasons unrelated to the quality of the hardware or software (e.g., subjectivity in the reporting of VRSE, individual differences in the experience of VRSE) (Kortum and Peres, 2014; Almeida et al., 2017). However, this modulation is again expected to have affected both new and old generation HMD studies in a similar way.

Beyond the difference between old and new generations HMDs, a substantial difference is observed between DK and CV new generation HMDs. Significantly fewer VRSE were present in the studies that used a CV, indicating the superiority of new generation CV HMDs compared to new generation DK HMDs. Furthermore, the studies (i.e., both old and new generation studies) which utilized VR software with ergonomic interactions had robustly less frequent VRSE and dropouts than the studies which implemented VR software with non-ergonomic interactions. However, the ergonomic interactions do not appear to mitigate the dropout frequency and the incidence of VRSE in old generation HMDs. In contrast, VRSE were present in more DK studies with non-ergonomic interactions compared to DK studies with ergonomic interactions. Similarly, more participants dropped out from DK studies with non-ergonomic interactions. Notably, there were no VRSE or dropouts in CV studies with ergonomic interactions. Therefore, the contribution of ergonomic interactions in the reduction of VRSE increases when newer and better HMDs are utilized. To conclude, the findings of the meta-analysis are aligned with the outcomes of the technological review.

## GENERAL DISCUSSION

### Technological Competence in VR Neuroscience and Neuropsychology

The findings of our technological literature review suggest that the hardware features of old generation HMDs and new generation DKs do not meet the minimum hardware features that alleviate or eradicate VRSE. Instead, the technological



literature review postulates the suitability of new generation CVs which have specific hardware capabilities to alleviate VRSE. However, VR software attributes (e.g., ergonomic interactions) are equally vital.

Secondly, the findings of our meta-analysis of 44 neuroscientific or neuropsychological studies using VR are aligned with the outcomes of our technological review, where VRSE were substantially less frequent in studies which utilized new generation VR HMDs. In particular, the studies which used new generation CVs accompanied by ergonomic interactions did not have any VRSE or dropouts. Therefore, the combined outcomes of the technological review and the meta-analysis indicate that the appropriate VR HMDs are those with hardware characteristics equal to or greater than the HTC Vive and Oculus Rift, though the VR HMD should be implemented in conjunction with VR software which offers ergonomic interactions.

However, researchers may have to opt for an HMD based on their available budget. For example, the Oculus Rift costs around \$400, while the HTC Vive costs around \$500. Moreover, the majority of HMDs also require a VR-ready desktop PC or a laptop to be operated, so a researcher needs to additionally spend around \$500–\$1,500 for a desktop computer or laptop to utilize these HMDs. Hence, the combined cost would be between \$800 and \$1,900. The cost of VR equipment (e.g., both HMD and computer) may lead researchers to use HMDs that are cheaper, albeit that they are more likely to result in VRSE. However, in the market, there are plenty of cost-effective alternatives that meet the minimum hardware criteria. For example, the Oculus Quest is a standalone HMD (i.e., it does not require a PC, a laptop, or a smartphone to be operated) and it costs approximately \$400. Hence, a researcher can spend the equivalent of the price of a neuropsychological test or a smartphone to acquire and use an HMD that meets the minimum hardware criteria to lower the presence of VRSE.

Nonetheless, the selection of an appropriate VR HMD and software requires technological competency from the researchers, clinicians, and/or research software developers. Unfortunately, the meta-analysis results do not indicate that technological knowledge of VR has been well-established in neuroscience. Of course, the utilization of old generation HMDs and new generation DKs pre-2016 is justified as the new generation CVs were not available. However, in our meta-analysis, 25 studies were conducted between 2016 and 2019, where half of these studies (13/25) implemented an inappropriate HMD (i.e., old generation HMD or new generation DK). However, 10 studies used a DK2 which has a marginally lower resolution than the minimum hardware criteria, while our meta-analysis results indicated that its utilization in conjunction with ergonomic interactions appears to alleviate the frequency of VRSE, but not as effectively as the CV HMDs. Furthermore, one fifth of the studies did use a new generation HMD, but they did not have ergonomic interactions in their VR software. Therefore, at this time, VR technological competence does not seem to have been well-established in neuroscience. As a result, in the studies since 2016, the health and safety of the participants may not be substantially guaranteed, and the reliability of the results may be

questionable, as VRSE substantially decreases reaction times and overall cognitive performance (Plant and Turner, 2009; Nalivaiko et al., 2015; Plant, 2016; Nesbitt et al., 2017; Mittelstaedt et al., 2018), as well as confounding neuroimaging and physiological data (Toschi et al., 2017; Arafat et al., 2018; Gavvani et al., 2018). The selection of an appropriate HMD is paramount for successfully implementing VR HMDs in cognitive neuroscience and neuropsychology.

However, the implementation of the currently available and appropriate HMDs in neuroscience and neuropsychology should be compatible with the research aims. For example, in research designs where the user should be active (i.e., navigating, walking, and interacting within the VE) instead of being idle, or in a standing or a seated position, the researcher should opt for the best HMD that permits intense body movement and activity. In this setting, the Oculus Rift was found to be inferior to the HTC Vive on pick-and-place (i.e., relocating objects) tasks, whilst the HTC Vive also provided a substantially superior VR experience for users compared to the Oculus Rift (Suznjevic et al., 2017). Moreover, the HTC Vive provides an interactive area that is twice the size (25 m<sup>2</sup>) of the Oculus Rift, albeit that both are very accurate in tracking (Borrego et al., 2018). Nevertheless, the HTC Vive was found to lose motion-tracking and the ground level becomes slanted when the user goes out of bounds (Niehorster et al., 2017). This shortcoming solely affects studies where the participant needs to go out of the tracking area. In most neuroscientific designs, the recommended maximum play area by HTC (6.25 m<sup>2</sup>) or by Borrego et al. (2018) (25 m<sup>2</sup>) are both substantially adequate for conducting ecological valid experiments (Borges et al., 2018; Borrego et al., 2018). Nonetheless, the slanted floor or lost tracking is not a hardware problem but a software one (Borges et al., 2018). In cases where the participant is required to go out of the tracking area, the tracking problem or the slanted floor may be easily corrected by adding 3 additional trackers (Peer et al., 2018), using software with an improved algorithm (freely distributed by NASA Ames Research Center) (Borges et al., 2018), or by simply updating the firmware of the lighthouse base stations. In summary, the researchers should be technologically competent to not only identify and implement a safe HMD and software, but an HMD and software that facilitate the optimal research methods pertinent to their research needs and aims.

As discussed in our technological review, the quality of the implemented VR software is equally important to avoid VRSE. Our meta-analysis of VR studies indicated that the utilization of ergonomic interactions is crucial albeit with the utilization of an appropriate HMD. For example, Detez et al. (2019) used the HTC Vive to investigate physiological arousal and behavior during gambling. However, the interactions and navigation within the VE were facilitated by using a typical controller (Detez et al., 2019). Hence, their VR software did not support the utilization of the ergonomic 6DoF controllers (both hands) of the HTC Vive, which facilitate naturalistic navigation (e.g., teleportation) and interaction within the VE. Consequently, Detez et al.'s (2019) participants experienced VRSE and 3 of their participants discontinued their sessions and so their data were discarded (Detez et al., 2019). Importantly, Detez et al. (2019) only reported

the presence of VRISE and dropout size. They did not provide any quantitative data on the intensity of VRISE, or the quality of their software attributes (e.g., graphics, sound, tutorials, in-game instructions and prompts) (Detez et al., 2019). Indeed, only six of the studies in the meta-analysis provided adequately explicit reports on VRISE and VR software. Since Detez et al. (2019) assessed reaction times and heart rates, these data are likely to be affected by VRISE, despite the study having a rigorous experimental design and using the HTC Vive. Therefore, it is important to use appropriate VR software and external hardware to prevent risks to the health and safety of the participants as well as the reliability of the results.

## Limitations and Future Studies

The above technological review and meta-analysis of VR studies evidenced the importance of technological and methodological features in VR research and clinical designs. However, our meta-analysis of VR studies has some limitations. The meta-analysis considered VR studies with diverse populations and designs, which may have affected the frequency of VRISE and the existence of dropouts. Uniformity across studies (e.g., considering only VRET, assessments or a specific clinical population) was not possible due to the scarcity of neuroscience studies involving VR, especially using new generation HMDs. Moreover, the review did not consider any software details due to the scarceness of such descriptions in published studies. Future VR studies should report software and hardware features to allow an in-depth meta-analysis. Equally, only six studies provided quantitative reports of VRISE intensity, consequently, only the presence or absence of VRISE was considered. The dichotomous consideration of VRISE is susceptible to reports based on subjective criteria and individual differences, but this is likely to have affected the VRISE rates in both old and new generation studies. Future studies should aim to appraise the quality of the software and intensity of VRISE (e.g., using questionnaires). Studies should also attempt to clarify the acceptable duration of immersive VR sessions, which will aid researchers in designing their studies appropriately. Importantly, the cost of the VR software development should also be considered. Finally, studies should attempt to provide software development guidelines that enable researchers and/or research software developers

to develop VR research software without depending on third parties (e.g., freelance developers or software development companies) and these guidelines should embed suggestions and instructions for VR software development, which meet the criteria discussed above.

## Conclusion

The use of VR HMDs is becoming more popular in neuroscience either for clinical or research purposes and VR technology and methods have been well-accepted by diverse populations in terms of age groups and clinical conditions. A more pleasant VR experience and a reduction in VRISE symptomatology has been found using new generation CV HMDs, which deliver an adequately high display resolution, rapid image refresh rate, ergonomic design and has controllers which allow naturalistic navigation and movement within the VE environment, especially when there is restricted teleportation. The outcomes of the current technological review and meta-analysis support the feasibility of new generation VR CV HMDs to be implemented in cognitive neuroscience and neuropsychology. The findings of the technological review suggest methods that should be considered in the development or selection of VR research software, as well as hardware and software features that should be included in the research protocol. The selected VR HMD and the VR research software should enable suitable ergonomic interactions, locomotion techniques (e.g., teleportation), and kinetic mechanics which ensure VRISE are reduced or completely avoided. A meticulous approach and technological competence are compulsory to consolidate the viability of VR research and clinical designs in cognitive neuroscience and neuropsychology.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

## AUTHOR CONTRIBUTIONS

PK had the initial idea and contributed to every aspect of this study. SC, LD, and SM contributed to the methodological aspects and the discussion of the results.

## REFERENCES

- Allen, C. L., and Bayraktutan, U. (2009). Oxidative stress and its role in the pathogenesis of ischaemic stroke. *Int. J. Stroke* 4, 461–470. doi: 10.1111/j.1747-4949.2009.00387.x
- Almeida, A., Rebelo, F., Noriega, P., and Vilar, E. (2017). “Virtual reality self induced cybersickness: an exploratory study,” in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 26–33. doi: 10.1007/978-3-319-60582-1\_3
- Anthes, C., García-Hernández, R. J., Wiedemann, M., and Kranzlmüller, D. (2016). “State of the art of virtual reality technology,” in *2016 IEEE Aerospace Conference* (Big Sky, MT), 1–19. doi: 10.1109/AERO.2016.7500674
- Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). “Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 1–120. doi: 10.1109/VR.2018.8446194
- Banakou, D., Kishore, S., and Slater, M. (2018). Virtually being Einstein results in an improvement in cognitive Task performance and a decrease in age bias. *Front. Psychol.* 9:917. doi: 10.3389/fpsyg.2018.00917
- Banville, F., Nolin, P., Lalonde, S., Henry, M., Dery, M. P., and Villemure, R. (2010). Multitasking and prospective memory: can virtual reality be useful for diagnosis? *Behav. Neurol.* 23, 209–211. doi: 10.1155/2010/320707
- Barberia, I., Oliva, R., Bourdin, P., and Slater, M. (2018). Virtual mortality and near-death experience after a prolonged exposure in a shared virtual reality may lead to positive life-attitude changes. *PLoS ONE* 13:e0203358. doi: 10.1371/journal.pone.0203358
- Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., and Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *Eur. J. Paediatr. Neurol.* 16, 514–521. doi: 10.1016/j.ejpn.2012.01.006
- Bland, M. (2015). *An Introduction to Medical Statistics, 4th Edn.* Oxford: Oxford University Press.

- Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12:752. doi: 10.1038/nrn3122
- Borges, M., Symington, A., Coltin, B., Smith, T., and Ventura, R. (2018). "HTC Vive: analysis and accuracy improvement," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 2610–2615. doi: 10.1109/IROS.2018.8593707
- Borrego, A., Latorre, J., Alcaniz, M., and Llorens, R. (2018). Comparison of oculus rift and HTC vive: feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation. *Games Health J.* 7, 151–156. doi: 10.1089/g4h.2017.0114
- Botella, C., García-Palacios, A., Villa, H., Baños, R. M., Quero, S., Alcañiz, M., et al. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: a controlled study. *Clin. Psychol. Psychother.* 14, 164–175. doi: 10.1002/cpp.524
- Bouchard, S., Robillard, G., Renaud, P., and Bernier, F. (2011). Exploring new dimensions in the assessment of virtual reality induced side effects. *J. Comput. Inform. Technol.* 1, 20–32. Available online at: <http://pubs.drdc.gc.ca/BASIS/pcandid/www/engpub/DDW?W%3DSYSNUM=536530&r=0>
- Bourdin, P., Barberia, I., Oliva, R., and Slater, M. (2017). A virtual out-of-body experience reduces fear of death. *PLoS ONE* 12:e0169343. doi: 10.1371/journal.pone.0169343
- Bozgeyikli, E., Raji, A., Katkooi, S., and Dubey, R. (2016). "Point and teleport locomotion technique for virtual reality," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, TX: ACM), 205–216. doi: 10.1145/2967934.2968105
- Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *AW&I Conference* (Chemnitz), 3.
- Brennesholtz, M. S. (2018). "3-1: Invited Paper: VR standards and guidelines," in *SID Symposium Digest of Technical Papers*, 49 (Pleasantville, NY), 1–4. doi: 10.1002/sdtp.12476
- Caputo, F., Greco, A., Egidio, D. A., Notaro, I., and Spada, S. (2017). "A preventive ergonomic approach based on virtual and immersive reality," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 3–15. doi: 10.1007/978-3-319-60582-1\_1
- Carlozzi, N. E., Gade, V., Rizzo, A. S., and Tulskey, D. S. (2013). Using virtual reality driving simulators in persons with spinal cord injury: three screen display versus head mounted display. *Disabil. Rehabil.* 8, 176–180. doi: 10.3109/17483107.2012.699990
- Christou, C. G., Michael-Grigoriou, D., Sokratous, D., and Tsiakoulia, M. (2018). "A virtual reality loop and wire game for stroke rehabilitation," in *28th International Conference on Artificial Reality and 23rd Telexistence and Eurographics Symposium on Virtual Environments* (Limassol).
- Collins, M. K., Ding, V. Y., Ball, R. L., Dolce, D. L., Henderson, J. M., and Halpern, C. H. (2018). Novel application of virtual reality in patient engagement for deep brain stimulation: a pilot study. *Brain Stimul.* 11, 935–937. doi: 10.1016/j.brs.2018.03.012
- Costello, P. J. (1997). *Health and Safety Issues Associated With Virtual Reality: A Review of Current Literature*. Loughborough: Advisory Group on Computer Graphics.
- Cox, D. R., and Donnelly, C. A. (2011). *Principles of Applied Statistics*. Cambridge: Cambridge University Press
- Davis, S., Nesbitt, K., and Nalivaiko, E. (2015). "Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters," in *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)*, 27 (Sydney), 30.
- de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1\_6
- Detz, L., Greenwood, L. M., Segrave, R., Wilson, E., Chandler, T., Ries, T., et al. (2019). A psychophysiological and behavioural study of slot machine near-misses using immersive virtual reality. *J. Gambl. Stud.* 35, 929–944. doi: 10.1007/s10899-018-09822-z
- Factor, S. A., Podskalny, G. D., and Molho, E. S. (1995). Psychogenic movement disorders: frequency, clinical profile, and characteristics. *J. Neurol. Neurosurg. Psychiatry* 59, 406–412. doi: 10.1136/jnnp.59.4.406
- Figueiredo, L., Rodrigues, E., Teixeira, J., and Techrieb, V. (2018). A comparative evaluation of direct hand and wand interactions on consumer devices. *Comput. Graph.* 77, 108–121. doi: 10.1016/j.cag.2018.10.006
- Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., and Schneider, W. X. (2016). Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. *Sci. Rep.* 6:37016. doi: 10.1038/srep37016
- Freeman, D., Bradley, J., Antley, A., Bourke, E., DeWeever, N., Evans, N., et al. (2016). Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. *Br. J. Psychiatry* 209, 62–67. doi: 10.1192/bjp.bp.115.176438
- Freeman, D., Evans, N., Lister, R., Antley, A., Dunn, G., and Slater, M. (2014). Height, social comparison, and paranoia: an immersive virtual reality experimental study. *Psychiatry Res.* 218, 348–352. doi: 10.1016/j.psychres.2013.12.014
- Frommel, J., Sonntag, S., and Weber, M. (2017). "Effects of controller-based locomotion on player experience in a virtual reality exploration game," in *Proceedings of the 12th International Conference on the Foundations of Digital Games* (Hyannis, MA: ACM), 30. doi: 10.1145/3102071.3102082
- Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., et al. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *J. Med. Internet Res.* 16:e167. doi: 10.2196/jmir.3235
- Gavani, A. M., Wong, R. H., Howe, P. R., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007
- George, C., Demmler, M., and Hussmann, H. (2018). "Intelligent interruptions for IVR: investigating the interplay between presence, workload and attention," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC: ACM), LBW511. doi: 10.1145/3170427.3188686
- Gómez-Jordana, L. I., Stafford, J., Peper, C. L. E., and Craig, C. M. (2018). Virtual footprints can improve walking performance in people with Parkinson's disease. *Front. Neurol.* 9:681. doi: 10.3389/fneur.2018.00681
- Goradia, I., Doshi, J., and Kurup, L. (2014). A review paper on oculus rift and project morpheus. *Int. J. Curr. Eng. Technol.* 4, 3196–3200. Available online at: <http://inpressco.com/wp-content/uploads/2014/09/Paper183196-3200.pdf>
- Gromala, D., Tong, X., Shaw, C., Amin, A., Ulas, S., and Ramsay, G. (2016). Mobius floe: an immersive virtual reality game for pain distraction. *Electron. Imaging* 2016, 1–5. doi: 10.2352/ISSN.2470-1173.2016.4.ERVR-413
- Hartanto, D., Kampmann, I. L., Morina, N., Emmelkamp, P. G., Neerincx, M. A., and Brinkman, W. P. (2014). Controlling social stress in virtual reality environments. *PLoS ONE* 9:e92804. doi: 10.1371/journal.pone.0092804
- Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS ONE* 12:e0174965. doi: 10.1371/journal.pone.0174965
- Hecht, J. (2016). Optical dreams, virtual reality. *Optics Photon. News* 27, 24–31. doi: 10.1364/OPN.27.6.000024
- Held, L., and Ott, M. (2018). On p-values and Bayes factors. *Annu. Rev. Stat. Appl.* 5, 393–419. doi: 10.1146/annurev-statistics-031017-100307
- Hsieh, T. J. T., Kuo, Y. H., and Niu, C. K. (2018). "Utilizing HMD VR to improve the spatial learning and wayfinding effects in the virtual maze," in *International Conference on Human-Computer Interaction* (Cham: Springer), 38–42. doi: 10.1007/978-3-319-92285-0\_6
- JASP Team (2018). *JASP (Version 0.9) [Computer Software]*.
- Jenner, P. (2003). Oxidative stress in Parkinson's disease. *Ann. Neurol.* 53, S26–S38. doi: 10.1002/ana.10483
- Jerald, J., LaViola, J. J. Jr., and Marks, R. (2017). "VR interactions," in *ACM SIGGRAPH 2017 Courses* (Los Angeles, CA: ACM), 19. doi: 10.1145/3084873.3084900
- Kane, R. L., and Parsons, T. D. (eds.). (2017). *The Role of Technology in Clinical Neuropsychology*. New York, NY: Oxford University Press.
- Kelly, J. W., Cherep, L. A., and Siegel, Z. D. (2017). Perceived space in the HTC vive. *ACM Trans. Appl. Percept.* 15:2. doi: 10.1145/3106155
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying



- simulator sickness. *Int. J. Aviation Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303\_3
- Kim, A., Darakjian, N., and Finley, J. M. (2017). Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *J. Neuroeng. Rehabil.* 14:16. doi: 10.1186/s12984-017-0225-2
- Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi: 10.1016/j.apergo.2017.12.016
- Kim, J. W., Choe, W. J., Hwang, K. H., and Kwag, J. O. (2017). "78-2: the optimum display for virtual reality," in *SID Symposium Digest of Technical Papers*, 48, 1146–1149. doi: 10.1002/sdtp.11845
- Kim, K., Kim, J., Ku, J., Kim, D. Y., Chang, W. H., Shin, D. I., et al. (2004). A virtual reality assessment and training system for unilateral neglect. *Cyberpsychol. Behav.* 7, 742–749. doi: 10.1089/cpb.2004.7.742
- Kortum, P., and Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the System Usability Scale. *Int. J. Human Comput. Int.* 30, 575–584. doi: 10.1080/10447318.2014.904177
- Lubetzky, A. V., Kary, E. E., Harel, D., Hujsak, B., and Perlin, K. (2018). Feasibility and reliability of a virtual reality oculus platform to measure sensory integration for postural control in young adults. *Physiother. Theory Pract.* 34, 935–950. doi: 10.1080/09593985.2018.1431344
- Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614
- Matheis, R. J., Schultheis, M. T., Tiersky, L. A., DeLuca, J., Millis, S. R., and Rizzo, A. (2007). Is learning and memory different in a virtual environment? *Clin. Neuropsychol.* 21, 146–161. doi: 10.1080/13854040601100668
- Meyerbroeker, K., Morina, N., Kerkhof, G. A., and Emmelkamp, P. M. G. (2013). Virtual reality exposure therapy does not provide any additional value in agoraphobic patients: a randomized controlled trial. *Psychother. Psychosom.* 82, 170–176. doi: 10.1159/000342715
- Mittelstaedt, J. M., Wacker, J., and Stelling, D. (2018). VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Real.* 23, 143–154. doi: 10.1007/s10055-018-0370-3
- Montenegro, J. M. F., and Argyriou, V. (2017). Cognitive evaluation for the diagnosis of Alzheimer's disease based on Turing test and virtual environments. *Physiol. Behav.* 173, 42–51. doi: 10.1016/j.physbeh.2017.01.034
- Moreau, G., Guay, M. C., Achim, A., Rizzo, A., and Lageix, P. (2006). The virtual classroom: an ecological version of the continuous performance test—A pilot study. *Annu. Rev. CyberTherapy Telemed.* 4, 59–66. doi: 10.1037/e695432011-008
- Mottelson, A., and Hornbæk, K. (2017). "Virtual reality studies outside the laboratory," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg: ACM), 9. doi: 10.1145/3139131.3139141
- Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043
- Nesbitt, K., Davis, S., Blackmore, K., and Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48, 1–8. doi: 10.1016/j.displa.2017.01.002
- Niehorster, D. C., Li, L., and Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research. *i-Perception* 8:2041669517708205. doi: 10.1177/2041669517708205
- Oagaz, H., Schoun, B., Pooji, M., and Choi, M. H. (2018). Neurocognitive assessment in virtual reality through behavioral response analysis. *IEEE J. Biomed. Health Inform.* 23, 1899–1910. doi: 10.1109/JBHI.2018.2881455
- Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the Oculus Rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Human Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660
- Parsons, T. D., and Barnett, M. (2017). Virtual Apartment-Based Stroop for assessing distractor inhibition in healthy aging. *Appl. Neuropsychol. Adult* 26, 144–154. doi: 10.1080/23279095.2017.1373281
- Parsons, T. D., and Carlew, A. R. (2016). Bimodal virtual reality stroop for assessing distractor inhibition in autism spectrum disorders. *J. Autism Dev. Disord.* 46, 1255–1267. doi: 10.1007/s10803-015-2663-7
- Parsons, T. D., Courtney, C. G., and Dawson, M. E. (2013). Virtual reality Stroop task for assessment of supervisory attentional processing. *J. Clin. Exp. Neuropsychol.* 35, 812–826. doi: 10.1080/13803395.2013.824556
- Parsons, T. D., and McMahan, T. (2017). An initial validation of the virtual of the virtual environment grocery store. *J. Neurosci. Methods* 291, 13–19. doi: 10.1016/j.jneumeth.2017.07.027
- Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932
- Parsons, T. D., Rizzo, A. A., Bamattre, J., and Brennan, J. (2007). Virtual reality cognitive performance assessment test. *Annu. Rev. CyberTherapy Telemed.* 5, 163–171. doi: 10.1037/e688982007-001
- Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016
- Peer, A., Ullich, P., and Ponto, K. (2018). "Vive tracking alignment and correction made easy," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen: IEEE), 653–654. doi: 10.1109/VR.2018.8446435
- Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: an open letter. *Behav. Res. Methods* 48, 408–411. doi: 10.3758/s13428-015-0577-0
- Plant, R. R., and Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: new hardware, new problems? *Behav. Res. Methods* 41, 598–614. doi: 10.3758/BRM.41.3.598
- Plouzeau, J., Paillot, D., Chardonnet, J. R., and Merienne, F. (2015). "Effect of proprioceptive vibrations on simulator sickness during navigation task in virtual environment," in *International Conference on Artificial Reality and Telexistence* (Kyoto).
- Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., and Valente, L. (2017). "Minimizing cyber sickness in head mounted display systems: design guidelines and applications," in *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)* (Perth, WA: IEEE), 1–6. doi: 10.1109/SeGAH.2017.7939283
- Quinlivan, B., Butler, J. S., Beiser, L., Williams, L., McGovern, E., O'Riordan, S., et al. (2016). Application of virtual reality head mounted display for investigation of movement: a novel effect of orientation of attention. *J. Neural Eng.* 13:056006. doi: 10.1088/1741-2560/13/5/056006
- Rakkolainen, I., Turk, M., and Höllerer, T. (2016). "A superwide-FOV optical design for head-mounted displays," in *Proceedings of the Joint 26th International Conference on Artificial Reality and Telexistence and the 21th Eurographics Symp. on Virtual Environments (ICAT-EGVE 2016)* (Little Rock, AR), 45–48.
- Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., Rizzo, A. A., et al. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *J. Traumat. Stress* 24, 93–96. doi: 10.1002/jts.20574
- Rizzo, A., Difede, J., Rothbaum, B. O., Reger, G., Spitalnick, J., Cukor, J., et al. (2010). Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. *Ann. N. Y. Acad. Sci.* 1208, 114–125. doi: 10.1111/j.1749-6632.2010.05755.x
- Rizzo, A. A., Schultheis, M., Kerns, K. A., and Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychol. Rehabil.* 14, 207–239. doi: 10.1080/09602010343000183
- Rooney, B., Burke, C., Bálint, K., O'Leary, T., Parsons, T., Lee, C. T., et al. (2017). "Virtual reality, presence and social cognition: the effect of eye-gaze and narrativity on character engagement," in *2017 23rd International Conference on Virtual System and Multimedia (VSMM)* (Dublin: IEEE), 1–6. doi: 10.1109/VSMM.2017.8346272
- Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., Dunlop, B., et al. (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic

- stress disorder in Iraq and Afghanistan War veterans. *Am. J. Psychiatry* 171, 640–648. doi: 10.1176/appi.ajp.2014.13121625
- Rouder, J. N., and Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behav. Res.* 47, 877–903. doi: 10.1080/00273171.2012.734737
- Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005
- Shiban, Y., Schelhorn, I., Pauli, P., and Mühlberger, A. (2015). Effect of combined multiple contexts and multiple stimuli exposure in spider phobia: a randomized clinical trial in virtual reality. *Behav. Res. Therapy* 71, 45–53. doi: 10.1016/j.brat.2015.05.014
- Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence* 6, 603–616. doi: 10.1162/pres.1997.6.6.603
- Smith, M. A., Rottkamp, C. A., Nunomura, A., Raina, A. K., and Perry, G. (2000). Oxidative stress in Alzheimer's disease. *Biochim. Biophys. Acta* 1502, 139–144. doi: 10.1016/S0925-4439(00)00040-5
- Sportillo, D., Paljic, A., Boukhris, M., Fuchs, P., Ojeda, L., and Roussarie, V. (2017). "An immersive Virtual Reality system for semi-autonomous driving simulation: a comparison between realistic and 6-DoF controller-based interaction," in *Proceedings of the 9th International Conference on Computer and Automation Engineering* (Sydney: ACM), 6–10. doi: 10.1145/3057039.3057079
- Stanney, K. M., and Hale, K. S. (2014). *Handbook of Virtual Environments: Design, Implementation, and Applications*. Boca Raton, FL: CRC Press. doi: 10.1201/b17360
- Suznjevic, M., Mandurov, M., and Matijasevic, M. (2017). "Performance and QoE assessment of HTC vive and oculus rift for pick-and-place tasks in VR," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)* (Erfurt: IEEE), 1–3. doi: 10.1109/QoMEX.2017.7965679
- Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003
- Veling, W., Brinkman, W. P., Dorrestijn, E., and Van Der Gaag, M. (2014). Virtual reality experiments linking social environment and psychosis: a pilot study. *Cyberpsychol. Behav. Social Netw.* 17, 191–195. doi: 10.1089/cyber.2012.0497
- Viirre, E., Price, B. J., and Chase, B. (2014). "Direct effects of virtual environments on users," in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton, FL: CRC Press), 527–535. doi: 10.1201/b17360-28
- Vorländer, M., and Shinn-Cunningham, B. (2014). "Virtual auditory displays," in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton, FL: CRC Press), 107–134. doi: 10.1201/b17360-7
- Wetzels, R., and Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychon. Bull. Rev.* 19, 1057–1064. doi: 10.3758/s13423-012-0295-x
- Yeh, S. C., Li, Y. Y., Zhou, C., Chiu, P. H., and Chen, J. W. (2018). Effects of virtual reality and augmented reality on induced anxiety. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 1345–1352. doi: 10.1109/TNSRE.2018.2844083
- Zimmer, P., Buttlar, B., Halbeisen, G., Walther, E., and Domes, G. (2018). Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses. *Psychoneuroendocrinology*. 101, 186–192. doi: 10.1016/j.psyneuen.2018.11.010

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

Copyright © 2019 Kourtesis, Collina, Doumas and MacPherson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# The Dynamics of Attention Shifts Among Concurrent Speech in a Naturalistic Multi-speaker Virtual Environment

**Keren Shavit-Cohen and Elana Zion Golumbic\***

*The Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat Gan, Israel*

## OPEN ACCESS

### Edited by:

Pietro Cipresso,  
Italian Auxological Institute (IRCCS),  
Italy

### Reviewed by:

Kristina C. Backer,  
University of California, Merced,  
United States  
Carlo Sestieri,  
Università degli Studi G. d'Annunzio  
Chieti e Pescara, Italy

### \*Correspondence:

Elana Zion Golumbic  
elana.zion-golumbic@biu.ac.il

### Specialty section:

This article was submitted to Speech  
and Language, a section of the  
journal *Frontiers in Human  
Neuroscience*

**Received:** 02 May 2019

**Accepted:** 16 October 2019

**Published:** 08 November 2019

### Citation:

Shavit-Cohen K and Zion Golumbic E  
(2019) The Dynamics of Attention  
Shifts Among Concurrent Speech in  
a Naturalistic Multi-speaker Virtual  
Environment.  
*Front. Hum. Neurosci.* 13:386.  
doi: 10.3389/fnhum.2019.00386

Focusing attention on one speaker on the background of other irrelevant speech can be a challenging feat. A longstanding question in attention research is whether and how frequently individuals shift their attention towards task-irrelevant speech, arguably leading to occasional detection of words in a so-called unattended message. However, this has been difficult to gauge empirically, particularly when participants attend to continuous natural speech, due to the lack of appropriate metrics for detecting shifts in internal attention. Here we introduce a new experimental platform for studying the dynamic deployment of attention among concurrent speakers, utilizing a unique combination of Virtual Reality (VR) and Eye-Tracking technology. We created a Virtual Café in which participants sit across from and attend to the narrative of a target speaker. We manipulated the number and location of distractor speakers by placing additional characters throughout the Virtual Café. By monitoring participant's eye-gaze dynamics, we studied the patterns of overt attention-shifts among concurrent speakers as well as the consequences of these shifts on speech comprehension. Our results reveal important individual differences in the gaze-pattern displayed during selective attention to speech. While some participants stayed fixated on a target speaker throughout the entire experiment, approximately 30% of participants frequently shifted their gaze toward distractor speakers or other locations in the environment, regardless of the severity of audiovisual distraction. Critically, preforming frequent gaze-shifts negatively impacted the comprehension of target speech, and participants made more mistakes when looking away from the target speaker. We also found that gaze-shifts occurred primarily during gaps in the acoustic input, suggesting that momentary reductions in acoustic masking prompt attention-shifts between competing speakers, in line with "glimpsing" theories of processing speech in noise. These results open a new window into understanding the dynamics of attention as they wax and wane over time, and the different listening patterns employed for dealing with the influx of sensory input in multisensory environments. Moreover, the novel approach developed here for tracking the locus of momentary attention in a naturalistic virtual-reality environment holds high promise for extending the study of human behavior and cognition and bridging the gap between the laboratory and real-life.

**Keywords:** speech processing, auditory attention, eye-tracking, virtual reality, cocktail party effect, distractability

## INTRODUCTION

Focusing attention on one speaker in a noisy environment can be challenging, particularly in the background of other irrelevant speech (McDermott, 2009). Despite the difficulty of this task, comprehension of an attended speaker is generally good and the content of distractor speech is rarely recalled explicitly (Cherry, 1953; Lachter et al., 2004). Preferential encoding of attended speech in multi-speaker contexts is also mirrored by enhanced neural responses to attended vs. distractor speech (Ding and Simon, 2012b; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013b; O'Sullivan et al., 2015). However, there are also indications that distractor speech is processed, at least to some degree. Examples for this are the Irrelevant Stimulus Effect, where distractor words exert priming effect on an attended task (Treisman, 1964; Neely and LeCompte, 1999; Beaman et al., 2007), as well as occasional explicit detection of salient words in distractor streams (Cherry, 1953; Wood and Cowan, 1995; Röer et al., 2017; Parmentier et al., 2018). These effects highlight a key theoretical tension regarding how processing resources are allocated among competing speech inputs. Whereas Late-Selection models of attention posit that attended and distractor speech can be fully processed, allowing for explicit detection of words in so-called unattended speech (Deutsch and Deutsch, 1963; Duncan, 1980; Parmentier et al., 2018), Limited-Resources models hold that there are inherent bottlenecks for linguistic processing of concurrent speech due to limited resources (Broadbent, 1958; Lachter et al., 2004; Lavie et al., 2004; Raveh and Lavie, 2015). The latter perspective reconciles indications for occasional processing of distractor speech as stemming from rapid shifts of attention toward distractor speech (Conway et al., 2001; Escera et al., 2003; Lachter et al., 2004). Yet, despite the parsimonious appeal of this explanation, to date, there is little empirical evidence supporting and characterizing the psychological reality of attention switches among concurrent speakers.

Establishing whether and when rapid shifts of attention towards distractor stimuli occur is operationally challenging since it refers to individuals' internal state that researchers do not have direct access to. Existing metrics for detecting shifts of attention among concurrent speech primarily rely on indirect measures such as prolongation of reaction times on an attended task (Beaman et al., 2007) or subjective reports (Wood and Cowan, 1995). Given these limitations, the current understanding of the dynamics of attention over time, and the nature and consequences of rapid attention-shifts among concurrent speech is extremely poor. Nonetheless, gaining insight into the dynamics of internal attention-shifts is critical for understanding how attention operates in naturalistic multi-speaker settings.

Here, we introduce a new experimental platform for studying the dynamic deployment of attention among concurrent speakers. We utilize Virtual Reality (VR) technology to simulate a naturalistic audio-visual multi-speaker environment, and track participant's gaze-position within the Virtual Scene as a marker for the locus of overt attention and as a means for detecting attention-shifts among concurrent speakers.

Participants experienced sitting in a "Virtual Café" across from a partner (avatar; animated target speaker) and were required to focus attention exclusively towards this speaker. Additional distracting speakers were placed at surrounding tables, with their number and location manipulated across conditions. Continuous tracking of gaze-location allowed us to characterize whether participants stayed focused on the target speaker as instructed or whether and how often they performed overt glimpses around the environment and toward distractor speakers. Critically, we tested whether shifting one's gaze around the environment and away from the target speaker impacted comprehension of target speech. We further tested whether gaze-shifts are associated with salient acoustic changes in the environment, such as onsets in distractor speech that can potentially grab attention exogenously (Wood and Cowan, 1995) or brief pauses that create momentary unmasking of competing sounds (Lavie et al., 2004; Cooke, 2006).

Gaze-shifts are often used as a proxy for attention shifts in natural vision (Anderson et al., 2015; Schomaker et al., 2017; Walker et al., 2017), however this measure has not been utilized extensively in dynamic contexts (Marius't Hart et al., 2009; Foulsham et al., 2011). This novel approach enabled us to characterize the nature of momentary attention-shifts in ecological multi-speaker listening conditions, as well as individual differences, gaining insight into the factors contributing to dynamic attention shifting and its consequences on speech comprehension.

## MATERIALS AND METHODS

### Participants

Twenty-six adults participated in this study (ages 18–32, median 24; 18 female, three left handed), all fluent in Hebrew, with self-reported normal hearing and no history of psychiatric or neurological disorders. Signed informed consent was obtained from each participant prior to the experiment, in accordance with the guidelines of the Institutional Ethics Committee at Bar-Ilan University. Participants were paid for participation or received class credit.

### Apparatus

Participants were seated comfortably in an acoustic-shielded room and viewed a 3D VR scene of a café, through a head-mounted device (Oculus Rift Development Kit 2). The device was custom-fitted with an embedded eye-tracker (SMI, Teltow, Germany; 60 Hz monocular sampling rate) for continuous monitoring of participants' eye-gaze position. Audio was presented through high-quality headphone (Sennheiser HD 280 pro).

### Stimuli

Avatar characters were selected from the Mixamo platform (Adobe Systems, San Jose, CA, USA). Soundtracks for the avatars' speech were 35–50 s long segments of natural Hebrew speech taken from podcasts and short stories<sup>1</sup>. Avatars' mouth and articulation movements were synced to the audio to

<sup>1</sup>www.icast.co.il

create a realistic audio-visual experience of speech (LipSync Pro, Rogo Digital, England). Scene animation and experiment programming was controlled using an open-source VR engine (Unity Software<sup>2</sup>). Speech loudness levels (RMS) were equated for all stimuli, in 10-s long bins (to avoid biases due to fluctuations in speech time-course). Audio was further manipulated within Unity using a 3D sound algorithm, so that it was perceived as originating from the spatial location of the speaking avatar, with overall loudness decreasing logarithmically with distance from the listener. Participant's head movements were not restricted, and both the graphic display and 3D sound were adapted on-line in accordance with head-position, maintaining a spatially-coherent audio-visual experience.

## Experiment Design

In the Virtual Café setting, participants experienced sitting at a café table facing a partner (animated speaking avatar) telling a personal narrative. They were told to focus attention exclusively on the speech of their partner (target speaker) and to subsequently answer four multiple-choice comprehension questions about the narrative (e.g., “What computer operating system was mentioned?”). Answers to the comprehension questions were evenly distributed throughout the narrative, and were pre-screened in a pilot study to ensure accuracy rates between 80% and 95% in a single-speaker condition. The time-period containing the answer to each question was recorded and used in subsequent analysis of performance as a function of gaze-shift behaviors (see below). Additional pairs of distracting speakers (avatars) were placed at surrounding tables, and we systematically manipulated the number and location of distractors in four conditions: No Distraction (NoD), Left Distractors (LD), Right Distractors (RD), Right and Left Distractors (RLD; **Figure 1**). Each condition consisted of five trials (~4 min per condition) and was presented in random order, which was different for each participant. The identity and voice of the main speaker were kept constant throughout the experiment, with different narratives in each trial, while the avatars and narratives serving as distractors varied from trial to trial. The allocation of each narrative to the condition was counter-balanced across participants, to avoid material-specific biases. Before starting the experiment itself, participants were given time to look around and familiarize themselves with the Café environment and the characters in it. During this familiarization stage, no audio was presented and participants terminated it when they were ready. They also completed two training-trials, in the NoD and RLD conditions, to familiarize them with the stimuli and task as well as the type of comprehension questions asked. This familiarization and training period lasted approximately 3-min.

## Analysis of Eye-Gaze Dynamics

Analysis of eye-gaze data was performed in Matlab (Mathworks, Natick, MA, USA) using functions from the fieldtrip toolbox<sup>3</sup> as well as custom-written scripts. The position of eye-gaze position in virtual space coordinates (x, y, z) was monitored continuously

throughout the experiment. Periods surrounding eye-blinks were removed from the data (250 ms around each blink). Clean data from each trial were analyzed as follows.

First, we mapped gaze-positions onto specific avatars/locations in the 3D virtual scene. For data reduction, we used a spatial clustering algorithm (k-means) to combine gaze data-points associated with similar locations in space. Next, each spatial cluster was associated with the closest avatar, by calculating the Euclidean distance between the center of the cluster and the center of each avatar presented in that condition. If two or more clusters were associated with looking at the same avatar, they were combined. Similarly, clusters associated with the members of the distractor avatar-pairs (left or right distractors) were combined. If a cluster did not fall within a particular distance-threshold from any of the avatars, it was associated with looking at “The Environment.” This resulted in a maximum of four clusters capturing the different possible gaze locations in each trial: (1) Target Speaker; (2) Left Distractors (when relevant); (3) Right Distractors (when relevant); and (4) Rest of the Environment. The appropriateness of cluster-to-avatar association and distance-threshold selection was verified through visual inspection.

Based on the clustered data, we quantified the percent of time that participants spent focusing at each location (*Percent Gaze Time*) in each trial, and detected the times of *Gaze-Shifts* from one cluster to another. Gaze-shifts lasting less than 250 ms were considered artifacts and removed from the analysis, as they are physiologically implausible (Bompas and Sumner, 2009; Gilchrist, 2011). The number of Gaze-shifts as well as the Percent Gaze Time spent at each of the four locations—Target Speaker, Left Distractors, Right Distractors and Environment—were averaged across trials, within condition. Since conditions differed in the type and number of distractors, comparison across conditions focused mainly on metrics pertaining to gazing at/away-from the target speaker.

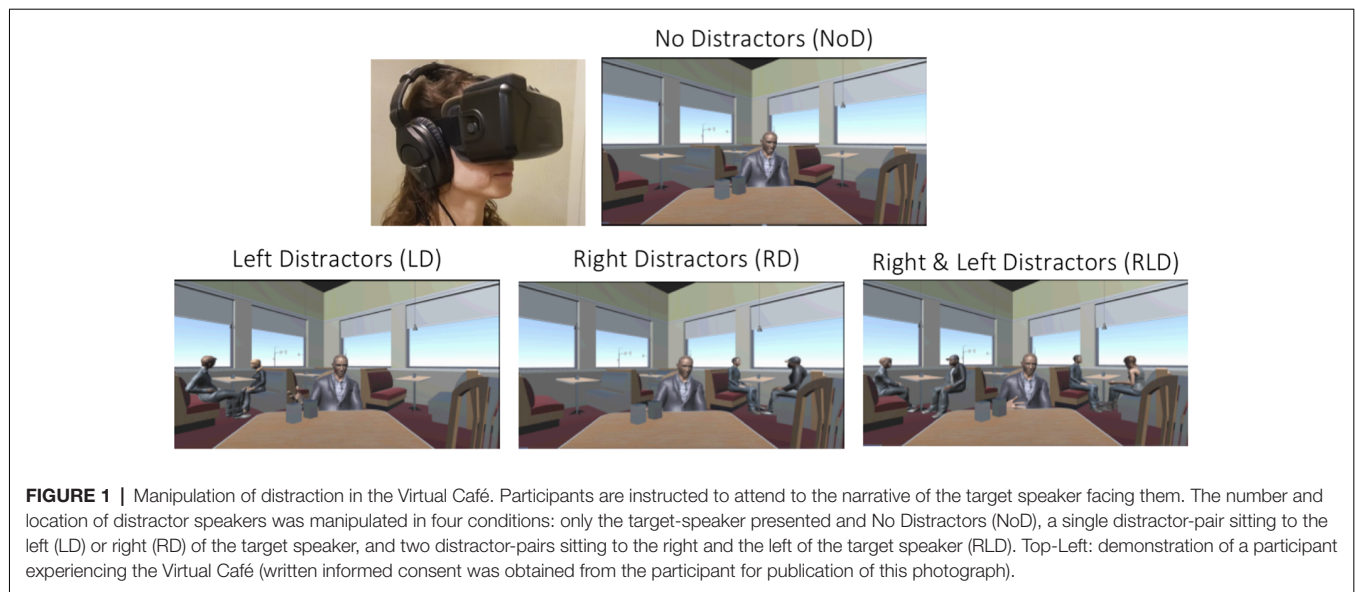
Mixed linear regression models were used in all analyses to fit the data and test for effects of Condition on gaze patterns (both Percent Gaze-Time Away and Gaze-Shifts), as well as possible correlations with speech comprehension accuracy measures. These analyses were conducted in R (R Development Core Team, 2012) and we report statistical results derived using both regular linear (lme4 package for R; Bates et al., 2015) and robust estimation approaches (robustlmm package for R; Koller, 2016), to control for possible contamination by outliers. The advantage of mixed-effects models is that they account for variability between subjects and correlations within the data, as well as possible differences in trial numbers across conditions (Baayen et al., 2008), which makes them particularly suitable for the type of data collected here.

## Analysis of Speech Acoustics Relative to Gaze-Shifts

A key question is what prompts overt gaze-shifts away from the target speakers, and specifically whether they are driven by changes in the acoustic input or if they should be considered more internally-driven. Two acoustic factors that have been suggested as inviting attention-shifts among concurrent speech

<sup>2</sup>unity3d.com

<sup>3</sup>fieldtriptoolbox.org



are: (a) onsets/loudness increases in distractor speech that can potentially grab attention exogenously (Wood and Cowan, 1995); and (b) brief pauses that create momentary unmasking of competing sounds (Lavie et al., 2004; Cooke, 2006). To test whether one or both of these factors account for the occurrence of gaze-shifts away from the target speaker in the current data, we performed a gaze-shift time-locked analysis of the speech-acoustics of target speech (in all conditions) and distractor speech (in the LD, RD and RLD conditions).

To this end, we first calculated the temporal envelope of the speech presented in each trial using a windowed RMS (30 ms smoothing). The envelopes were segmented relative to the times where gaze-shifts *away from the target speaker* occurred in that particular trial (−400 to +200 ms around each shift). Given that the initiation-time for executing saccades is ~200 ms (Gilchrist, 2011), the time-window of interest for looking at possible influences of the acoustics on gaze-shifts is prior to that, i.e., 400–200 ms prior to the gaze-shift itself.

Since the number of gaze-shifts varied substantially across participants, we averaged the gaze-shift-locked envelope-segments across all trials and participants, within condition. The resulting average acoustic-loudness waveform in each condition was compared to a distribution of non-gaze-locked loudness levels, generated through a permutation procedure as follows: the same acoustic envelopes were segmented randomly into an equal number of segments as the number of gaze-shifts in each condition (sampled across participants with the same proportion as the real data). These were averaged, producing a non-gaze-locked average waveform. This procedure was repeated 1,000 times and the real gaze-shift locked waveform was compared to the distribution of non-gaze-locked waveforms. We identified time-points where the loudness level fell above or below the top/bottom 5% tile of the non-gaze-locked distribution, signifying that the speech acoustics were particularly quiet or loud relative (relative to the rest of the presented speech stimuli). We also quantified the signal-to-noise

ratio (SNR) between the time-resolved spectrograms of target and distractor speech surrounding gaze-shifts, according to:  $SNR(f, t) = \log \left( \frac{P_{\text{target}}(f, t)}{P_{\text{distractor}}(f, t)} \right)$ , with  $P(f, t)$  depicting the power at frequency  $f$  at time  $t$ . This was calculated for target-distractor combinations surrounding each gaze-shift, and averaged across shifts and trials.

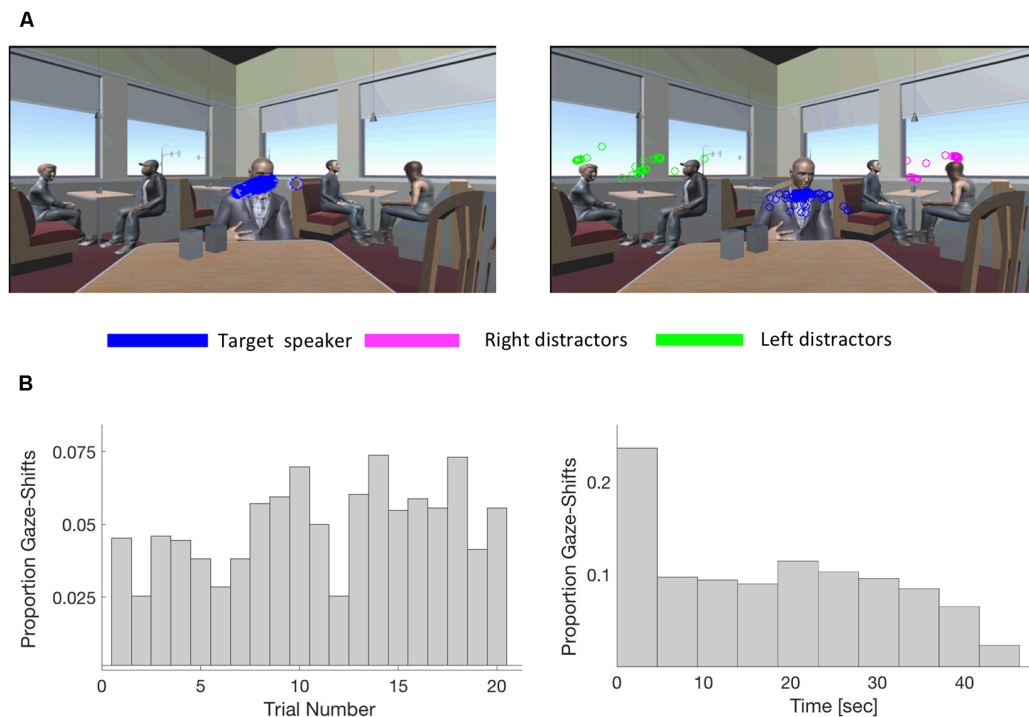
## RESULTS

### Gaze-Patterns and Speech Comprehension

On an average, participants spent −7.6% of each trial (−3 s in a 40-s-long trial) looking at locations other than the target speaker and they performed an average of 2.5 gaze-shifts per trial. **Figure 2A** shows the distribution of eye-gaze location in two example trials taken from different participants, demonstrating that sometimes gaze was fixated on the target-speaker throughout the entire trial, and sometimes shifted occasionally towards the distractors. The distribution of Gaze-shifts was relatively uniform over the course of the entire experiment (**Figure 2B**, left). Twenty-three percentage of gaze-shifts were performed near the onset of the trial, however, the majority of gaze-shifts occurred uniformly throughout the entire trial (**Figure 2B**, right).

**Figures 3A,B** show how the average *Gaze Time Away* from the target speaker (i.e., time spent looking at distractor avatars or other locations in the Environment) and the number of *Gaze-Shifts* away from the target speaker, varied across the four conditions. To test whether gaze patterns (number of *Gaze-Shifts* and/or proportion *Gaze-Time Away*) differed across conditions, we estimated each of them separately using linear mixed effect model with the factor Condition as a fixed effect (*Gaze-Shifts* Condition and *Gaze-Time* Condition), where each of the three distraction conditions (RD, LD and RLD) was compared to the NoD condition. By-subject intercepts were





**FIGURE 2 |** Characterization of gaze-shift patterns. **(A)** Illustration of the variability in gaze-patterns across individuals. The figure depicts all gaze data points in a specific trial in the RLD condition for two example participants. While the participant shown in the left panel remained focused exclusively on the target speaker throughout the trial (blue dots), the participant in the right panel spent a substantial portion of the trial looking at the distractor speakers on both the left (green) and the right (magenta). **(B)** Left: distribution of all gaze-shifts across the duration of the experiment, collapsed across participants. Gaze-shifts occurred throughout the experiment and were not more prevalent the beginning/end. Right: distribution of gaze-shifts over the course of a trial, collapsed across all trials and participants. Twenty-three percentage of gaze-shifts occurred during the first 5 s of each trial, and the remainder could occur with similar probability throughout the entire trial.

included as random effects. No significant effects of Condition were found on *Gaze-Time*, however, participants performed significantly more *Gaze-Shifts* in the RLD condition relative to the NoD condition (lmer:  $\beta = 0.8$ ,  $t = 2.5$ ,  $p = 0.01$ ; robustlmm:  $\beta = 0.54$ ,  $t = 2.5$ ).

Of critical interest is whether the presence of distractors and gaze-shifts towards them impacted behavioral outcomes of speech comprehension. Accuracy on the multiple-choice comprehension questions of the target speaker was relatively good in all conditions (mean accuracy  $82\% \pm 3$ ; **Figure 3C**). A mixed linear model estimating  $\text{Accuracy} \sim \text{Condition}$  did not reveal any significant differences in Accuracy between conditions (lmer: all  $t$ 's  $< 0.199$ ,  $p > 0.6$ ; robustlmm: all  $t$ 's  $< 0.05$ ). However, adding Percent Gaze-Time as a second fixed effect to the  $\text{Accuracy} \sim \text{Condition}$  model, improved the model significantly ( $\chi^2 = 9.14$ ,  $p < 10^{-3}$ ), with Percent Gaze-Time showing a significant correlation with Accuracy (lmer:  $\beta = -0.19$ ,  $t = -3.13$ ,  $p = 0.001$ ; robustlmm:  $\beta = -0.23$ ,  $t = -3.77$ ; **Figure 3D**). Adding Number of Shifts to the  $\text{Accuracy} \sim \text{Condition}$  model, however, did not yield any additional significant advantage (likelihood ratio test  $\chi^2 = 2.4$ ,  $p > 0.1$ ; **Figure 3E**), suggesting that the number of gaze-shifts performed *per se* did not affect speech comprehension.

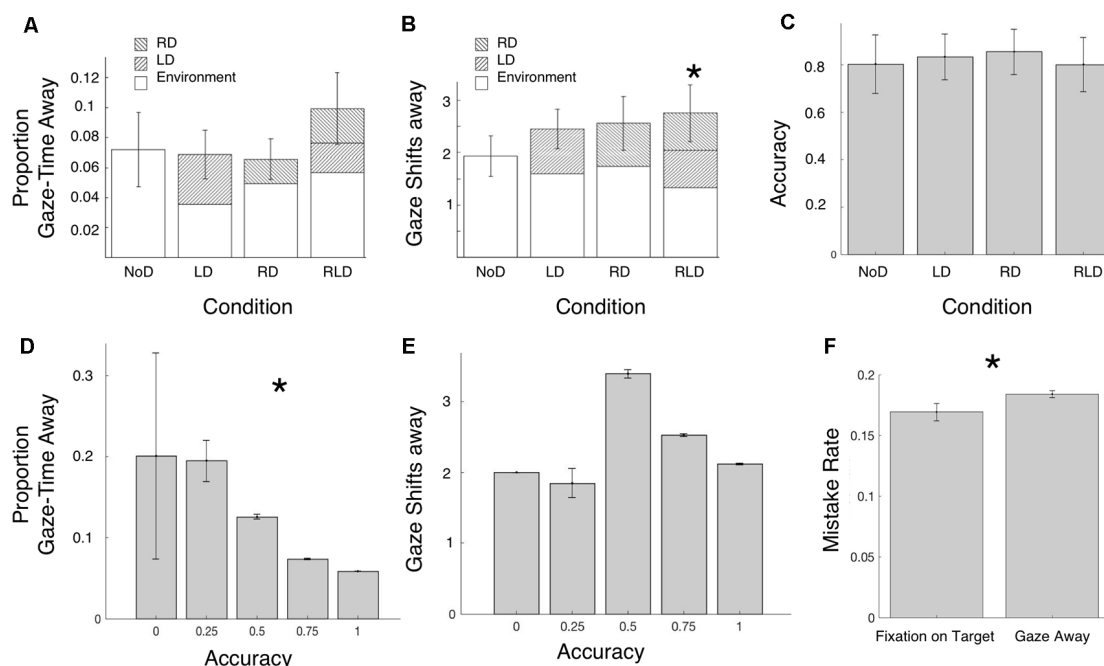
To further assess the link between performance on the comprehension questions and gaze-shifts, we tested whether

participants were more likely to make mistakes on specific questions if they happened to be looking away from the target-speaker when the critical information for answering that question was delivered. Mistake rates were slightly lower when participants fixated on the target speaker when the critical information was delivered (16% miss-rate) vs. when they looked away (18% miss-rate). To evaluate this effect statistically, we fit a linear mixed model to the accuracy results on individual questions testing whether they were mediated by the presence of a gaze-shift when the answer was given, as well as the condition [ $\text{Accuracy} \sim \text{Shift (yes/no)} + \text{Condition}$  as fixed effects], with by-subject intercepts included as random effects. This analysis demonstrated a small yet significant effect of the presence of a gaze-shift during the period when the answer was given (lmer  $\beta = -0.05$ ,  $t = -2.16$ ,  $p < 0.04$ ; robustlmm  $t = -3$ ; **Figure 3F**), however there was no significant effect of Condition (all  $t$ 's  $< 0.5$ ).

## Individual Differences in Gaze Patterns

When looking at gaze-patterns across participants, we noted substantial variability in the number of gaze-shift performed and percent time spent gazing away from the target speaker. As illustrated in **Figures 2A, 4**, some participants stayed completely focused on the main speaker throughout the entire experiment, whereas others spent a substantial portion of each trial gazing





**FIGURE 3 |** Summary of gaze-shift patterns and behavioral outcomes across conditions. **(A,B)** Proportion of Gaze-Time and Number of Gaze-Shifts Away from target speaker, per trial and across conditions. Results within each condition are broken down by gaze-location (Right Distractors, Left Distractors or Environment in blank, left and right diagonals, respectively). There was no significant difference between conditions in the total Gaze-time away from the target speaker or number of gaze-shifts. Significantly more Gaze-Shifts were performed in the RLD condition relative to the NoD condition. No other contrasts were significant. **(C)** Mean accuracy on comprehension questions, across condition. Difference between conditions was not significant. **(D,E)** Analysis of Accuracy as a function of Gaze-Shift Patterns, at the whole trial level. Trials where participants spent a larger proportion of the time looking away from the target-speaker were associated with lower accuracy rates. No significant correlation was found between accuracy rates and the number of Gaze-Shifts performed. **(F)** Analysis of Accuracy on single question as a function of Gaze-Shift Patterns. Mistake rates were significantly higher if participants were looking away from the target speaker vs. fixating on the target speaker during the time-window when the information critical for answering the question was delivered. Error bars indicate Standard Error of the Mean (SEM). \* $p < 0.05$ .

around the environment (*range across participants*: 0–18 average number gaze-shifts per trial; 0–34.52% average percent of trial spent looking away from the target speaker). This motivated further inspection of gaze-shift behavior at the individual level. Specifically, we tested whether individual behavior of performing many or few gaze-shifts away from the target was stable across conditions. We calculated Cronbach's  $\alpha$  between conditions and found high internal consistency across conditions in the number of gaze-shifts performed as well as in the percent of gaze-time away from the target speaker ( $\alpha = 0.889$  and  $\alpha = 0.832$ , respectively). This was further demonstrated by strong positive correlations between the percent time spent gazing away from the target speaker in No Distraction condition vs. each of the Distraction conditions (lmer: all  $r$ 's  $> 0.5$ ; robustlmm all  $r$ 's  $> 0.6$ ) as well as the number of gaze-shifts (lmer and robustlmm: all  $r$ 's  $> 0.5$ ; **Figures 4C,D**). This pattern suggests that individuals have characteristic tendencies to either stay focused or gaze-around the scene, above and beyond the specific sensory attributes or degree of distraction in a particular scenario.

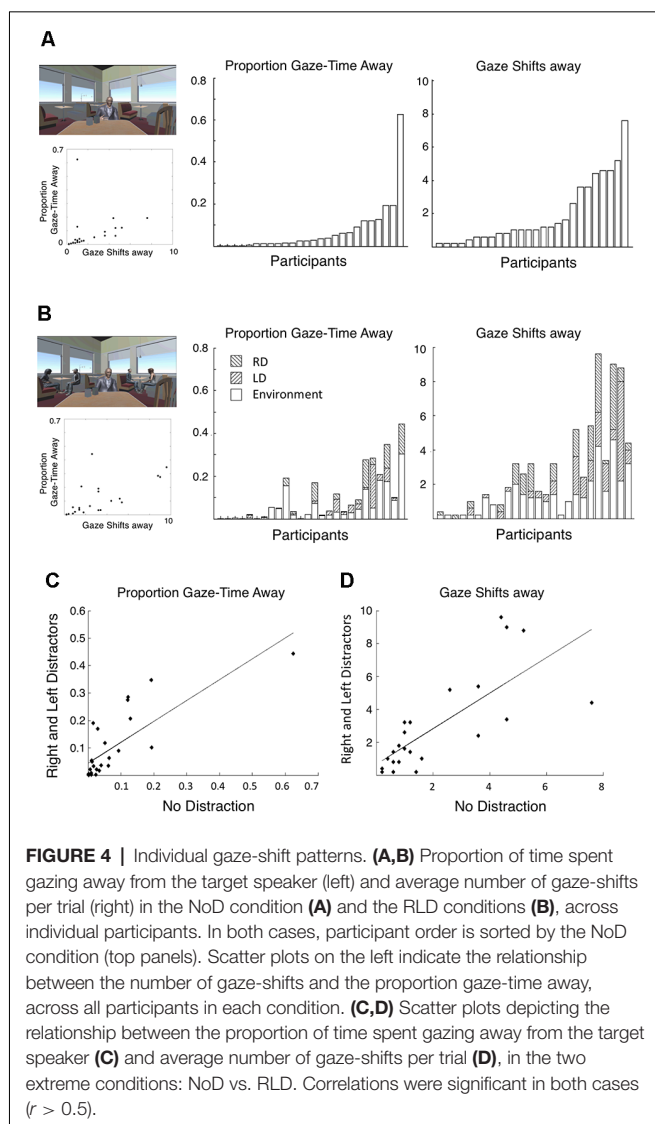
## Gaze-Locked Analysis of Speech Acoustics

Last, we tested whether there was any relationship between the timing of gaze-shifts and the local speech-acoustics. To this

end, we performed a gaze-shift-locked analysis of the envelope of the target or distractor speech (when present). Analysis of distractor speech envelope consisted only of eye-gaze shifts *toward that distractor* (i.e., excluding shifts to other places in the environment). **Figure 5** shows the average time-course of the target and distractor speech envelopes relative to the onset of a gaze-shift. For both target speech (top row) as well as for distractor speech (bottom row), gaze-shifts seem to have been preceded by a brief period of silence (within the lower 5% tile; red shading) between 200 and 300 ms prior to the shift.

Frequency-resolved analysis of the SNR between target and distractor speech similarly indicates low SNR in the period preceding gaze-shifts. A reduction in SNR prior to gaze-shifts was primarily evident in the 3–8 kHz range (sometimes considered the “unvoiced” part of the speech spectrum; Atal and Hanauer, 1971), whereas SNR in the lower part of the spectrum (0–2 kHz) was near 1 dB both before and after gaze-shifts. Although SNR does not take into account the overall loudness-level of each speaker but only the ratio between the speakers, the observed SNR modulation is consistent with momentary periods of silence/drops in the volume of *both* concurrent speakers.

This pattern is in line with an acoustic release-from-masking account, suggesting that gaze-shifts are prompted by momentary gaps in the speech, and particularly when gaps in concurrent



speech coincide-temporally (as seen here in the Single and Two Distractor conditions). Conversely, the suggestion that attention-shifts are a product of exogenous capture by salient events in distracting speech does not seem to be supported by the current data, since the acoustics of the distractor speech that participants shifted their gaze towards did not seem to contain periods with consistently loud acoustics. We did, however, find increases in loudness of the target speech acoustics near gaze-shift onset (within the top 5% tile; red shading between  $-100$  and  $+50$  ms).

## DISCUSSION

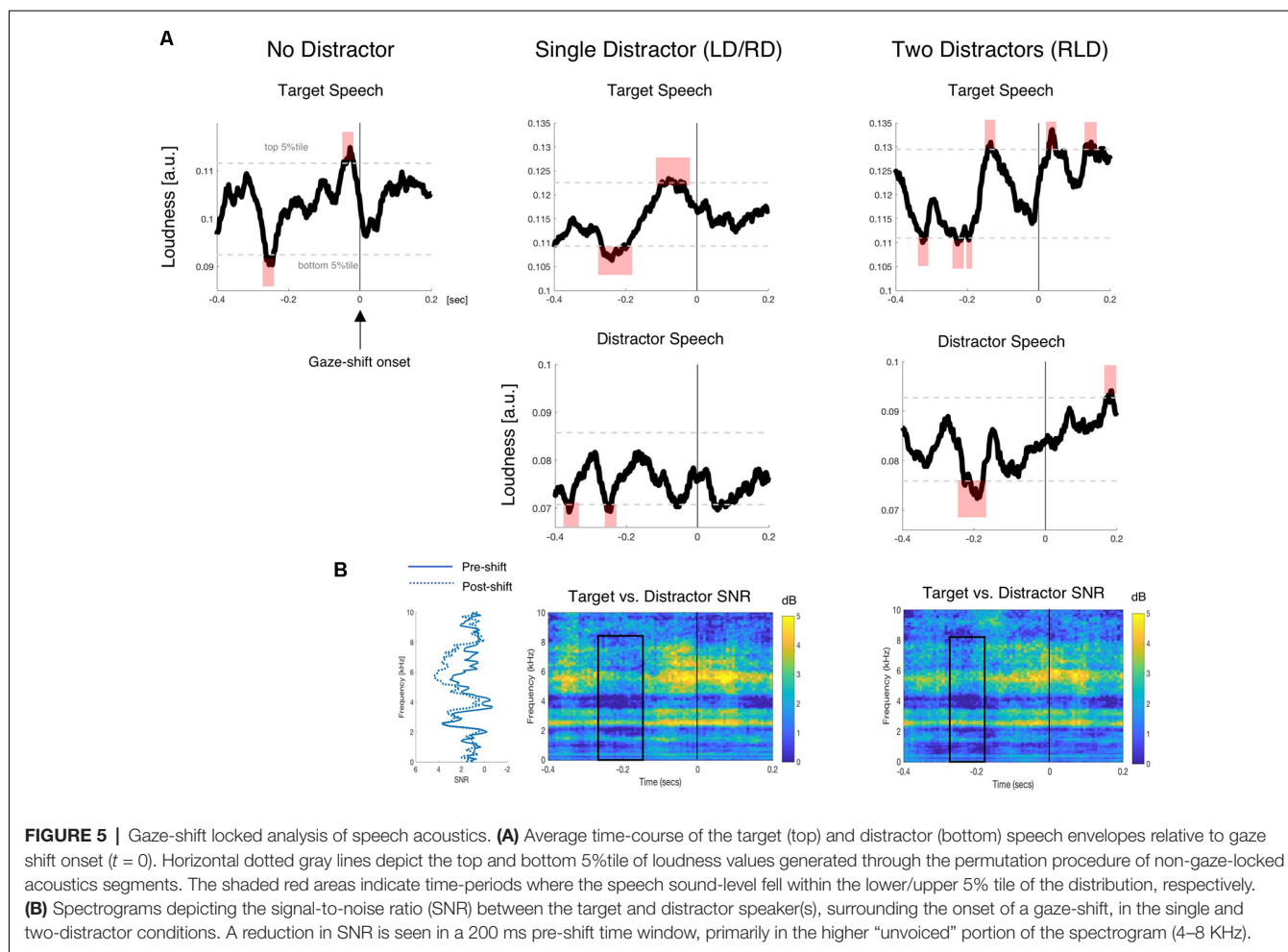
The current study is a first and novel attempt to characterize how individuals deploy overt attention in naturalistic audiovisual settings, laden with rich and competing stimuli. By monitoring eye-gaze dynamics in our Virtual Café, we studied the patterns of gaze-shifts and its consequences for speech comprehension. Interestingly, we found that the presence and number of

competing speakers in the environment did not, on average, affect the amount of time spent looking at the target speaker, nor did it impair comprehension of the target speaker, although participants did perform slightly more gaze-shifts away in the two-distractor RLD condition. This demonstrates an overall resilience of the attention and speech-processing systems for overcoming the acoustic-load posed by distractors in naturalistic audio-visual conditions. This ability is of utmost ecological value, and likely benefits both from the availability of visual and spatial cues (Freyman et al., 2004; Zion Golumbic et al., 2013a) as well as the use of semantic context to maintain comprehension despite possible reductions in speech intelligibility (Simpson and Cooke, 2005; Vergauwe et al., 2010; Ding and Simon, 2012a; Calandruccio et al., 2018). At the same time, our results also suggest that the ability to maintain attention on the designated speaker under these conditions is highly individualized. Participants displayed characteristic patterns of either staying focused on a target speaker or sampling other locations in the environment overtly, regardless of the severity of the so-called sensory distraction. Critically, the amount of time that individuals spent looking around the environment and away from the target speaker was negatively correlated with speech comprehension, directly linking overt attention to speech comprehension. We also found that gaze-shifts away from the target speaker occurred primarily following gaps in the acoustic input, suggesting that momentary reductions in acoustic masking can prompt attention-shifts between competing speakers, in line with “glimpsing” theories of processing speech in noise. These results open a new window into understanding the dynamics of attention as they wax and wane over time, and the listening patterns exhibited by individuals for dealing with the influx of sensory input in complex naturalistic environments.

## Is Attention Stationary?

An underlying assumption of many experimental studies is that participants allocate attention solely to task-relevant stimuli, and that attention remains stationary over time. However, this assumption is probably unwarranted (Weissman et al., 2006; Esterman et al., 2013) since sustaining attention over long periods of time is extremely taxing (Schweizer and Moosbrugger, 2004; Warm et al., 2008; Avisar and Shalev, 2011), and individuals spend a large proportion of the time mind-wandering or “off-task” (Killingsworth and Gilbert, 2010; Boudewyn and Carter, 2018; but see Seli et al., 2018). Yet, empirically testing the studying the frequency and characteristics of attention shifts is operationally difficult since it pertains to participants’ internal state that experimenters do not have direct access to. The use of eye-gaze position as a continuous metric for the locus of momentary overt attention in a dynamic scene in the current study contributes to this endeavor.

Here, we found that indeed, in many participants eye-gaze was not maintained on the target speaker throughout the entire trial. Roughly 30% of participants spent over 10% of each trial looking at places in the environment other than the to-be-attended speaker, across all conditions. Interestingly, this proportion is similar to that reported in previous studies for



**FIGURE 5 |** Gaze-shift locked analysis of speech acoustics. **(A)** Average time-course of the target (top) and distractor (bottom) speech envelopes relative to gaze shift onset ( $t = 0$ ). Horizontal dotted gray lines depict the top and bottom 5%tile of loudness values generated through the permutation procedure of non-gaze-locked acoustics segments. The shaded red areas indicate time-periods where the speech sound-level fell within the lower/upper 5% tile of the distribution, respectively. **(B)** Spectrograms depicting the signal-to-noise ratio (SNR) between the target and distractor speaker(s), surrounding the onset of a gaze-shift, in the single and two-distractor conditions. A reduction in SNR is seen in a 200 ms pre-shift time window, primarily in the higher “unvoiced” portion of the spectrogram (4–8 KHz).

the prevalence of detecting ones’ own name in a so-called unattended message (Cherry, 1953; Wood and Cowan, 1995), an effect attributed by some to rapid attention shifts (Lachter et al., 2004; Beaman et al., 2007; Lin and Yeh, 2014). Although in the current study we did not test whether these participants also gleaned more information from distractors’ speech, we did find that comprehension of the target speaker was reduced as a function of the time spent looking away from the target speaker. Participants were also more likely to miss information from the target-speech during gaze-shifts away, yielding slightly higher mistake-rates. These results emphasize the dynamic nature of attention and attention-shifts, and demonstrate that brief overt attention-shifts can negatively impact speech processing in ecological multi-speaker and multisensory contexts.

They also highlight the importance of studying individual differences in attentional control. In the current study set, we did not collect additional personal data from participants which may have shed light on the source of the observed variability in gaze-patterns across individuals. However, based on previous literature, individual differences may stem from factors such as susceptibility to distraction (Ellermeier and Zimmer, 1997; Cowan et al., 2005; Avisar and Shalev, 2011; Bourel-Ponchel

et al., 2011; Forster and Lavie, 2014; Hughes, 2014), working memory capacity (Conway et al., 2001; Kane and Engle, 2002; Tsuchida et al., 2012; Sörqvist et al., 2013; Hughes, 2014; Naveh-Benjamin et al., 2014; Wiemers and Redick, 2018) or personality traits (Rauthmann et al., 2012; Risko et al., 2012; Baranes et al., 2015; Hoppe et al., 2018). Additional dedicated research is needed to resolve the source of the individual differences observed here.

## Is Eye-Gaze a Good Measure for Attention-Shifts Among Concurrent Speech?

One may ask, to what extent do the current results fully capture the prevalence of attention-shifts, since it is known that these can also occur covertly (Posner, 1980; Petersen and Posner, 2012)? This is a valid concern and indeed the current results should be taken as representing a *lower-bound* for the frequency of attention-shifts and we should assume that attention-shifts are probably more prevalent than observed here. This motivates the future development of complementary methods for quantifying covert shifts of attention among concurrent speech, given the current absence of a reliable metrics.

Another concern that may be raised with regard to the current results is that individuals may maintain attention to the target speaker even while looking elsewhere, and hence the gaze-shifts measured here might not reflect true shifts of attention. Although in principle this could be possible, previous research shows that this is probably not the default mode of listening under natural audiovisual conditions. Rather, a wealth of studies demonstrate a tight link between gaze-shifts and attention-shifts (Chelazzi et al., 1995; Deubel and Schneider, 1996; Grosbras et al., 2005; Szinte et al., 2018) and gaze is widely utilized experimentally as a proxy for the locus of visuospatial attention (Gredebäck et al., 2009; Linse et al., 2017). In multi-speaker contexts, it has been shown that participants tend to move their eyes towards the location of attended speech sounds (Gopher and Kahneman, 1971; Gopher, 1973). Similarly, looking towards the location of distractor-speech significantly reduces intelligibility and memory for attended speech and increases intrusions from distractor speech (Reisberg et al., 1981; Spence et al., 2000; Yi et al., 2013). This is in line with the current finding of a negative correlation between the time spent looking at the target speaker and speech comprehension, and higher mistake-rates during gaze-shifts, which further link overt gaze to selective attention to speech. Studies on audiovisual speech processing further indicate that looking at the talking face increases speech intelligibility and neural selectivity for attended speech (Sumby and Pollack, 1954; Zion Golumbic et al., 2013a; Lou et al., 2014; Crosse et al., 2016; Park et al., 2016), even when the video is not informative about the content of speech (Kim and Davis, 2003; Schwartz et al., 2004), and eye-gaze is particularly utilized for focusing attention to speech under adverse listening condition (Yi et al., 2013). Taken together, current findings support the interpretation that gaze-shifts reflect shifts in attention away from the target speaker, in line with the limited resources perspective of attention (Lavie et al., 2004; Esterman et al., 2014), making eye-gaze a useful and reliable metric for studying the dynamics of attention to naturalistic audio-visual speech. Interestingly, this metric has recently been capitalized on for use in assistive listening devices, utilizing eye-gaze direction to indicate the direction of a listener's attention (Favre-Felix et al., 2017; Kidd, 2017). That said, gaze-position is likely only one of several factors in determining successful speech comprehension in multi-speaker environments (e.g., SNR level, audio-visual congruency, engagement in content etc.), as suggested by the significant yet still moderate effect-sizes found here.

## Listening Between the Gaps—What Prompts Attention Shifts Among Concurrent Speech?

Besides characterizing the prevalence and behavioral consequences of attention-shifts in audio-visual multi-talker contexts, it is also critical to understand what prompts these shifts. Here we tested whether there are aspects of the scene acoustics that can be associated with attention-shifts away from the target speaker. We specifically tested two hypotheses: (1) that attention is captured exogenously by highly salient sensory events in distracting speech (Wood and Cowan, 1995; Itti and

Koch, 2000; Kayser et al., 2005); and (2) that attention-shifts occur during brief pauses in speech acoustics that momentarily unmask the competing sounds (Lavie et al., 2004; Cooke, 2006).

Regarding the first hypothesis, the current data suggest that distractor saliency is not a primary factor in prompting gaze-shifts. Since gaze-shifts were just as prevalent in the NoD condition as in conditions that contained distractors and since no consistent increase in distractor loudness was observed near gaze-shifts, we conclude that the gaze-shifts performed by participants do not necessarily reflect exogenous attentional capture by distractor saliency. This is in line with previous studies suggesting that sensory saliency is less effective in drawing exogenous attention in dynamic scenarios relative to the stationary contexts typically used in laboratory experiments (Smith et al., 2013).

Rather, our current results seem to support the latter hypothesis that attention-shifts are prompted by momentary acoustic release-from-masking. We find that gaze-shifts occurred more consistently ~200–250 ms after instances of low acoustic intensity in both target and distractor sounds and low SNR. This time-scale is on-par with the initiation time for saccades (Gilchrist, 2011), and suggests that momentary reduction in masking provide an opportunity for the system to shift attention between speakers. This pattern fits with accounts for comprehension of speech-in-noise, suggesting that listeners utilize brief periods of unmasking or low SNR to glean and piece together information for deciphering speech content (“acoustic glimpsing”; Cooke, 2006; Li and Loizou, 2007; Vestergaard et al., 2011; Rosen et al., 2013). Although this acoustic-glimpsing framework is often used to describe how listeners maintain intelligibility of target-speech in noise, it has not been extensively applied to studying *shifts* of attention among concurrent speech. The current results suggest that brief gaps in the audio or periods of low SNR may serve as triggers for momentary attention shifts, which can manifest overtly (as demonstrated here), and perhaps also covertly. Interestingly, a previous study found that eye-blinks also tend to occur more often around pauses when viewing and listening to audio-visual speech (Nakano and Kitazawa, 2010), pointing to a possible link between acoustic glimpsing and a reset in the oculomotor system, creating optimal conditions for momentary attention-shifts.

## CONCLUSION

There is growing understanding that in order to really understand the human cognitive system, it needs to be studied in contexts relevant for real-life behavior, and that tightly constrained artificial laboratory paradigms do not always generalize to real-life (Kingstone et al., 2008; Marius't Hart et al., 2009; Foulsham et al., 2011; Risko et al., 2016; Rochais et al., 2017; Hoppe et al., 2018). The current study represents the attempt to bridge this gap between the laboratory and real-life, by studying how individuals spontaneously deploy overt attention in a naturalistic virtual-reality environment. Using this approach, the current study highlights the characteristics and individual differences in selective attention to speech under



naturalistic listening conditions. This pioneering work opens up new horizons for studying how attention operates in real-life and understanding the factors contributing to success as well as the difficulties in paying attention to speech in noisy environments.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The study was approved by the Institutional Ethics Committee at Bar-Ilan University, and the research was conducted according

to the guidelines of the committee. Signed informed consent was obtained from each participant prior to the experiment.

## AUTHOR CONTRIBUTIONS

EZG designed the study, oversaw data collection and analysis. KS-C collected and analyzed the data. Both authors wrote the article.

## FUNDING

This work was supported by the Israel Science Foundation I-Core Center for Excellence 51/11, and by the United States–Israel Binational Science Foundation grant #2015385.

## REFERENCES

- Anderson, N. C., Ort, E., Kruijine, W., Meeter, M., and Donk, M. (2015). It depends on *when* you look at it: salience influences eye movements in natural scene viewing and search early in time. *J. Vis.* 15:9. doi: 10.1167/15.5.9
- Atal, B. S., and Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *J. Acoust. Soc. Am.* 50, 637–655. doi: 10.1121/1.1912679
- Avisar, A., and Shalev, L. (2011). Sustained attention and behavioral characteristics associated with ADHD in adults. *Appl. Neuropsychol.* 18, 107–116. doi: 10.1080/09084282.2010.547777
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Baranes, A., Oudeyer, P.-Y., and Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Res.* 117, 81–90. doi: 10.1016/j.visres.2015.10.009
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Beaman, C. P., Bridges, A. M., and Scott, S. K. (2007). From dichotic listening to the irrelevant sound effect: a behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex* 43, 124–134. doi: 10.1016/s0010-9452(08)70450-7
- Bompas, A., and Sumner, P. (2009). Temporal dynamics of saccadic distraction. *J. Vis.* 9, 17–17. doi: 10.1167/9.9.17
- Boudewyn, M. A., and Carter, C. S. (2018). I must have missed that:  $\alpha$ -band oscillations track attention to spoken language. *Neuropsychologia* 117, 148–155. doi: 10.1016/j.neuropsychologia.2018.05.024
- Bourel-Ponchel, E., Querné, L., Le Moing, A. G., Delignières, A., de Broca, A., and Berquin, P. (2011). Maturation of response time and attentional control in ADHD: evidence from an attentional capture paradigm. *Eur. J. Paediatr. Neurol.* 15, 123–130. doi: 10.1016/j.ejpn.2010.08.008
- Broadbent, D. E. (1958). “Selective listening to speech,” in *Perception and Communication*, ed. D. E. Broadbent (Elmsford, NY: Pergamon Press), 11–35.
- Calandruccio, L., Buss, E., Bencheck, P., and Jett, B. (2018). Does the semantic content or syntactic regularity of masker speech affect speech-on-speech recognition? *J. Acoust. Soc. Am.* 144, 3289–3302. doi: 10.1121/1.5081679
- Chelazzi, L., Biscaldi, M., Corbetta, M., Peru, A., Tassinari, G., and Berlucchi, G. (1995). Oculomotor activity and visual spatial attention. *Behav. Brain Res.* 71, 81–88. doi: 10.1016/0166-4328(95)00134-4
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Conway, A. R. A., Cowan, N., and Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon. Bull. Rev.* 8, 331–335. doi: 10.3758/bf03196169
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.* 119, 1562–1573. doi: 10.1121/1.2166600
- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi: 10.1016/j.cogpsych.2004.12.001
- Crosse, M. J., Di Liberto, G. M., and Lalor, E. C. (2016). Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *J. Neurosci.* 36, 9888–9895. doi: 10.1523/jneurosci.1396-16.2016
- Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* 36, 1827–1837. doi: 10.1016/0042-6989(95)00294-4
- Deutsch, J. A., and Deutsch, D. (1963). Attention: some theoretical considerations. *Psychol. Rev.* 70, 80–90. doi: 10.1037/h0039515
- Ding, N., and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U S A* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Ding, N., and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychol. Rev.* 87, 272–300. doi: 10.1037/0033-295x.87.3.272
- Ellermeier, W., and Zimmer, K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. *J. Acoust. Soc. Am.* 102, 2191–2199. doi: 10.1121/1.419596
- Escera, C., Yago, E., Corral, M.-J., Corbera, S., and Nuñez, M. I. (2003). Attention capture by auditory significant stimuli: semantic analysis follows attention switching. *Eur. J. Neurosci.* 18, 2408–2412. doi: 10.1046/j.1460-9568.2003.02937.x
- Esterman, M., Noonan, S. K., Rosenberg, M., and Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* 23, 2712–2723. doi: 10.1093/cercor/bhs261
- Esterman, M., Rosenberg, M. D., and Noonan, S. K. (2014). Intrinsic fluctuations in sustained attention and distractor processing. *J. Neurosci.* 34, 1724–1730. doi: 10.1523/JNEUROSCI.2658-13.2014
- Favre-Felix, A., Graversen, C., Dau, T., and Lunner, T. (2017). “Real-time estimation of eye gaze by in-ear electrodes. in,” in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Seogwipo: IEEE), 4086–4089.
- Forster, S., and Lavie, N. (2014). Distracted by your mind? Individual differences in distractibility predict mind wandering. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 251–260. doi: 10.1037/a0034108
- Foulsham, T., Walker, E., and Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Res.* 51, 1920–1931. doi: 10.1016/j.visres.2011.07.002
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.* 115, 2246–2256. doi: 10.1121/1.1689343



- Gilchrist, I. D. (2011). "Saccades," in *The Oxford Handbook of Eye Movements*, eds S. P. Liversedge, I. D. Gilchrist and S. Everling (Oxford, UK: Oxford University Press), 85–94.
- Gopher, D. (1973). Eye-movement patterns in selective listening tasks of focused attention. *Percept. Psychophys.* 14, 259–264. doi: 10.3758/bf03212387
- Gopher, D., and Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Percept. Mot. Skills* 33, 1335–1342. doi: 10.2466/pms.1971.33.3f.1335
- Gredebäck, G., Johnson, S., and von Hofsten, C. (2009). Eye tracking in infancy research. *Dev. Neuropsychol.* 35, 1–19. doi: 10.1080/87565640903325758
- Grosbras, M. H., Laird, A. R., and Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention and gaze perception. *Hum. Brain Mapp.* 25, 140–154. doi: 10.1002/hbm.20145
- Hoppe, S., Loetscher, T., Morey, S. A., and Bulling, A. (2018). Eye movements during everyday behavior predict personality traits. *Front. Hum. Neurosci.* 12:105. doi: 10.3389/fnhum.2018.00105
- Hughes, R. W. (2014). Auditory distraction: a duplex-mechanism account. *Psych J.* 3, 30–41. doi: 10.1002/pchj.44
- Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/s0042-6989(99)00163-7
- Kane, M. J., and Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention and general fluid intelligence: an individual-differences perspective. *Psychon. Bull. Rev.* 9, 637–671. doi: 10.3758/bf03196323
- Kayser, C., Petkov, C. I., Lippert, M., and Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: an auditory saliency map. *Curr. Biol.* 15, 1943–1947. doi: 10.1016/j.cub.2005.09.040
- Kidd, G. Jr. (2017). Enhancing auditory selective attention using a visually guided hearing aid. *J. Speech Lang. Hear. Res.* 60, 3027–3038. doi: 10.1044/2017\_JSLHR-H-17-0071
- Killingsworth, M. A., and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science* 330:932. doi: 10.1126/science.1192439
- Kim, J., and Davis, C. (2003). Hearing foreign voices: does knowing what is said affect visual-masked-speech detection? *Perception* 32, 111–120. doi: 10.1068/p3466
- Kingstone, A., Smilek, D., and Eastwood, J. D. (2008). Cognitive ethology: a new approach for studying human cognition. *Br. J. Psychol.* 99, 317–340. doi: 10.1348/000712607X251243
- Koller, M. (2016). robustlmm: an R package for robust estimation of linear mixed-effects models. *J. Stat. Softw.* 75, 1–24. doi: 10.18637/jss.v075.i06
- Lachter, J., Forster, K. I., and Ruthruff, E. (2004). Forty-five years after broadbent (1958): still no identification without attention. *Psychol. Rev.* 111, 880–913. doi: 10.1037/0033-295X.111.4.880
- Lavie, N., Hirst, A., de Fockert, J. W., and Viding, E. (2004). Load theory of selective attention and cognitive control. *J. Exp. Psychol. Gen.* 133, 339–354. doi: 10.1037/0096-3445.133.3.339
- Li, N., and Loizou, P. C. (2007). Factors influencing glimpsing of speech in noise. *J. Acoust. Soc. Am.* 122, 1165–1172. doi: 10.1121/1.2749454
- Lin, S.-H., and Yeh, Y.-Y. (2014). Attentional load and the consciousness of one's own name. *Conscious. Cogn.* 26, 197–203. doi: 10.1016/j.concog.2014.03.008
- Linse, K., Rüger, W., Joos, M., Schmitz-Peiffer, H., Storch, A., and Hermann, A. (2017). Eye-tracking-based assessment suggests preserved well-being in locked-in patients. *Ann. Neurol.* 81, 310–315. doi: 10.1002/ana.24871
- Lou, Y., Yoon, J. W., and Huh, H. (2014). Modeling of shear ductile fracture considering a changeable cut-off value for stress triaxiality. *Int. J. Plast.* 54, 56–80. doi: 10.1016/j.iplas.2013.08.006
- Marius's Hart, B. M., Vockeroth, J., Schumann, F., Bartl, K., Schneider, E., König, P., et al. (2009). Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. *Vis. cogn.* 17, 1132–1158. doi: 10.1080/13506280902812304
- McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027. doi: 10.1016/j.cub.2009.09.005
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Nakano, T., and Kitazawa, S. (2010). Eyeblick entrainment at breakpoints of speech. *Exp. Brain Res.* 205, 577–581. doi: 10.1007/s00221-010-2387-z
- Naveh-Benjamin, M., Kilb, A., Maddox, G. B., Thomas, J., Fine, H. C., Chen, T., et al. (2014). Older adults do not notice their names: a new twist to a classic attention task. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 1540–1550. doi: 10.1037/xlm0000020
- Neely, C., and LeCompte, D. (1999). The importance of semantic similarity to the irrelevant speech effect. *Mem. Cogn.* 27, 37–44. doi: 10.3758/bf03201211
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355
- Park, H., Kayser, C., Thut, G., and Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *Elife* 5:e14521. doi: 10.7554/elife.14521
- Parmentier, F. B. R., Pacheco-Unguetti, A. P., and Valero, S. (2018). Food words distract the hungry: evidence of involuntary semantic processing of task-irrelevant but biologically-relevant unexpected auditory words. *PLoS One* 13:e0190644. doi: 10.1371/journal.pone.0190644
- Petersen, S. E., and Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89. doi: 10.1146/annurev-neuro-062111-150525
- Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/00335558008248231
- Rauthmann, J. F., Seubert, C. T., Sachse, P., and Furtner, M. R. (2012). Eyes as windows to the soul: gazing behavior is related to personality. *J. Res. Pers.* 46, 147–156. doi: 10.1016/j.jrp.2011.12.010
- Raveh, D., and Lavie, N. (2015). Load-induced inattention deafness. *Atten. Percept. Psychophys.* 77, 483–492. doi: 10.3758/s13414-014-0776-2
- Reisberg, D., Scheiber, R., and Potemkin, L. (1981). Eye position and the control of auditory attention. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 318–323. doi: 10.1037/0096-1523.7.2.318
- Risko, E. F., Anderson, N. C., Lanthier, S., and Kingstone, A. (2012). Curious eyes: individual differences in personality predict eye movement behavior in scene-viewing. *Cognition* 122, 86–90. doi: 10.1016/j.cognition.2011.08.014
- Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the fourth wall of cognitive science. *Curr. Dir. Psychol. Sci.* 25, 70–74. doi: 10.1177/0963721415617806
- Rochais, C., Henry, S., and Hausberger, M. (2017). Spontaneous attention-capture by auditory distractors as predictor of distractibility: a study of domestic horses (*Equus caballus*). *Sci. Rep.* 7:15283. doi: 10.1038/s41598-017-15654-5
- Röer, J. P., Körner, U., Buchner, A., and Bell, R. (2017). Attentional capture by taboo words: a functional view of auditory distraction. *Emotion* 17, 740–750. doi: 10.1037/emo0000274
- Rosen, S., Souza, P., Ekelund, C., and Majeed, A. A. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *J. Acoust. Soc. Am.* 133, 2431–2443. doi: 10.1121/1.4794379
- R Development Core Team. (2012). R: A Language and Environment for Statistical Computing. Vienna: R foundation for Statistical Computing. Available online at: <http://www.R-project.org/>.
- Schomaker, J., Walper, D., Wittmann, B. C., and Einhäuser, W. (2017). Attention in natural scenes: affective-motivational factors guide gaze independently of visual salience. *Vision Res.* 133, 161–175. doi: 10.1016/j.visres.2017.02.003
- Schwartz, J.-L., Berthommier, F., and Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, B69–B78. doi: 10.1016/s0010-0277(04)00054-x
- Schweizer, K., and Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence* 32, 329–347. doi: 10.1016/j.intell.2004.06.006
- Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., and Schacter, D. L. (2018). How pervasive is mind wandering, really? *Conscious. Cogn.* 66, 74–78. doi: 10.1016/j.concog.2018.10.002
- Simpson, S. A., and Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *J. Acoust. Soc. Am.* 118, 2775–2778. doi: 10.1121/1.2062650

- Smith, T. J., Lamont, P., and Henderson, J. M. (2013). Change blindness in a dynamic scene due to endogenous override of exogenous attentional cues. *Perception* 42, 884–886. doi: 10.1068/p7377
- Sörqvist, P., Marsh, J. E., and Nösl, A. (2013). High working memory capacity does not always attenuate distraction: bayesian evidence in support of the null hypothesis. *Psychon. Bull. Rev.* 20, 897–904. doi: 10.3758/s13423-013-0419-y
- Spence, C., Ranson, J., and Driver, J. (2000). Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. *Percept. Psychophys.* 62, 410–424. doi: 10.3758/bf03205560
- Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Szinte, M., Jonikaitis, D., Rangelov, D., and Deubel, H. (2018). Pre-saccadic remapping relies on dynamics of spatial attention. *Elife* 7:e37598. doi: 10.7554/elifesciences.37598
- Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *Am. J. Psychol.* 77, 533–546. doi: 10.2307/1420765
- Tsuchida, Y., Katayama, J., and Murohashi, H. (2012). Working memory capacity affects the interference control of distractors at auditory gating. *Neurosci. Lett.* 516, 62–66. doi: 10.1016/j.neulet.2012.03.057
- Vergauwe, E., Barrouillet, P., and Camos, V. (2010). Do mental processes share a domain-general resource? *Psychol. Sci.* 21, 384–390. doi: 10.1177/0956797610361340
- Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *J. Acoust. Soc. Am.* 130, 429–439. doi: 10.1121/1.3596462
- Walker, F., Buckner, B., Anderson, N. C., Schreij, D., and Theeuwes, J. (2017). Looking at paintings in the vincent van gogh museum: eye movement patterns of children and adults. *PLoS One* 12:e0178912. doi: 10.1371/journal.pone.0178912
- Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Hum. Factors J. Hum. Factors Ergon. Soc.* 50, 433–441. doi: 10.1518/001872008X312152
- Weissman, D. H., Roberts, K. C., Visscher, K. M., and Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9, 971–978. doi: 10.1038/nn1727
- Wiemers, E. A., and Redick, T. S. (2018). Working memory capacity and intra-individual variability of proactive control. *Acta Psychol.* 182, 21–31. doi: 10.1016/j.actpsy.2017.11.002
- Wood, N., and Cowan, N. (1995). The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 255–260. doi: 10.1037/0278-7393.21.1.255
- Yi, A., Wong, W., and Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *J. Speech Lang. Hear. Res.* 56, 471–480. doi: 10.1044/1092-4388(2012/10-0288)
- Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., and Poeppel, D. (2013a). Visual input enhances selective speech envelope tracking in auditory cortex at a “Cocktail Party”. *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013b). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

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

Copyright © 2019 Shavit-Cohen and Zion Golumbic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology

Panagiotis Kourtesis<sup>1,2,3,4\*</sup>, Simona Collina<sup>3,4</sup>, Leonidas A. A. Doulmas<sup>2</sup> and Sarah E. MacPherson<sup>1,2</sup>

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Eugene Nalivaiko,  
University of Newcastle, Australia  
Mark Dennison,  
United States Army Research  
Laboratory, United States  
Justin Maximilian Mittelstädt,  
Institute of Aerospace Medicine,  
German Aerospace Center (DLR),  
Germany

### \*Correspondence:

Panagiotis Kourtesis  
pkourtes@exseed.ed.ac.uk

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 12 August 2019

**Accepted:** 11 November 2019

**Published:** 26 November 2019

### Citation:

Kourtesis P, Collina S,  
Doulmas LAA and MacPherson SE  
(2019) Validation of the Virtual Reality  
Neuroscience Questionnaire:  
Maximum Duration of Immersive  
Virtual Reality Sessions Without  
the Presence of Pertinent Adverse  
Symptomatology.  
Front. Hum. Neurosci. 13:417.  
doi: 10.3389/fnhum.2019.00417

There are major concerns about the suitability of immersive virtual reality (VR) systems (i.e., head-mounted display; HMD) to be implemented in research and clinical settings, because of the presence of nausea, dizziness, disorientation, fatigue, and instability (i.e., VR induced symptoms and effects; VRISE). Research suggests that the duration of a VR session modulates the presence and intensity of VRISE, but there are no suggestions regarding the appropriate maximum duration of VR sessions. The implementation of high-end VR HMDs in conjunction with ergonomic VR software seems to mitigate the presence of VRISE substantially. However, a brief tool does not currently exist to appraise and report both the quality of software features and VRISE intensity quantitatively. The Virtual Reality Neuroscience Questionnaire (VRNQ) was developed to assess the quality of VR software in terms of user experience, game mechanics, in-game assistance, and VRISE. Forty participants aged between 28 and 43 years were recruited (18 gamers and 22 non-gamers) for the study. They participated in 3 different VR sessions until they felt weary or discomfort and subsequently filled in the VRNQ. Our results demonstrated that VRNQ is a valid tool for assessing VR software as it has good convergent, discriminant, and construct validity. The maximum duration of VR sessions should be between 55 and 70 min when the VR software meets or exceeds the parsimonious cut-offs of the VRNQ and the users are familiarized with the VR system. Also, the gaming experience does not seem to affect how long VR sessions should last. Also, while the quality of VR software substantially modulates the maximum duration of VR sessions, age and education do not. Finally, deeper immersion, better quality of graphics and sound, and more helpful in-game instructions and prompts were found to reduce VRISE intensity. The VRNQ facilitates the brief assessment and reporting of the quality of VR software features and/or the intensity of VRISE, while its minimum and

parsimonious cut-offs may appraise the suitability of VR software for implementation in research and clinical settings. The findings of this study contribute to the establishment of rigorous VR methods that are crucial for the viability of immersive VR as a research and clinical tool in cognitive neuroscience and neuropsychology.

**Keywords:** virtual reality, VRISE, VR sickness, cybersickness, neuroscience, neuropsychology, psychology, motion sickness

## INTRODUCTION

Immersive virtual reality (VR) has emerged as a novel tool for neuroscientific and neuropsychological research (Bohil et al., 2011; Parsons, 2015; Parsons et al., 2018). Nevertheless, there are concerns pertinent to implementing VR in research and clinical settings, especially regarding the head-mounted display (HMD) systems (Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). A primary concern is the presence of adverse physiological symptoms (i.e., nausea, dizziness, disorientation, fatigue, and postural instability), which are referred to as motion-sickness, cybersickness, VR sickness or VR induced symptoms and effects (VRISE) (Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017).

Longer durations in a virtual environment have been associated with a higher probability of experiencing VRISE, while the intensity of VRISE also appears to increase proportionally with the duration of the VR session (Sharples et al., 2008). However, extensive linear and angular accelerations provoke intense VRISE, even in a short period of time (McCauley and Sharkey, 1992; LaViola, 2000; Gavvani et al., 2018). VRISE may place the health and safety of the participants or patients at risk of experiencing adverse physiological symptoms (Parsons et al., 2018). Research has also shown that VRISE induce significant decreases in reaction times and overall cognitive performance (Nalivaiko et al., 2015; Nesbitt et al., 2017; Mittelstaedt et al., 2019), as well as substantially increasing body temperatures and heart rates (Nalivaiko et al., 2015), which may compromise physiological data acquisition. Furthermore, the presence of VRISE has been found to significantly augment cerebral blood flow and oxyhemoglobin concentration (Gavvani et al., 2018), electrical brain activity (Arafat et al., 2018), and the connectivity between stimulus-response regions and nausea-processing regions (Toschi et al., 2017). Thus, VRISE appear to confound the reliability of neuropsychological, physiological, and neuroimaging data (Kourtesis et al., 2019).

To our knowledge, there do not appear to be any guidelines as to the appropriate maximum duration of VR research and clinical sessions to evade or alleviate the presence of VRISE. Recently, our work has suggested that VRISE are substantially reduced or prevented by VR software that facilitates ergonomic navigation (e.g., physical movement) and interaction (e.g., direct-hand tracking) facilitated by the hardware capabilities (e.g., motion tracking) of commercial, contemporary VR HMDs comparable to or more advanced than the HTC Vive and/or Oculus Rift (Kourtesis et al., 2019). However, there are other factors such as the type of display and its features that may also induce or

reduce VRISE (Mittelstaedt et al., 2018; Kourtesis et al., 2019). Nevertheless, we note that adequate technological competence is required to be able to implement appropriate VR hardware and/or software. In an attempt to reach a methodological consensus, we have proposed minimum hardware and software features, which appraise the suitability of VR hardware and software (see **Table 1**; Kourtesis et al., 2019).

While VRISE may occur for various reasons, they are predominantly the undesirable outcomes of hardware and software insufficiencies (e.g., low resolution and refresh rates of the image, a narrow field of view, non-ergonomic interactions, and inappropriate navigation modes) (de França and Soares, 2017; Palmisano et al., 2017; Kourtesis et al., 2019). In terms of hardware, the technical specifications of the computer (e.g., processing power and graphics card), and VR HMD (e.g., the field of view, refresh rate, and resolution) suffice to appraise their suitability (Kourtesis et al., 2019). However, there is not a tool to quantify the software's recommended features, as well as the intensity of VRISE (Kourtesis et al., 2019). Currently, the most frequently used measure of VRISE is the simulator sickness questionnaire (SSQ), which only considers the symptoms pertinent to simulator sickness (Kennedy et al., 1993). However, the SSQ does not assess software attributes (Kennedy et al., 1993), and there is an argument that simulator sickness symptomatology may not be identical to VRISE (Stanney et al., 1997). There is thus a need for a tool, which will enable researchers to assess both the suitability of VR software, as well as the intensity of VRISE.

Our recent technological literature review of VR hardware and software pinpointed four domains that should be considered in the development or selection of VR research/clinical software (Kourtesis et al., 2019). The domains are user experience, game mechanics, in-game assistance, and VRISE. Each domain has five criteria that should be met to ensure the appropriateness of the software (see **Table 1**). Also, in the same study, the meta-analysis of 44 VR neuroscientific studies revealed that most of the studies did not report quantitatively VR software's quality and/or VRISE intensity (Kourtesis et al., 2019). In an attempt to provide a brief tool for the appraisal of VR research/clinical software features and VRISE intensity, we developed the virtual reality neuroscience questionnaire (VRNQ), which includes twenty questions that address five criteria under each domain. This study aimed to validate the VRNQ and provide suggestions for the duration of VR research/clinical sessions. We also considered the gaming experience of the participants to examine whether this may affect the duration of the VR sessions. Lastly, we investigated the software predictors of VRISE as measured by the VRNQ.



## MATERIALS AND METHODS

### Participants

Forty participants (21 males) aged between 28 and 43 years ( $M = 32.08$ ;  $SD = 3.54$ ) and an educational level between 12 and 16 full-time years of education ( $M = 14.25$ ;  $SD = 1.37$ ) were recruited for the study. Eighteen participants (10 males) identified themselves as gamers through self-report and 22 as non-gamers (11 males). The gamer experience was a dichotomous variable (i.e., gamer or non-gamer) based on the participants' response to a question asking whether they played games on a weekly basis. The participants responded to a call disseminated through mailing lists at the University of Edinburgh and social media. The study was approved by the Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. All participants provided written informed consent prior to taking part.

### Material

#### Hardware

An HTC Vive HMD with two lighthouse-stations for motion tracking was used with two HTC Vive's wands with 6 degrees of freedom (DoF) to facilitate navigation and interactions within the environment (Kourtesis et al., 2019). The VR area where the participants were immersed and interacted with the virtual environments was 4.4 m<sup>2</sup>. Additionally, the HMD was connected to a laptop with an Intel Core i7 7700HQ processor at 2.80 GHz, 16 GB RAM, a 4095 MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100 (SATA) hard disk, and Realtek High Definition Audio.

#### Software

Three VR games were selected, which included ergonomic navigation (i.e., teleportation and physical mobility) and interactions (i.e., 6 DoF wands simulating hand movements) with the virtual environment. In line with Kourtesis et al. (2019), the VR software inclusion criteria (see **Table 1**) were: (1) ergonomic interactions which simulate real-life hand movements; (2) a navigation system which uses teleportation and physical mobility; (3) comprehensible tutorials pertinent to the controls; and (4) in-game instructions and prompts which assist the user in

orientating and interacting with the virtual environment. The suitability of the VR software for both gamers and non-gamers was also considered. The selected VR games which met the above software criteria were: (1) "Job Simulator" (Session 1)<sup>1</sup>; (2) "The Lab" (Session 2)<sup>2</sup>; and (3) "Rick and Morty: Virtual Rick-ality" (Session 3)<sup>3</sup>. In "Job Simulator," the participant becomes an employee who has several occupations, such as a cook (preparing simply recipes), car mechanic (doing rudimentary tasks e.g., replacing faulty parts), and an office worker (making calls and sending emails). In "The Lab," the participant needs to complete several mini-games like slingshot (shooting down piles of boxes), longbow (shooting down invaders), xortex (spaceship-battles), postcards (visiting exotic places), human medical scan (exploring the human body), solar system (exploring the solar system), robot repair (repairing a robot), and secret shop (exploring a magical shop). In "Rick and Morty: Virtual Rick-ality," the participant needs to complete several imaginary home-chores as in "Job Simulator," though, in this case, the participant is required to follow a sequence of tasks according to a fictional storyline.

### Virtual Reality Neuroscience Questionnaire (VRNQ)

The VRNQ measures the quality of user experience, game mechanics, and in-game assistance, as well as the intensity of VRUSE. The VRNQ involves 20 questions where each question corresponds to one of the criteria for appropriate VR research/clinical software (e.g., the level of immersion; see **Table 1**). The 20 questions are grouped under four domains, where each domain encompasses five questions. Hence, VRNQ produces a total score corresponding to the overall quality of VR software, as well as four sub-scores (i.e., user experience, game mechanics, in-game assistance, VRUSE). The user experience score is based on the intensity of the immersion, the level of enjoyment, as well as the quality of the graphics, sound, and VR technology (i.e., internal and external hardware). The game mechanics' score depends on the ease to navigate, physically move, and interact with the virtual environment (i.e., use, pick

<sup>1</sup>[https://store.steampowered.com/app/448280/Job\\_Simulator/](https://store.steampowered.com/app/448280/Job_Simulator/)

<sup>2</sup>[https://store.steampowered.com/app/450390/The\\_Lab/](https://store.steampowered.com/app/450390/The_Lab/)

<sup>3</sup>[https://store.steampowered.com/app/469610/Rick\\_and\\_Morty\\_Virtual\\_Rickality/](https://store.steampowered.com/app/469610/Rick_and_Morty_Virtual_Rickality/)

**TABLE 1** | Domains and criteria for VR research/clinical software.

Domains	User experience	Game mechanics	In-game assistance	VRUSE
CRITERIA	An adequate level of immersion	A suitable navigation system (e.g., teleportation)	Digestible tutorials	Absence or insignificant presence of nausea
	Pleasant VR experience	Availability of physical movement	Helpful tutorials	Absence or insignificant presence of disorientation
	High quality graphics	Naturalistic picking/placing of items	Adequate duration of tutorials	Absence or insignificant presence of dizziness
	High quality sounds	Naturalistic use of items	Helpful in-game instructions	Absence or insignificant presence of fatigue
	Suitable hardware (HMD and computer)	Naturalistic 2-handed interaction	Helpful in-game prompts	Absence or insignificant presence of instability

Derived from Kourtesis et al. (2019).



and place, and hold items; two-handed interactions). The in-game assistance score appraises the quality of the tutorial(s), in-game instructions (e.g., description of the aim of the task), and prompts (e.g., arrows showing the direction). The VRISE are evaluated by the intensity of primary adverse symptoms and effects pertinent to VR (i.e., nausea, disorientation, dizziness, fatigue, and instability). VRNQ responses are indicated on a 7-point Likert style scale, ranging from 1 = extremely low to 7 = extremely high. The higher scores indicate a more positive outcome; this also applies to the evaluation of VRISE intensity. Hence, the higher VRISE score indicates a lower intensity of VRISE (i.e., 1 = extremely intense feeling, 2 = very intense feeling, 3 = intense feeling, 4 = moderate feeling, 5 = mild feeling, 6 = very mild feeling, 7 = absent). The VRNQ also includes space under each question, where the participant may provide optional qualitative feedback. For further details, please see the VRNQ in **Supplementary Material**.

## Procedure

The participants individually attended three separate VR sessions; in each session, they were immersed in different VR software. The period between each session was 1 week for each participant (i.e., 3 weeks in total). The participants went through an induction pertinent to the VR software for that session and the specific HMD and controllers used (i.e., HTC Vive and its 6DoF wands-controllers) before being immersed. Subsequently, the participants were asked to play the respective VR game until they completed it, or they felt any discomfort or fatigue. The duration of each VR session was recorded from the time the software was started until the participant expressed that they wanted to discontinue. At the end of each session, participants were asked to complete the VRNQ. The “Job Simulator” was always used in the 1st session, “The Lab” was always used in the 2nd session, and “Rick and Morty: Virtual Rick-ality” was always used in the 3rd session.

## Statistical Analyses

A reliability analysis of the VRNQ was conducted to calculate Cronbach's alpha and inspect whether the items have adequate internal consistency for research and clinical purposes. A Cronbach's alpha of 0.70–1.00 indicates good to excellent internal consistency (Nunally and Bernstein, 1994). A confirmatory factor analysis (CFA) was performed to examine the construct validity of the VRNQ in terms of convergent and discriminant validity (Cole, 1987). The reliability analysis and CFA were conducted using AMOS (version 24) (Arbuckle, 2014), and IBM Statistical Package for the Social Sciences (SPSS) 24.0 (Ibm Corp, 2016). Several tests for goodness of fit were implemented to allow the evaluation of VRNQ's structure. The (CFI), Tucker Lewis index (TLI), standardized root mean square residual (SRMR), and the root mean squared error of approximation (RMSEA) were used to assess model fit. A CFI and TLI equal to or greater than 0.90 indicate good structural model fit to the data (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). An SRMR and RMSEA less than 0.08 postulate a good fit to the data (Hu and Bentler, 1999; Hopwood and Donnellan, 2010). Lastly,

the variance of the results was assessed by dividing the  $\chi^2$  by the degrees of freedom (df), which is an indicator of the sample distribution (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010).

The reliability and confirmatory factor analyses were conducted based on 120 observations (40 participants \* 3 sessions with different software). The *a priori* sample size calculator for structural equation models was used to calculate the minimum sample size for model structure. This calculator uses the error function formula, the lower bound sample size formula for a structural equation model, and the normal distribution cumulative distribution function (Soper, 2019a), which are in perfect agreement with the recommendations for statistical power analysis for the behavioral sciences (Cohen, 2013). A sample size of 100 observations was suggested as the minimum for conducting CFA to examine the model structure with statistical power equal to or greater than 0.80. Hence, the 120 observations in our sample appear adequate to conduct a CFA with statistical power equal to or greater than 0.80.

Bayesian Pearson correlation analyses were conducted to examine whether any of the demographic variables were significantly associated with the VRNQ total score and sub-scores, or the length of the VR sessions. Bayesian paired samples *t*-tests were performed to investigate possible differences between each session's duration, as well as the VRNQ results for each VR game. Also, a Bayesian independent samples *t*-test examined whether there were any differences between gamers and non-gamers in the duration of the session. Lastly, a Bayesian linear regression was performed to examine the predictors of VRISE, where the Jeffreys–Zellner–Siow (JZS) mixed g-prior was used for the selection of the best model. JZS has the computational advantages of a g-prior in conjunction with the theoretical advantages of a Cauchy prior, which are valuable in variable selection for the best model (Liang et al., 2008; Rouder and Morey, 2012). For all the analyses, a Bayes Factor ( $BF_{10}$ )  $\geq 10$  was set for statistical inference, which indicates strong evidence in favor of the alternative hypothesis (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). All the Bayesian analyses were performed using JASP (Version 0.8.1.2) (Jasp Team, 2017). The Bayesian Pearson correlation analyses and Bayesian linear regression analysis were conducted based on 120 observations (40 participants \* 3 different software sessions). The *post hoc* statistical power calculator was used to calculate the observed power of the best model using Bayesian linear regression analysis (Soper, 2019b).

## RESULTS

### Reliability Analysis and CFA

The reliability analysis demonstrated good to excellent Cronbach's  $\alpha$  for each domain of the VRNQ (i.e., user experience –  $\alpha = 0.89$ , game mechanics –  $\alpha = 0.89$ , in-game assistance –  $\alpha = 0.90$ , VRISE –  $\alpha = 0.89$ ; see **Table 2**), which indicate very good internal reliability (Nunally and Bernstein, 1994). VRNQ's fit indices are displayed in **Table 2** with their respective thresholds. The  $\chi^2/df$  was 1.61, which indicates good

**TABLE 2 |** Internal reliability and goodness of fit for the VRNQ.

Statistics	Thresholds	Results
Cronbach's $\alpha$	$\geq 0.70$	USER – 0.886 GM – 0.888 GA – 0.895 VR – 0.891
$\chi^2/df$	$\leq 2.00$	1.610
Comparative fit index (CFI)	$\geq 0.90$	0.954
Tuckers Lewis index (TLI)	$\geq 0.90$	0.938
Standardized root mean square residual (SRMR)	$< 0.08$	0.076
Root mean square error of approximation (RMSEA)	$\leq 0.08$	0.071

VRNQ Domains: USER, user experience; GM, game mechanics; GA, in-game assistance; VR VRISE.

variance in the sample (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). Both CFI and TLI were close to 0.95, which suggest a good fit for the VRNQ model (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). Comparably, SPMR and RMSEA values were between 0.06 and 0.08, which also support a good fit (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). The VRNQ's path diagram is displayed in **Figure 1**, where from left to right are depicted the correlations among the factors/domains of the VRNQ, the correlations between each factor/domain and its items, and the error terms for each item. The VRNQ items/questions are efficiently associated with their respective factor/domain, which shows good convergent validity (Cole, 1987). Furthermore, there was not any significant correlation amongst the factors/domains, which indicates good discriminant validity (Cole, 1987).

## Descriptive Statistics of Sessions' Duration and VRNQ Scores

The descriptive statistics for the sessions' durations and the VRNQ scores are displayed in **Table 3**. In session 1, the participants were immersed for 59.65 (8.42) minutes. In session 1, the average time of gamers seems more than the average time of non-gamers (**Table 3**). In session 2, the participants spent 64.72 (6.24) minutes (**Table 3**). In session 3, gamers spent 70.44 (7.78) minutes, while non-gamers spent 65.73 (6.75) minutes (**Table 3**). The average total score of the VRNQ for all software was 126.30 (7.55) (maximum score is 140), where gamers and non-gamers scores did not appear to differ. Similarly, the median scores for each domain were 30–32 out of 35, where again gamers and non-gamers scores did not appear to differ. Importantly, all the VRISE scores (per item) for both gamers and non-gamers were equal to 5 (i.e., mild feeling), or 6 (i.e., very mild feeling), or 7 (absent feeling). The vast majority of scores were equal to 6 (i.e., very mild feeling) or 7 (absent feeling) (see **Figure 2**).

## Minimum and Parsimonious Cut-Off Scores of VRNQ

Cut-off scores were calculated for the VRNQ total score and sub-scores to inspect the suitability of the assessed VR

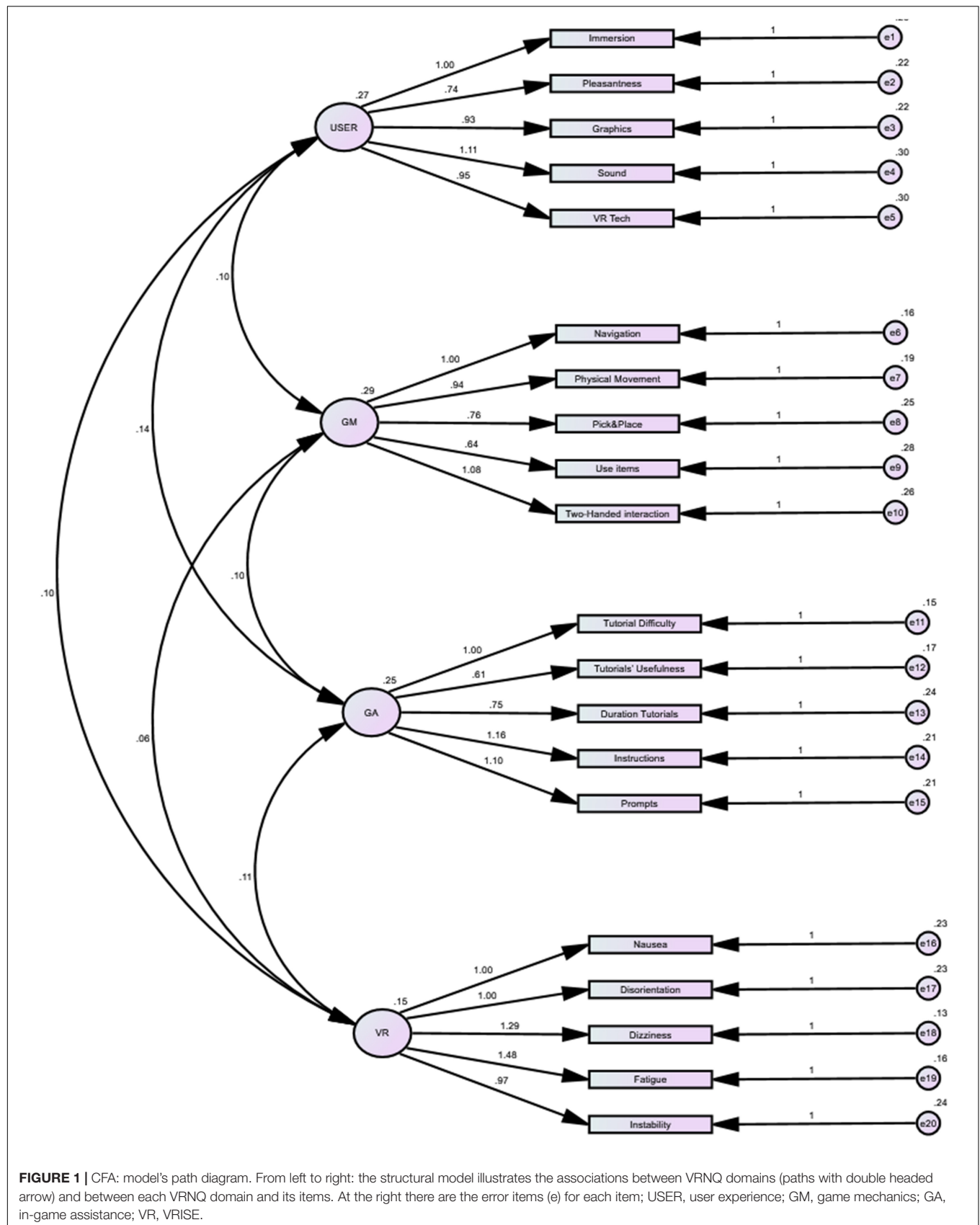
software (see **Table 4**). In the VRNQ, the ordinal 1–3 responses are paired with negative qualities, response 4 is paired with neutral/moderate qualities, and 5–7 responses are paired with positive qualities (see **Supplementary Material**). The minimum cut-offs suggest that if the median of the responses is 25 for every sub-score, and 100 in the total score (i.e., at least a median of 5 for every item), then the VRNQ outcomes indicate that the evaluated VR software is of an adequate quality not to cause any significant VRISE. Furthermore, the parsimonious cut-offs suggest that, if the median of the responses is 30 for every sub-score, and 120 for the total score (i.e., at least a median of 6 for every item) then the utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software. The minimum and parsimonious cut-offs hence appear adequate to guarantee the safety, pleasantness, and appropriateness of the VR software for research and/or clinical purposes.

## Bayesian T-Tests

The Bayesian independent samples *t*-test between gamers and non-gamers indicated that the former spent significantly more time in VR across the total duration for the 3 sessions ( $BF_{10} = 14.99$ ), as well as the duration of the 1st session ( $BF_{10} = 2,532$ ; see **Table 4**) (Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). The difference is much smaller in the total duration than the difference in the 1st session. Thus, the difference between the gamers and non-gamers in the total duration appears to be driven by the substantial difference in the 1st session's duration (see **Table 5**). Conversely, the Bayesian paired samples *t*-test (i.e., differences between the VR games) indicated significant differences in the total score and every sub-score of VRNQ (see **Table 6**) between the VR software. The VR software in the 3rd session was evaluated higher than the VR software in the 1st and 2nd sessions, while the VR software in the 2nd session was rated better than the VR software in the 1st session. There was also an important difference between the duration of the 3rd session (longer) and the duration of the 1st session (shorter;  $BF_{10} = 103,568$ ), while there was not a substantial difference between the duration of the 2nd and 3rd sessions ( $BF_{10} = 2.78$ ), as well as between the duration of 1st and 2nd sessions ( $BF_{10} = 7.05$ ; see **Table 6**) (Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017).

## Bayesian Pearson Correlation Analyses and Regression Analysis

The Bayesian Pearson correlation analyses did not show any significant correlation between age and any of the VRNQ scores, between age and duration of the sessions, between education and any of the VRNQ scores, or between education and duration of the sessions. However, the duration of the session was positively correlated with the total VRNQ score [ $BF_{10} = 81.54$ ;  $r(120) = 0.310$ ,  $p < 0.001$ ]. Furthermore, the VRISE score substantially correlated with the following VRNQ items: immersion, pleasantness, graphics, sound, pick and place, tutorial's difficulty, tutorial's usefulness, tutorial's duration, instructions, and prompts (see **Table 7**). In contrast, VRISE did not significantly correlate with the following VRNQ



**FIGURE 1 |** CFA: model's path diagram. From left to right: the structural model illustrates the associations between VRNQ domains (paths with double headed arrow) and between each VRNQ domain and its items. At the right there are the error items (e) for each item; USER, user experience; GM, game mechanics; GA, in-game assistance; VR, VR/RISE.

items: VR tech, navigation, physical movement, use items, or two-handed interactions (see **Table 7**). Moreover, the Bayesian regression analysis indicated the five best models that predicted the VRNQ's VRISE score (see **Table 8**). The best model includes the following items from the VRNQ: immersion, graphics, sound, instructions, and prompts. All the predictors exceeded the prior inclusion probabilities (see **Figure 3**). The best model showed a  $BF_M = 117.42$ , whereas the second-best model displayed a  $BF_M = 56.40$  (see **Table 8**); hence, the difference between the best model compared to the second-best model was robust (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012; Marsman and Wagenmakers, 2017). Also, the best model has an  $R^2 = 0.324$  (see **Table 8**), which postulates that the model explains the 32.4% of the variance of VRISE score (Rouder and Morey, 2012; Wetzels and Wagenmakers, 2012). Lastly, the *post hoc* statistical power analysis for the best model indicated an observed statistical power of 0.998,  $p < 0.001$ , which postulates a high efficiency, precision, reproducibility, and reliability

of the regression analysis and results (Button et al., 2013; Cohen, 2013).

## DISCUSSION

### The VRNQ as a Research and Clinical Tool

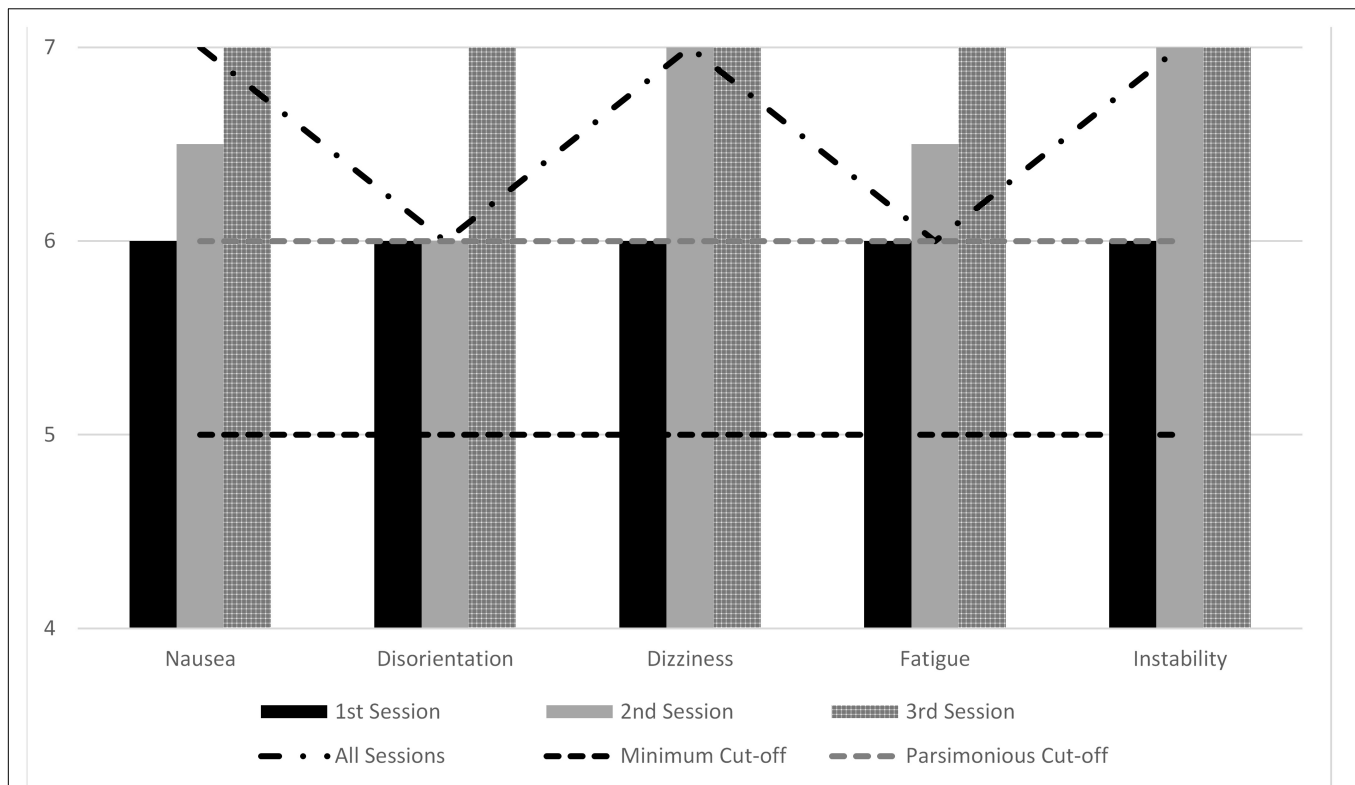
The VRNQ is a short questionnaire (5–10 min administration time) which assesses the quality of VR software in terms of user experience, game mechanics, in-game assistance, and VRISE. The values of the fit indices of CFA (i.e., CFI, TLI, SPMR, and RMSEA) indicated that the VRNQ's structure was a good fit to the data, which postulates good construct validity for the VRNQ (Hu and Bentler, 1999; Jackson et al., 2009; Hopwood and Donnellan, 2010). In addition, the construct validity of the VRNQ was supported by its convergent and discriminant validity (Cole, 1987). VRNQ items were strongly correlated with their grouping factor, which indicates robust convergent validity, while there were substantially poor correlations between the factors, which postulates very good discriminant validity (Cole, 1987). Furthermore, the Cronbach's  $\alpha$  for each VRNQ domain (i.e., user experience –  $\alpha = 0.89$ , game mechanics –  $\alpha = 0.89$ , in-game assistance –  $\alpha = 0.90$ , VRISE –  $\alpha = 0.89$ ; see **Table 2**) suggest very good construct validity (Nunnally and Bernstein, 1994). Henceforth, the VRNQ emerges as a valid and suitable tool to evaluate the quality of the VR research/clinical software as well as the intensity of the adverse VRISE.

Furthermore, minimum and parsimonious cut-off scores were calculated for the VRNQ total score and sub-scores to inspect the suitability of the assessed VR software. The minimum cut-offs indicate the lowest acceptable quality that VR research/clinical software should be, while the parsimonious cut-offs are offered for more robust support of the VR software's suitability, which may be required in experimental and clinical designs with more conservative standards. However, the individual scores from the VRNQ may be modulated by individual differences and preferences unrelated to the quality of the software (Kortum and Peres, 2014). In addition, the VRNQ produces ordinal data; therefore, the median is the appropriate measure for their analysis (Harpe, 2015). Hence, the median VRNQ scores for the whole sample should be used to assess the VR software's quality effectively. Also, the medians of the VRNQ total score and sub-scores allow the generalization of the results and comparison between different VR software (Kortum and Peres, 2014; Harpe, 2015). Researchers, clinicians, and/or research software developers should use the medians of the VRNQ total score and sub-scores to assess whether the implemented VR software exceed the minimum or parsimonious cut-offs. Hence, if the medians of the VRNQ sub-scores and totals score for VR research software meet the minimum cut-offs, then these results support the VR software's suitability. Likewise, if the medians of VRNQ sub-scores and totals score for VR research software meet the parsimonious cut-offs, then these results provide even stronger support for its suitability. However, median scores below these

**TABLE 3 |** Descriptive statistics: duration of VR sessions and VRNQ scores.

	Group	N	Mean (SD)	SE
Total duration	Gamers	18	199.39 (13.63)	3.21
	Non-Gamers	22	186.36 (11.76)	2.51
	Total	40	192.2 (14.09)	2.23
Duration of session 1	Gamers	18	65.61 (7.14)	1.68
	Non-Gamers	22	54.77 (5.91)	1.26
	Total	40	59.65 (8.42)	1.33
Duration of session 2	Gamers	18	63.33 (6.16)	1.45
	Non-Gamers	22	65.86 (6.21)	1.32
	Total	40	64.72 (6.24)	0.99
Duration of session 3	Gamers	18	70.44 (7.78)	1.83
	Non-Gamers	22	65.73 (6.75)	1.44
	Total	40	67.85 (7.52)	0.69
VRNQ total score out of 140 (across 3 sessions)	Gamers	18	127.2 (7.32)	0.99
	Non-Gamers	22	125.6 (7.71)	0.95
	Total	40	126.3 (7.55)	0.69
User's experience (across 3 sessions) out of 35	Gamers	18	31.37 (2.73)	0.34
	Non-Gamers	22	30.91 (2.73)	0.37
	Total	40	31.12 (2.73)	0.25
Game mechanics (across 3 sessions) out of 35	Gamers	18	31.50 (2.68)	0.37
	Non-Gamers	22	31.32 (2.61)	0.32
	Total	40	31.40 (2.63)	0.24
In-game assistance (across 3 sessions) out of 35	Gamers	18	31.70 (2.59)	0.35
	Non-Gamers	22	31.65 (2.52)	0.31
	Total	40	31.68 (2.54)	0.23
VRISE (across 3 sessions) out of 35	Gamers	18	32.67 (2.17)	0.30
	Non-Gamers	22	31.71 (2.56)	0.32
	Total	40	32.14 (2.43)	0.22





**FIGURE 2 |** VRSE intensity in VR sessions as measured by VRNQ. Median scores of VRSE items of VRNQ; VRNQ Minimum Cut-off ( $\geq$ ); VRNQ Parsimonious Cut-off ( $\geq$ ); 1, Extreme intense feeling; 2, Very intense feeling; 3, Intense feeling; 4, Moderate feeling; 5, Mild feeling; 6, Very mild feeling; 7, Absent feeling.

cut-offs suggest that the suitability of the VR software is questionable, but they do not indicate that this VR software is certainly unsuitable.

Also, VRNQ appears as an appropriate tool to measure both VRSE and VR software features compared to other questionnaires. The SSQ is the most implemented questionnaire in VR studies. However, the SSQ only considers the symptoms pertinent to simulator sickness and it does not assess software attributes (Kennedy et al., 1993), while there is a dispute that simulator sickness symptomatology may not be the same as VRSE (Stanney et al., 1997). Alternatively, Virtual reality sickness questionnaire (VRSQ) was recently developed (Kim et al., 2018). The development of VRSQ was based on the SSQ, where the researchers attempted to isolate the items which are pertinent to VRSE (Kim et al., 2018). However, their sample size was relatively small (i.e., 24 participants \* 4 sessions = 96 observations) (Kim et al., 2018). Notably, the factor analyses of Kim et al. (2018) accepted only items pertinent to oculomotor and disorientation components of SSQ, and rejected all the items pertinent to nausea (i.e., 7 items) (Kim et al., 2018), while nausea is the most frequent symptom in VRSE (Stanney et al., 1997; Sharples et al., 2008; Bohil et al., 2011; de França and Soares, 2017; Palmisano et al., 2017). Also, comparable to SSQ, VRSQ does not consider software features. Hence, the VRNQ appears to be the only valid and suitable tool to evaluate both the intensity of predominant VRSE and the quality of VR software features.

The VRNQ allows researchers to report the quality of VR software and/or the intensity of VRSE in their VR studies. However, an in-depth assessment of the numerous software features requires a questionnaire with more than the 20 questions of the VRNQ (Zarour et al., 2015). For an in-depth software analysis, questionnaires with more questions pertinent to the whole spectrum of software features should be preferred (Zarour et al., 2015). Additionally, the VRNQ has solely five items pertinent to VRSE. Hence, it does not offer an exhaustive assessment of VRSE. Studies that aim to investigate VRSE in depth should opt for a tool which contains more items pertinent to VRSE than VRNQ (e.g., SSQ). The VRNQ is a brief questionnaire (5–10 min administration time) including 20 items, which enables researchers, clinicians, and research software developers to evaluate and report the quality of the VR software and the intensity of VRSE for research and clinical purposes.

## Maximum Duration of VR Sessions

The duration of the VR session is a crucial factor in research and/or clinical design. In our sample, the participants discontinued the VR session due to loss of interest, while none discontinued due to VRSE. In the 1st session, gamers spent significantly more time immersed than the non-gamers; a difference which modulated the difference between the two

**TABLE 4 |** VRNQ cut-offs.

Score	Minimum cut-offs	Parsimonious cut-offs
User experience	≥25/35	≥30/35
Game mechanics	≥25/35	≥30/35
In-game assistance	≥25/35	≥30/35
VRISE	≥25/35	≥30/35
VRNQ total score	≥100/140	≥120/140

*The median of each sub-score and totals scores should meet the suggested cut-offs to support that the evaluated VR software has an adequate quality without any significant VRISE. The utilization of the parsimonious cut-offs more robustly supports the suitability of the VR software.*

**TABLE 5 |** Bayesian independent samples *t*-test: gamers against non-gamers.

Variables	Significance	BF10	Error %
Age		0.323	0.006
Education		0.325	0.006
Total duration	*	14.987	7.044e-6
Session 1 duration	***	2531.886	7.491e-8
Session 2 duration		0.595	0.006
Session 3 duration		1.580	0.003
VRNQ total		0.425	0.007
User's experience		0.359	0.006
Game mechanics		0.315	0.006
In-game assistance		0.315	0.006
VRISE		0.745	0.003

*BF<sub>10</sub> = Bayes Factor; \* BF<sub>10</sub> > 10, \*\*\* BF<sub>10</sub> > 100.*

groups in the summed duration across all sessions. However, it is worth noting that there was not a significant difference between the two groups in the time spent in VR for the 2nd and 3rd sessions. The observed difference in the 1st session and the absence of a difference in the later sessions' durations postulates that when users are familiarized with the VR technology, while the influence of their gaming experience on the session's duration becomes insignificant. In support of this, a recent study showed that user gaming experience does not affect the perceived workload of the users in VR (Lum et al., 2018). Hence, the level of familiarization of the participants with the VR technology appears to affect substantially the duration of the VR session.

Nevertheless, in the whole sample, irrespective of participants' gaming experience, the durations of the 2nd and 3rd sessions are sufficiently longer than the duration of the 1st session. The duration of the 3rd session is not significantly longer than the duration of the 2nd session. Furthermore, given that in each session, a different VR software was administered, the VRNQ correspondingly pinpointed significant differences amongst the implemented VR software' quality. All the VRNQ scores for the 3rd session's VR software are greater than the 2nd session's VR software scores. Similarly, all the VRNQ scores for the 2nd session's VR software are greater than the 1st session's VR software scores. Also, the duration of VR session was positively correlated with the total score of VRNQ. Thus, the quality of the VR software as measured by the

VRNQ seems to be significantly associated with the duration of the VR session.

Overall, in every session, the intensity of VRISE was reported as very mild to absent by the vast majority of the sample. However, comparable to the rest of the VRNQ scores, the VRISE score for the 3rd VR session was significantly higher (i.e., milder feeling) than the 2nd and 3rd sessions. Similarly, the VRISE score for the 2nd session's VR software was substantially higher than the 1st session's VR software score. Notably, there was not any difference between gamers and non-gamers in the VRNQ scores across the three sessions. Equally, the age and education of participants did not correlate with any of the VRNQ scores or the duration of sessions. Thus, the age, education, and gaming experience of the participants did not affect the responses in the VRNQ. Therefore, the observed differences in the VRISE scores between the VR sessions support that the quality of the VR software as measured by the VRNQ and the level of familiarization of the participants with the VR technology also affect the intensity of VRISE.

The findings postulate that the implementation of VR software with a maximum duration between 55 and 70 min is substantially feasible. However, long exposures in VR have been found to increase the probability of experiencing VRISE and the intensity of VRISE (Sharples et al., 2008). In our sample, especially in the 3rd session, which was substantially longer than the other sessions, the intensity of VRISE was significantly lower than the rest of the sessions. As discussed above, the substantially lower intensity of VRISE in the 3rd session appears to be a result of increased VR familiarity, and the better quality of the implemented VR software as measured by the VRNQ. Hence, researchers and/or clinicians should consider the quality of their VR software to define the appropriate duration of their VR session. In research and clinical designs where the duration of the VR session is required to be between 55 and 70 min, the researchers and/or clinicians should opt for the parsimonious cut-offs of the VRNQ to ensure adequate quality of their VR software to facilitate longer sessions without significant VRISE. Additionally, an extended introductory tutorial which allows participants to familiarize themselves with the VR technology and mechanics would assist with the implementation of longer (i.e., 55–70 min) VR sessions, where the presence and intensity of VRISE would not be significant.

## The Quality of VR Software and VRISE

The VRISE score substantially correlated with almost every item under the section of user experience and in-game assistance (see **Table 6**). However, the VRISE score did not correlate with VR tech (the item under the user experience's domain) or most of the items under the section of game mechanics. The quality of VR hardware (i.e., the HMD and its controllers) and interactions (i.e., ergonomic or non-ergonomic) with the virtual environment are crucial for the alleviation or evasion of VRISE (Kourtesis et al., 2019). Nevertheless, in this sample, the VR tech item (i.e., the quality of the internal and external VR hardware) was not expected to correlate with the VRISE score, because the HMD and its 6DoF controllers were the

**TABLE 6 |** Bayesian paired samples *t*-tests: differences between the VR software.

Pairs		Significance	BF10	Error %
Session 2 duration	Session 1 duration		7.049	~0.001
Session 3 duration	Session 2 duration		2.783	~3.276e-4
Session 3 duration	Session 1 duration	***	103568.858	NaN
S3 VRNQ total	S2 VRNQ total	***	6.942e + 12	NaN
S3 VRNQ total	S1 VRNQ total	***	3.520e + 20	NaN
S2 VRNQ total	S1 VRNQ total	***	8.500e + 17	NaN
S3 VRISE	S2 VRISE	***	22075.036	NaN
S3 VRISE	S1 VRISE	***	1.322e + 10	NaN
S2 VRISE	S1 VRISE	***	1.160e + 7	NaN
S3 in-game assistance	S2 in-game assistance	***	207216.904	NaN
S2 in-game assistance	S1 in-game assistance	***	1.197e + 7	NaN
S3 in-game assistance	S1 in-game assistance	***	8.028e + 10	NaN
S3 game mechanics	S2 game mechanics	***	274310.417	NaN
S2 game mechanics	S1 game mechanics	***	4.883e + 14	NaN
S3 game mechanics	S1 game mechanics	***	2.876e + 14	NaN
S3 user's experience	S2 user's experience	***	2.873e + 7	NaN
S3 in-game assistance	S1 user's experience	***	2.597e + 7	NaN
S2 user's experience	S1 user's experience	***	1.708e + 6	NaN

BF<sub>10</sub> = Bayes Factor; \* BF<sub>10</sub> > 10, \*\* BF<sub>10</sub> > 30, \*\*\* BF<sub>10</sub> > 100; S1, Session 1; S2, Session 2; S3, Session 3.

same for all 3 VR software versions and sessions. Hence, the variance in the responses to this item was limited. Also, the three VR software games share common game mechanics, especially the same navigation system (i.e., teleportation) and a similar amount of physical mobility. Likewise, apart from some controls (i.e., the button to grab items), the interaction systems of the implemented VR software were very proximal. Therefore, the absence of a correlation between VRISE scores and most of the items in the game mechanics' section was also an expected outcome. Nonetheless, the VRISE score was strongly associated with the level of immersion and enjoyment, the quality

of graphics and sound, the comfort to pick and place 3D objects, and the usefulness of in-game assistance modes (i.e., tutorials, instructions, and prompts).

The items which correlated with the VRISE score were also included in the best models of predicting its value (see **Table 7**). Importantly, the best model includes as predictors of VRISE, the level of immersion, the quality of graphics and sound, and the helpfulness of in-game instructions and prompts (see **Table 7**). The higher scores for prompts and instructions indicate that the user was substantially assisted by the in-game assistance (e.g., an arrow showing the direction that the user should follow) to orientate and guide his or herself from one point of interest to the next in accordance with the scenario of the VR experience. This may be interpreted as ease to orient and interact with the virtual environment, as well as a significant decrease in confusion (Brade et al., 2018). The quality of the in-game assistance methods is essential for the usability and enjoyment that VR software offers (Brade et al., 2018). Equally, the quality of the graphics is predominantly dependent upon rendering which encompasses the in-game quality of the image known as perceptual quality, and the exclusion of redundant visual information known as occlusion culling (Lavoué and Mantiuk, 2015). The improvement of these two factors not only results in improved quality of the graphics but also in improved performance of the software (Brennesholtz, 2018). Furthermore, the spatialized sound of VR software, which assists the user to orient his or herself (Ferrand et al., 2017), deepens the experienced immersion (Riecke et al., 2011), and enriches the geometry of the virtual space without affecting the performance of the software (Kobayashi et al., 2015). Lastly, the level of immersion appears to be negatively correlated with the frequency and intensity of VRISE (Milleville-Pennel and Charron, 2015; Weech et al., 2019). The best model hence aligns with the relevant

**TABLE 7 |** Bayesian Pearson correlations analyses: VRISE score with VRNQ items.

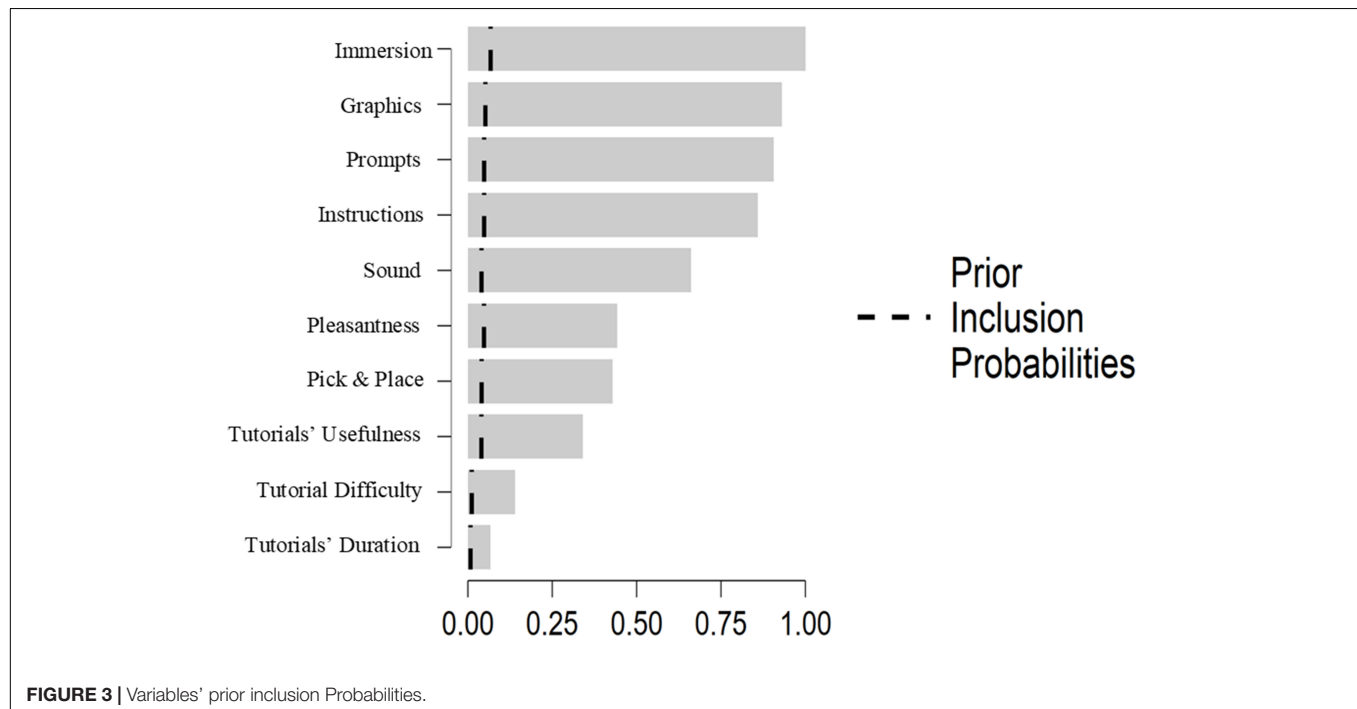
Pairs		Significance	BF10	r
VRISE	Immersion	***	1226.538	0.371
VRISE	Pleasantness	*	20.504	0.273
VRISE	Graphics	***	1629.195	0.377
VRISE	Sound	***	18586.578	0.421
VRISE	VR Tech		5.094	0.228
VRISE	Navigation		4.808	0.226
VRISE	Physical movement		2.229	0.197
VRISE	Pick and place	***	175.087	0.329
VRISE	Use items		0.405	0.109
VRISE	Two-handed interaction		0.506	0.123
VRISE	Tutorial difficulty	***	28252.587	0.428
VRISE	Tutorials usefulness	***	161.949	0.327
VRISE	Tutorials' duration	***	128.539	0.322
VRISE	Instructions	***	952.871	0.366
VRISE	Prompts	***	706510.726	0.476

BF<sub>10</sub> = Bayes Factor; \* BF<sub>10</sub> > 10, \*\* BF<sub>10</sub> > 30, \*\*\* BF<sub>10</sub> > 100;

**TABLE 8 |** Models' comparison: predictors of VRISE score.

Models	P(M)	P(M  data)	BF <sub>M</sub>	BF <sub>10</sub>	R <sup>2</sup>
Prompts + Sound + Graphics + Immersion + Instructions	0.004	0.304	117.42***	1.000	0.324
Prompts + Graphics + Immersion + Instructions + Pleasantness	0.004	0.173	56.47**	0.571	0.317
Prompts + Sound + Graphics + Immersion + Instructions + Pick and Place	0.004	0.161	43.15*	0.443	0.330
Prompts + Sound + Graphics + Immersion + Instructions + Pick and Place + Tutorials Usefulness + Pleasantness	0.021	0.123	6.62	0.072	0.337
Prompts + Graphics + Immersion + Instructions + Pick and Place + Tutorials Usefulness + Pleasantness	0.008	0.077	10.72*	0.121	0.329

P, Probability; M, Model; BF<sub>M</sub>, Model's Bayesian Factor; \* BF<sub>M</sub> > 10, \*\* BF<sub>M</sub> > 30, \*\*\* BF<sub>M</sub> > 100; BF<sub>10</sub> = BF against null model.

**FIGURE 3 |** Variables' prior inclusion Probabilities.

literature and provides further evidence in support of the utility of the VRNQ as a valid and efficient tool to appraise the quality of the VR software and intensity of VRISE.

## Limitations and Future Studies

This study also has some limitations. In this study, construct validity for the VRNQ is provided. However, future work should endeavor to provide convergent validation of the VRNQ with tools that measure VRISE symptomatology (e.g., SSQ) and/or VR software attributes. Moreover, the sample size was relatively small, but it offered an adequate statistical power for the conducted analyses. Also, the VRNQ does not directly quantify linear or angular accelerations, which may induce intense VRISE in a relatively short period of time (McCauley and Sharkey, 1992; LaViola, 2000; Gavvani et al., 2018). However, the VRNQ quantifies the effect(s) of linear and angular accelerations (i.e., VRISE), where VR software with a highly provocative content (e.g., linear and angular accelerations) would fail to meet or exceed the VRNQ cut-offs for the VRISE domain. Furthermore, the study utilized only one type of VR hardware, which did not allow us to inspect the effect of VR HMD's quality on VRISE presence

and intensity. Similarly, our VR software did not allow us to compare different ergonomic interactions or levels of provocative potency pertaining to VRISE. Future studies with a larger sample, various types of VR hardware, and VR software with substantially more diverse features will offer further insights on the impact of software features on VRISE intensity, as well as provide additional support for the VRNQ's structural model. Lastly, neuroimaging (e.g., electroencephalography) and physiological data (e.g., heart rates) may correlate, classify, and predict VRISE symptomatology (Kim et al., 2005; Dennison et al., 2016, 2019). Hence, future studies should consider collecting neuroimaging and/or physiological data that could further elucidate the relationship between VRNQ's VRISE score(s) and brain region activation or cardiovascular responses (e.g., heart rate).

## CONCLUSION

This study showed that the VRNQ is a valid and reliable tool which assesses the quality of VR software and intensity of VRISE. Our findings support the viability of VR sessions with a duration up to 70 min, when the participants are



familiarized with VR tech through an induction session, and the quality of the VR software meets the parsimonious cut-offs of VRNQ. Also, our results offered insights on the software-related predictors of VRSE intensity, such as the level of immersion, the quality of graphics and sound, and the helpfulness of in-game instructions and prompts. Finally, the VRNQ enables researchers to quantitatively assess and report the quality of VR software features and intensity of VRSE, which are vital for the efficacious implementation of immersive VR systems in cognitive neuroscience and neuropsychology. The minimum and parsimonious cut-offs of VRNQ may appraise the suitability of VR software for implementation in research and clinical settings. The VRNQ and the findings of this study contribute to the endeavor of establishing thorough VR research and clinical methods that are crucial to guarantee the viability of implementing immersive VR systems in cognitive neuroscience and neuropsychology.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## REFERENCES

- Arafat, I. M., Ferdous, S. M. S., and Quarles, J. (2018). "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Reutlingen: IEEE), 1–120.
- Arbuckle, J. L. (2014). *Amos (Version 23.0)*. Chicago: IBM SPSS.
- Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 752–762. doi: 10.1038/nrn3122
- Brade, J., Dudczig, M., and Klimant, P. (2018). "Using virtual prototyping technologies to evaluate human-machine-interaction concepts," in *aw&I Conference*, Vol. 3, (Chemnitz).
- Brennesholtz, M. S. (2018). 3–1: invited paper: vr standards and guidelines. *SID Symposium Dig. Tech. Pap.* 49, 1–4. doi: 10.1002/sdtp.12476
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Confidence and precision increase with high statistical power. *Nat. Rev. Neurosci.* 14:585. doi: 10.1038/nrn3475-c4
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge.
- Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *J. Consul. Clin. Psychol.* 55, 584–594. doi: 10.1037//0022-006x.55.4.584
- de França, A. C. P., and Soares, M. M. (2017). "Review of virtual reality technology: an ergonomic approach and current challenges," in *International Conference on Applied Human Factors and Ergonomics*, (Cham: Springer), 52–61. doi: 10.1007/978-3-319-60582-1\_6
- Dennison, M. Jr., D'Zmura, M., Harrison, A., Lee, M., and Raglin, A. (2019). "Improving motion sickness severity classification through multi-modal data fusion," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, Vol. 11006, (Baltimore, MA.), 110060T.
- Dennison, M. S., Wisti, A. Z., and D'Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays* 44, 42–52. doi: 10.1016/j.displa.2016.07.002
- Ferrand, S., Alouges, F., and Aussal, M. (2017). "Binaural spatialization methods for indoor navigation," in *Audio Engineering Society Convention*, Vol. 142, (New York, NY: Audio Engineering Society).
- Gavagni, A. M., Wong, R. H., Howe, P. R., Hodgson, D. M., Walker, F. R., and Nalivaiko, E. (2018). Cybersickness-related changes in brain hemodynamics: a pilot study comparing transcranial Doppler and near-infrared spectroscopy assessments during a virtual ride on a roller coaster. *Physiol. Behav.* 191, 56–64. doi: 10.1016/j.physbeh.2018.04.007
- Harpe, S. E. (2015). How to analyze likert and other rating scale data. *Curr. Pharm. Teach. Learn.* 7, 836–850. doi: 10.1016/j.cptl.2015.08.001
- Hopwood, C. J., and Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personal. Soc. Psychol. Rev.* 14, 332–346. doi: 10.1177/1088868310361240
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struc. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118
- Ibm Corp. (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.
- Jackson, D. L., Gillaspay, J. A. Jr., and Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychol. Methods* 14, 6–23. doi: 10.1037/a0014694
- Jasp Team, (2017). *JASP (Version 0.8.1.2)*.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303\_3
- Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi: 10.1016/j.apergo.2017.12.016
- Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., and Kim, H. T. (2005). Characteristic changes in the physiological components of cybersickness. *Psychophysiology* 42, 616–625.
- Kobayashi, M., Ueno, K., and Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence* 24, 163–174. doi: 10.1162/pres\_a\_00226
- Kortum, P., and Peres, S. C. (2014). The relationship between system effectiveness and subjective usability scores using the system usability scale. *Int. J. Hum. Comput. Int.* 30, 575–584. doi: 10.1080/10447318.2014.904177
- Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019). Technological competence is a precondition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Front. Hum. Neurosci.* 13:342.
- LaViola, J. J. Jr. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bull.* 32, 47–56. doi: 10.1145/333329.333344

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Philosophy, Psychology and Language Sciences Research Ethics Committee of the University of Edinburgh. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

PK had the initial idea and contributed to every aspect of this study. SC, LD, and SM contributed to the methodological aspects and the discussion of the results. The VRNQ may be downloaded from **Supplementary Material**.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2019.00417/full#supplementary-material>

- Lavoué, G., and Mantiuk, R. (2015). "Quality assessment in computer graphics," in *Visual Signal Quality Assessment*, (Cham: Springer), 243–286. doi: 10.1007/978-3-319-10368-6\_9
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *J. Am. Stat. Assoc.* 103, 410–423.
- Lum, H. C., Greatbatch, R., Waldfogle, G., and Benedict, J. (2018). "How Immersion, Presence, Emotion, & Workload Differ in Virtual Reality and Traditional Game Mediums," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 62, (Los Angeles, CA: SAGE Publications), 1474–1478. doi: 10.1177/1541931218621334
- Marsman, M., and Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614
- McCauley, M. E., and Sharkey, T. J. (1992). Cybersickness: perception of self-motion in virtual environments. *Presence* 1, 311–318. doi: 10.1162/pres.1992.1.3.311
- Milleville-Pennel, I., and Charron, C. (2015). Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accid. Anal. Prev.* 74, 192–202. doi: 10.1016/j.aap.2014.10.021
- Mittelstaedt, J., Wacker, J., and Stelling, D. (2018). Effects of display type and motion control on cybersickness in a virtual bike simulator. *Displays* 51, 43–50. doi: 10.1016/j.displa.2018.01.002
- Mittelstaedt, J. M., Wacker, J., and Stelling, D. (2019). VR aftereffect and the relation of cybersickness and cognitive performance. *Vir. Real.* 23, 143–154. doi: 10.1007/s10055-018-0370-3
- Nalivaiko, E., Davis, S. L., Blackmore, K. L., Vakulin, A., and Nesbitt, K. V. (2015). Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. *Physiol. Behav.* 151, 583–590. doi: 10.1016/j.physbeh.2015.08.043
- Nesbitt, K., Davis, S., Blackmore, K., and Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48, 1–8. doi: 10.1016/j.displa.2017.01.002
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.
- Palmisano, S., Mursic, R., and Kim, J. (2017). Vection and cybersickness generated by head-and-display motion in the oculus rift. *Displays* 46, 1–8. doi: 10.1016/j.displa.2016.11.001
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* 9:660. doi: 10.3389/fnhum.2015.00660
- Parsons, T. D., McMahan, T., and Kane, R. (2018). Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. *Clin. Neuropsychol.* 32, 16–41. doi: 10.1080/13854046.2017.1337932
- Riecke, B. E., Feuereissen, D., Rieser, J. J., and McNamara, T. P. (2011). "Spatialized sound enhances biomechanically-induced self-motion illusion (vection)," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, NY: ACM.), 2799–2802.
- Rouder, J. N., and Morey, R. D. (2012). Default bayes factors for model selection in regression. *Multivar. Behav. Res.* 47, 877–903. doi: 10.1080/00273171.2012.734737
- Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005
- Soper, D. S. (2019a). *A-priori Sample Size Calculator for Structural Equation Models*. Available at <http://www.danielsoper.com/statcalc> (accessed July 10, 2019).
- Soper, D. S. (2019b). *Post-hoc Statistical Power Calculator for Multiple Regression*. Available at <http://www.danielsoper.com/statcalc> (accessed July 10, 2019).
- Stanney, K. M., Kennedy, R. S., and Drexler, J. M. (1997). "Cybersickness is not simulator sickness," In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, (Los Angeles, CA: SAGE Publications). 41, 1138–1142.
- Toschi, N., Kim, J., Sclocco, R., Duggento, A., Barbieri, R., Kuo, B., et al. (2017). Motion sickness increases functional connectivity between visual motion and nausea-associated brain regions. *Auton. Neurosci.* 202, 108–113. doi: 10.1016/j.autneu.2016.10.003
- Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* 10:158. doi: 10.3389/fpsyg.2019.00158
- Wetzels, R., and Wagenmakers, E. J. (2012). A default bayesian hypothesis test for correlations and partial correlations. *Psychon. Bull. Rev.* 19, 1057–1064. doi: 10.3758/s13423-012-0295-x
- Zarour, M., Abran, A., Desharnais, J. M., and Alarifi, A. (2015). An investigation into the best practices for the successful design and implementation of lightweight software process assessment methods: a systematic literature review. *J. Syst. Softw.* 101, 180–192. doi: 10.1016/j.jss.2014.11.041

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

Copyright © 2019 Kourtesis, Collina, Doumas and MacPherson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Virtual Reality Analgesia With Interactive Eye Tracking During Brief Thermal Pain Stimuli: A Randomized Controlled Trial (Crossover Design)

Najood A. Al-Ghamdi<sup>1</sup>, Walter J. Meyer III<sup>2,3</sup>, Barbara Atzori<sup>4</sup>, Wadee Alhalabi<sup>5,6,7</sup>, Clayton C. Seibel<sup>8</sup>, David Ullman<sup>8</sup> and Hunter G. Hoffman<sup>8,9\*</sup>

<sup>1</sup> Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, <sup>2</sup> Shriners Hospitals for Children, Galveston, TX, United States, <sup>3</sup> Department of Psychiatry, The University of Texas Medical Branch at Galveston, Galveston, TX, United States, <sup>4</sup> Department of Health Sciences, School of Psychology, University of Florence, Florence, Italy, <sup>5</sup> Department of Computer Science, College of Engineering, Effat University, Jeddah, Saudi Arabia, <sup>6</sup> The Virtual Reality Research Group, King Abdulaziz University, Jeddah, Saudi Arabia, <sup>7</sup> Department of Computer Science, Dar Al-Hekma University, Jeddah, Saudi Arabia, <sup>8</sup> Virtual Reality Research Center, Human Photonics Lab, University of Washington, Seattle, WA, United States, <sup>9</sup> Department of Mechanical Engineering, College of Engineering, University of Washington, Seattle, WA, United States

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Luana Colloca,  
University of Maryland, Baltimore,  
United States  
Karel Allegaert,  
University Hospitals Leuven, Belgium  
Paula Goolkasian,  
University of North Carolina  
at Charlotte, United States  
Elisa Pedrolì,  
Italian Institute for Auxology (IRCCS),  
Italy

### \*Correspondence:

Hunter G. Hoffman  
hoontair@gmail.com

### Specialty section:

This article was submitted to  
Sensory Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 24 April 2019

**Accepted:** 19 December 2019

**Published:** 23 January 2020

### Citation:

Al-Ghamdi NA, Meyer WJ III, Atzori B, Alhalabi W, Seibel CC, Ullman D and Hoffman HG (2020) Virtual Reality Analgesia With Interactive Eye Tracking During Brief Thermal Pain Stimuli: A Randomized Controlled Trial (Crossover Design). *Front. Hum. Neurosci.* 13:467. doi: 10.3389/fnhum.2019.00467

In light of growing concerns about opioid analgesics, developing new non-pharmacologic pain control techniques has become a high priority. Adjunctive virtual reality can help reduce acute pain during painful medical procedures. However, for some especially painful medical procedures such as burn wound cleaning, clinical researchers recommend that more distracting versions of virtual reality are needed, to further amplify the potency of virtual reality analgesia. The current study with healthy volunteers explores for the first time whether interacting with virtual objects in Virtual Reality (VR) via “hands free” eye-tracking technology integrated into the VR helmet makes VR more effective/powerful than non-interactive/passive VR (no eye-tracking) for reducing pain during brief thermal pain stimuli.

**Method:** Forty eight healthy volunteers participated in the main study. Using a within-subject design, each participant received one brief thermal pain stimulus during interactive eye tracked virtual reality, and each participant received another thermal pain stimulus during non-interactive VR (treatment order randomized). After each pain stimulus, participants provided subjective 0–10 ratings of cognitive, sensory and affective components of pain, and rated the amount of fun they had during the pain stimulus.

**Results:** As predicted, interactive eye tracking increased the analgesic effectiveness of immersive virtual reality. Compared to the passive non-interactive VR condition, during the interactive eye tracked VR condition, participants reported significant reductions in worst pain ( $p < 0.001$ ) and pain unpleasantness ( $p < 0.001$ ). Participants reported a significantly stronger illusion of presence ( $p < 0.001$ ), and significantly more fun in VR ( $p < 0.001$ ) during the interactive condition compared to during passive VR. In summary, as predicted by our primary hypothesis, in the current laboratory acute pain analog study with healthy volunteers, increasing the immersiveness of the VR system

via interactive eye tracking significantly increased how effectively VR reduced worst pain during a brief thermal pain stimulus. Although attention was not directly measured, the pattern of pain ratings, presence ratings, and fun ratings are consistent with an attentional mechanism for how VR reduces pain. Whether the current results generalize to clinical patient populations is another important topic for future research. Additional research and development is recommended.

**Keywords:** virtual reality, analgesia, pain, distraction, non-pharmacologic analgesic techniques, opioid analgesia

## INTRODUCTION

Excessive pain during medical procedures is a frequent problem, worldwide, and recovering from a severe burn is often unusually painful. As part of the natural healing process, severe burns exfoliate dead skin cells. To prevent infection and speed up healing, wound care nurses remove the bandages, and wipe/scrub the burn wound with a wet washcloth to remove/clean off the thin layer of sloughed off dead skin cells, and other debris.

Despite giving patients powerful analgesia pain medications shortly before wound care, patients typically remain conscious during burn wound care, pain during non-surgical wound debridement of severe burn wounds is often severe to excruciating, and wound care procedures are repeated frequently. Children with large burns typically have their burn wounds cleaned several days per week, during several weeks of hospitalization (Hoffman et al., 2019).

Psychological factors such as fear, anxiety and depression can increase or amplify how much pain patients subjectively experience during painful medical procedures (Hemington et al., 2017), making pain management even more challenging. What people are thinking about, and where patients direct their attention during medical procedures (Heathcote et al., 2017), expectations of pain, and memories of previous painful procedures can increase pain intensity (Melzack and Wall, 1965; Noel et al., 2012, 2015a,b; Fields, 2018; Fischer et al., 2018).

Opioid analgesics are widely regarded as effective and essential tools for acute pain management such as burn wound care (Malchow and Black, 2008; McIntyre et al., 2016; Ballantyne, 2018) but opioid side effects such as nausea, constipation, urinary retention, drowsiness, and lack of appetite (Mendez-Romero et al., 2018) limit opioid dose levels (Cherny et al., 2001). In addition, there are growing concerns about over-prescription of opioids in the United States and Europe (Krane, 2019; Yaster et al., 2019). There are also growing concerns about the opposite problem of limited availability of opioids in under-developed countries (Vijayan, 2011), recent shortages of medical opioid analgesics in the United States, and possible large reductions in availability in the future, in light of stricter laws and large lawsuits against pharmaceutical companies that sell opioid analgesics. In light of growing problems with opioid analgesics, developing effective new non-pharmacologic pain control treatments has become a national and international priority (Keefe et al., 2012, 2018).

Psychological pain control techniques can help reduce reliance on opioids, and may help compensate for undermedication.

There is growing evidence that adjunctive immersive virtual reality distraction can help reduce the suffering of patients during medical procedures with little or no side effects from the VR (Hoffman, 1998; Hoffman et al., 2000, 2011; Garrett et al., 2014; Atzori et al., 2018a,b; Indovina et al., 2018). Virtual reality goggles with eye tracking technology embedded into the goggles, have recently become commercially available, and could potentially help make VR more distracting. However, to date, one important gap in the scientific VR analgesia literature is that there have been no PubMed indexed studies using eye tracked Virtual Reality to treat pain.

In previous clinical studies, virtual reality has typically been used in addition to traditional pain medications to help reduce the pain experienced by patients during painful severe burn wound cleaning sessions (Hoffman et al., 2000, 2004a, 2019; Maani et al., 2011a,b; Kipping et al., 2012; Faber et al., 2013; Jeffs et al., 2014; Khadra et al., 2018; McSherry et al., 2018)<sup>1</sup>. In a military study, while in virtual reality, soldiers with combat-related burn injuries spent less time thinking about their pain, patients reported reductions in worst pain intensity and reductions in the emotional component of pain (pain unpleasantness) during VR, and patients reported having more fun when they went into virtual reality during medical procedures, compared to standard of care pain medications alone (Maani et al., 2008, 2011a,b).

The essence of immersive virtual reality analgesia is the patient's illusion of going to a different place, the subjective experience of "feeling present" in the computer generated world, as if the virtual world is a place the patient is visiting (Slater et al., 1994; Slater and Wilbur, 1997). Researchers argue that the illusion of "being there" in virtual reality is unusually attention grabbing (Hoffman et al., 2000, 2001, 2003b, 2004a,b,c, 2007; Law et al., 2011). For example, people perform more poorly on a divided attention task while in virtual reality (Hoffman et al., 2003b). The perception of pain requires attentional resources. Researchers (e.g., Hoffman, 1998; Hoffman et al., 2000; Birnie et al., 2017), propose that VR uses up so much attention, that the patient's brain has less attention available to process incoming nociceptive signals traveling from the burn wound to the brain while the wound is being cleaned.

One clinical research study recently explored the use of virtual reality to help reduce the pain of pediatric patients with large severe burn injuries during burn wound cleaning sessions in the

<sup>1</sup><https://www.nationalgeographic.com/magazine/2020/01/scientists-are-unraveling-the-mysteries-of-pain-feature/>



intensive care unit of a regional pediatric burn center (Hoffman et al., 2019). Although patients reported large and significant reductions in pain during burn wound care, the researchers recommended that stronger versions of virtual reality need to be developed, in order to better distract patients experiencing such high levels of pain during burn wound cleaning.

Researchers have described and tested design guidelines for how to make VR more effective. Several analog laboratory thermal pain studies have shown that more immersive VR systems designed to elicit a stronger illusion of feeling present in the virtual world are more effective at reducing pain (Hoffman et al., 2004c, 2006, 2014; Dahlquist et al., 2007; Wender et al., 2009; Law et al., 2011; Zeroth et al., 2019). Interacting with the virtual world increases the immersiveness of the VR system, potentially increasing the amount of attention drawn into the virtual world (e.g., Wender et al., 2009). The results of these laboratory studies have helped guide the design of effective VR analgesia systems, e.g., fMRI magnet-friendly wide field of view VR helmets, e.g., Hoffman et al. (2004b, 2007) and the development robot-like arm VR goggle holders for severe burn patients (Maani et al., 2008, 2011a,b; Hoffman et al., 2019).

Unfortunately, increasing the immersiveness of VR systems used in the ICU for children with large severe burn injuries is both highly recommended, but also technically challenging. Burns on their heads and face often preclude burn patients from wearing a traditional VR helmet, so head tracking is not possible. Children with large severe burn injuries also often have severe burns on their hands/fingers, making it difficult for them to use mouse tracking to interact with the virtual world during burn wound care.

In the current pilot laboratory study, we used a new VR helmet that allows participants to use their eye movements as a “hands free” input device to interact with the virtual world. We predicted that adding interactivity via an eye tracking system embedded into the VR goggles would increase the immersiveness of the VR system, and would increase the participants illusion of “being there” in the virtual world, making VR more attention grabbing and more effective at reducing the acute pain of healthy volunteers during brief thermal pain stimuli. Several large computer companies are developing and marketing new virtual reality and augmented reality eye tracking technologies. <https://www.forbes.com/sites/solrogers/2019/02/05/seven-reasons-why-eye-tracking-will-fundamentally-change-vr/#16bfb2c83459>. For example, Apple Computers recently purchased SMI. Note that SMI is the company that made the eye tracking technology used in the current study.

The current analog laboratory pain study with healthy volunteer participants is the first controlled study in the PubMed literature to explore whether interactive eye tracking can enhance the analgesic effectiveness of virtual reality distraction. The SMI eye tracked HTC VIVE VR helmet starts with a standard HTC VIVE helmet. But in addition, each eyepiece of the goggles is trimmed with a small ring of eye tracking technology<sup>2</sup>. Six infrared lights are positioned in a circle around each eye. In addition to the low energy infrared lights, miniature infrared

cameras mounted onto the same ring record the pattern of red lights with an infrared camera (see **Figure 3**). These miniature cameras can make real time digital video streams of the six small red dots of infrared light reflected off of the outer surface of the patient’s eyes (the cornea). As the participants look at different objects in the computer generated world, the pattern of infrared dots changes shape. The VR computer can tell from the pattern of dots, where the patient is looking (search [www.youtube.com](http://www.youtube.com) for “SMI eyetracked virtual reality” for related informational videos). Because the eye tracking system only uses light in the narrow bandwidth of infra-red, the video camera is able to ignore confusing reflection noise from the visible spectrum and infrared thus improves eye tracking accuracy.

The information from the miniature infrared cameras mounted in the VR helmet is transmitted to the VR software program in the VR computer. In the current study, participants interact with the virtual world by aiming virtual snowballs at virtual objects in the 3D virtual canyon. Just as a computer mouse input device can be used to move a computer cursor around a computer screen, using eye tracking technology embedded into the VR goggles, in the current study, the participant in virtual reality can aim snowballs at objects in virtual reality by simply looking at the virtual objects. Essentially, the “cursor” or reticle crosshair, follows the patient’s eye fixations. So if the patient looks at a Snowman in virtual reality, the virtual snowballs hit the Snowman, and the virtual snowman reacts (with special animated effects) when hit by a snowball.

The current laboratory thermal pain study with healthy volunteers explores for the first time, whether interactive eye tracking can enhance the analgesic effectiveness of virtual reality distraction.

## MATERIALS AND METHODS

### Subjects

Forty-eight female college student volunteers from Effat University (age range 18–30 years old, mean = 21.77, *SD* = 2.07) participated in the main study, and an additional 24 students (from the identical context as the main study participants) were randomized to participate in a small pilot side study, to test our pain paradigm assumption that pain ratings were stable over repeated stimulations for people who received two test pain stimuli with No VR. Effat University is an all female institution of higher education for women in Jeddah Saudi Arabia. All data was collected by female research assistants and all participants were female, an understudied gender. This research was conducted in accordance with the Declaration of the World Medical Association<sup>3</sup>. All subjects gave written informed consent in accordance with the Declaration of Helsinki. Both written and verbal informed consent were obtained using a protocol approved by the Effat University’s Human Subjects Review Committee.

<sup>2</sup>[https://en.wikipedia.org/wiki/SensoMotoric\\_Instruments](https://en.wikipedia.org/wiki/SensoMotoric_Instruments)

<sup>3</sup>[www.wma.net](http://www.wma.net)

## Within-Subjects Design

Each of the 48 participants who received VR rated their pain during “eye tracked VR” during one thermal pain stimulus (e.g., Test 1), and rated their pain again during a second thermal stimulus during “passive VR” (e.g., Test 2). Treatment order of passive VR vs. interactive eye tracked VR was randomized using random number sequences from [www.random.org](http://www.random.org). Some people received “eye tracked VR” first and “passive VR” second, and some people received “passive VR” first and “eye tracked VR” second. Each individual participant’s pain during passive VR was compared to that same participant’s pain during interactive eye tracked VR.

## Measures and Procedures

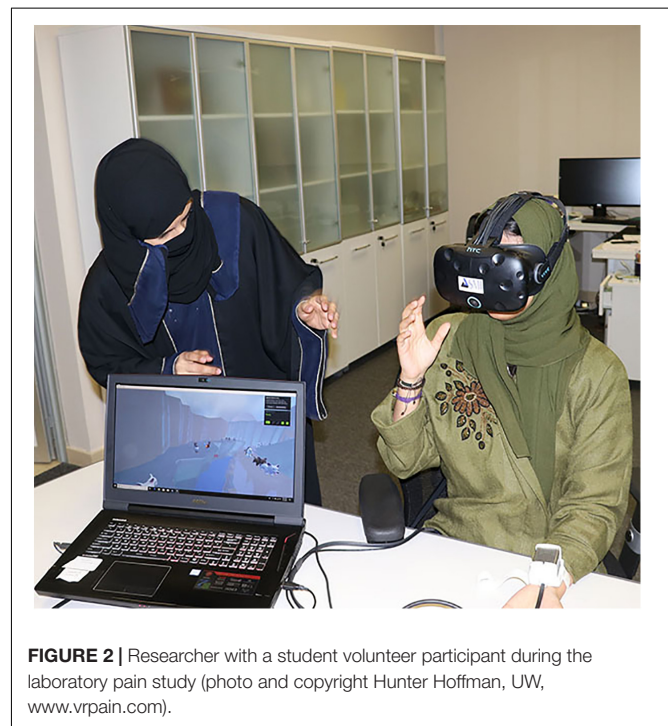
### Experimental Thermal Pain Model

Controlled thermal pain stimulation was applied using a commercially available computerized Medoc thermal pain stimulator<sup>4</sup> (Medoc Q-Sense Ramp and Hold program). During the first phase of the study, each participant selected the temperature they would use in this study. The stimulus temperature (range = 44 – 48.5°C in the present study) of each 10 s heat stimulus temperature was individually determined for each subject using the psychophysical method of ascending levels (Hoffman et al., 2004b, 2007). A 10-s heat stimulus (always 44°C for the first stimulus) was delivered via a thermode attached to the participant’s forearm (by a female researcher), and the subject was asked to rate their pain during the stimulus using a 0–10 graphic rating scale. With the subject’s permission, the temperature for the next stimulus was then increased by 1°C (or less, if the participant was approaching their maximum) and participants again rated their pain. This sequence was repeated until the subject reported a stimulus that was “painful but tolerable” for the brief stimulus duration, and that the subject was willing to receive for two additional 10-s thermal pain stimuli. This final stimulus temperature that the participant selected for the baseline pain condition (10-s thermal stimulus with no distraction) also served as the pain stimulus temperature used during the subsequent VR interventions (10 s of thermal pain during passive VR distraction, and 10 s of thermal pain during interactive eye-tracked VR, VR treatment order randomized). Allowing participants to select the temperature they would use in this experiment was popular with the participants.

The VR system was carried out using a gaming laptop: MSI GeForce GTX 1080 8 GB, Intel Core i7 7th (2.80 GHz), 16 GB RAM, Windows 10 operating system connected to an SMI HTC VIVE VR helmet with FOV 110° from HTC, with 1080 × 1200 pixels per eye resolution and a refresh rate of 90 Hz. The head mounted display VR helmet, integrated with SMI eye-tracking 250 Hz, works with the SDK C++ for various VR engines like Unity. A new VR world, SnowCanyon<sup>5,6</sup> was integrated with the eye tracking hardware, enabling participants to use the eye-tracker to select a virtual object by simply looking at the



**FIGURE 1** | A still shot from SnowCanyon (image by bigenvironments.com, copyright Hunter Hoffman, [www.vrpain.com](http://www.vrpain.com)).



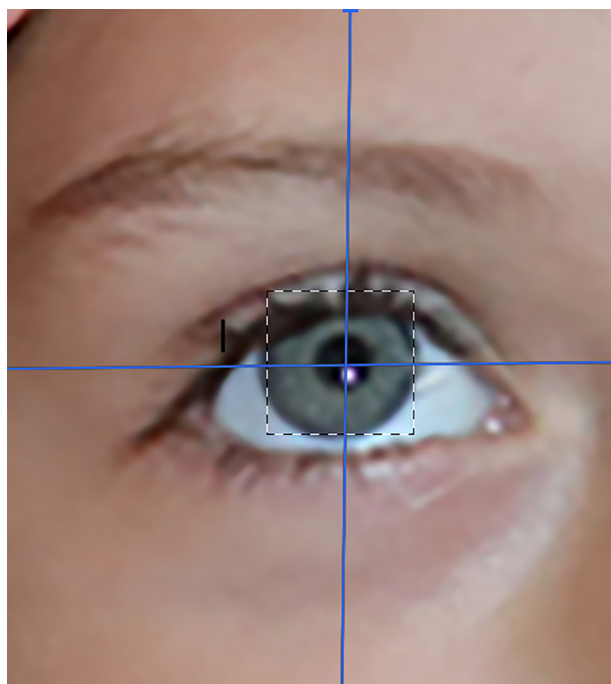
**FIGURE 2** | Researcher with a student volunteer participant during the laboratory pain study (photo and copyright Hunter Hoffman, UW, [www.vrpain.com](http://www.vrpain.com)).

virtual object target (e.g., a virtual snowman) in the VR goggles. The SnowCanyon virtual environment<sup>5</sup> presents a virtual arctic canyon to the user, complete with flowing river below, blue sky above, and terraced canyon walls to the sides containing virtual penguins, igloos, and snowmen (see **Figure 1**). Subjects in both treatment conditions wore an HTC VIVE VR helmet head-mounted display with an integrated SMI eye tracking system (see **Figures 2, 3**). For all participants in the current study, sound was muted and the helmets were immobilized (no head tracking), as an analog to the robot-like articulated arm goggle held eye-tracked VR goggles that will eventually be used with actual burn patients in future clinical studies. The eye-tracking interactions in the VR game were designed to be easy for the

<sup>4</sup>[www.medoc-web.com](http://www.medoc-web.com)

<sup>5</sup>[bigenvironments.com](http://bigenvironments.com)

<sup>6</sup>[www.vrpain.com](http://www.vrpain.com)



**FIGURE 3 |** An artist's rendition of an eye-tracked eye. Photo, image and copyright Hunter Hoffman, www.vrpain.com.

participants to make sense of how the players interact with game objects in that game/environment.

In both VR treatment conditions, each subject “glided” through the virtual world along a pre-determined path. During the interactive VR treatment condition, participants in SnowCanyon could target and shoot virtual objects in the virtual world by moving their eyes (i.e., simply looking at an object to aim snowballs at that object). The Passive VR treatment condition consisted of the identical SnowCanyon software and VR system, but with no eye-tracking and no interactivity/no snowballs. Subjects passively glided through the snowy 3D canyon in the 110 degree field of view HTC VIVE VR goggles.

## Measures

After each pain stimulus, subjects received the following instructions prior to answering six separate subjective questions: “Please indicate how you felt during the most recent 10-s pain stimulus by making a mark anywhere on the line. Your response does not have to be a whole number.”

After each pain stimulus, participants rated their pain using Graphic Rating Scales (GRS). Such pain rating scales have been shown to be valid through their strong associations with other measures of pain intensity, as well as through their ability to detect treatment effects (Jensen and Karoly, 2001; Jensen, 2003; Hoffman et al., 2004c). The GRS is a 10-unit horizontal line labeled with number and word descriptors. In the current study, the tool was used to assess three reports of the pain experience (“worst pain,” “pain unpleasantness,” and “time spent thinking about pain”) that correspond to three separable components of

the pain experience; sensory pain, affective pain, and cognitive pain, respectively.

Descriptor labels were associated with each mark to help the participant rate their pain magnitude in each domain. For pain intensity, the GRS descriptors were *no pain at all*, *mild pain*, *moderate pain*, *severe pain*, and *excruciating pain*. For pain unpleasantness, the GRS descriptors were *not unpleasant at all*, *mildly unpleasant*, *moderately unpleasant*, *severely unpleasant*, and *excruciatingly unpleasant*. For time spent thinking about pain, the GRS descriptors were *none of the time*, *some of the time*, *half of the time*, *most of the time*, *all of the time*.

The Graphic Rating Scale has previously been used to assess pain intensity in children eight and older and has been documented to be the preferred report method for young children (Tesler et al., 1991). The GRS is more sensitive than simple descriptive pain scales and participants can easily answer these pain ratings despite having no previous experience. Visual Analog Scales have been validated for use in children aged 7 and higher (Bringuier et al., 2009).

A single rating “to what extent did you feel like you ‘went into’ the virtual world,” adapted from Slater et al. (1994) was also used in the present study to assess user presence in the virtual world. Descriptor labels were *I did not feel like I went inside at all*, *mild sense of going inside*, *moderate sense of going inside*, *strong sense of going inside*, *I went completely inside the computer generated world*. Hendrix and Barfield (1995) showed the reliability of a similar VR presence rating. The measure’s ability to detect treatment effects (Hoffman et al., 2003a, 2004c) is preliminary evidence of our VR presence measure’s validity. Participants also rated how real the objects seemed in virtual reality, descriptors were *completely fake*, *somewhat real*, *moderately real*, *very real*, *indistinguishable from a real object*. Participants rated nausea as a result of VR, using a graphic rating scale with descriptors *no nausea at all*, *mild nausea*, *moderate nausea*, *severe nausea*, *vomit*. As a surrogate measure of positive affect, participants rated how much fun they had during the painful stimulus (Hoffman et al., 2004a). The verbal descriptors associated with the fun rating were *no fun at all* (0), *mildly fun* (1–4), *moderately fun* (5–6), *pretty fun* (7–9), and *extremely fun* (10). Previous studies indicate that these secondary measures are sensitive to manipulations of the immersiveness of the VR system (Hoffman et al., 2004a,c; Wender et al., 2009).

The specific questions used in the current study were designed to assess the cognitive component of pain (amount of time spent thinking about pain), the affective component of pain (pain unpleasantness), and the sensory component of pain (worst pain). Nausea/Dizziness was assessed in an effort to identify the incidence of this component of simulator sickness sometimes associated with VR use.

Participants individually identified and pre-approved a baseline thermal pain stimulus temperature they found “painful but tolerable for 10 s, that they were willing to tolerate for two more 10 s thermal pain stimuli at that same temperature.” The mean thermal stimulation temperature was 47.18°C (SD = 0.93).

Subjective pain ratings were obtained from each healthy volunteer participant after brief thermal pain stimuli at three time



points, with an interstimulus interval of approximately 4 min, using the same temperature each time: (a) baseline, (b) Test 1, and (c) Test 2.

## RESULTS

As mentioned earlier, treatment order was randomized, using random number sequences from random.org. Twenty-four participants received passive VR first and interactive eye tracked VR second (treatment order 1) and the other 24 participants received interactive eye tracked VR first and passive VR second (treatment order 2).

A two way Mixed ANOVA was conducted to test for undesired treatment order effects. The repeated measures ANOVA factor was (passive VR vs. interactive eye tracked VR), and the between groups factor was treatment order 1 vs. treatment order 2. Mixed ANOVAs showed there was no significant interaction between treatment order and worst pain ratings (i.e., no significant treatment order effects for Worst pain),  $F(1,46) = 1.81$ ,  $p = 0.19$  ns,  $MS = 2.04$ ,  $\eta_p^2 = 0.04$ . There was no significant interaction between treatment order and pain unpleasantness,  $F(1,46) < 1$ ,  $p = 0.69$  ns,  $MS = 0.26$ ,  $\eta_p^2 = 0.004$ . There was no significant interaction between treatment order and participants ratings of Time spent thinking about pain during the thermal stimulus,  $F(1,46) = 1.00$ ,  $p = 0.32$  ns,  $MS = 0.51$ ,  $\eta_p^2 = 0.02$ . And, there was no significant interaction between treatment order and participants ratings of Fun during the thermal stimulus,  $F(1,46) = 1.45$ ,  $p = 0.23$  ns,  $MS = 1.76$ ,  $\eta_p^2 = 0.03$ .

Because no significant order effects were found, the results were collapsed across treatment order for all of the analyses below.

### Statistical Analyses Collapsed Across Treatment Order

One way repeated measure ANOVAs were performed to test if there were significant main effects of No VR vs. passive VR vs. interactive eye-tracked VR. Paired  $t$ -test analyses were performed for the primary outcome measure (worst pain), as well as for the secondary pain measures (i.e., pain unpleasantness, and time spent thinking about pain). For these three pain ratings, alpha was conservatively set at  $0.05/3 = 0.017$ . Any  $p$ -value less than 0.017 was considered significant (Bonferroni corrected for familywise error). Additional paired  $t$ -test analyses were performed for the other secondary graphic rating scale measures (fun, nausea, presence, real) with  $\alpha = 0.05$ .

#### WORST PAIN

A one way repeated measure ANOVA indicated a significant main effect of No VR vs. passive VR vs. interactive eye-tracked VR for worst pain,  $F(2,94) = 61.41$ ,  $p < 0.001$ ,  $MS = 92.15$ , partial  $\eta^2 = 0.57$ . As predicted, compared to No VR, worst pain ratings were significantly lower during passive VR. Compared to No VR, worst pain was significantly lower during interactive eye tracked VR. And compared to passive VR, worst pain was significantly lower during interactive eye tracked VR.

#### Worst pain

No VR	Passive VR	Eye tracked VR	Paired $t$ -tests
6.35 (1.14)	5.00 (1.91)		$t(47) = 5.66$ , $p < 0.001$ , $SD = 1.66$
6.35 (1.14)		3.58 (2.03)	$t(47) = 9.64$ , $p < 0.001$ , $SD = 1.99$
	5.00 (1.91)	3.58 (2.03)	$t(47) = 6.49$ , $p < 0.001$ , $SD = 1.51$

#### Pain Unpleasantness

A one way repeated measure ANOVA showed a significant main effect of No VR vs. passive VR vs. interactive eye tracked VR for pain unpleasantness,  $F(2,94) = 44.33$ ,  $p < 0.001$ ,  $MS = 68.92$ ,  $\eta_p^2 = 0.49$ .

*Post hoc* paired comparisons (paired  $t$ -tests) are shown below. As predicted, compared to No VR, pain unpleasantness ratings were significantly lower during passive VR. Compared to No VR, pain unpleasantness was significantly lower during interactive eye tracked VR. And compared to passive VR, pain unpleasantness was significantly lower during interactive eye tracked VR.

#### PAIN UNPLEASANTNESS

##### Pain unpleasantness

No VR	Passive VR	Eye tracked VR	Paired $t$ -tests
4.73 (1.62)	3.48 (2.18)		$t(47) = 5.15$ , $p < 0.001$ , $SD = 1.68$
4.73 (1.62)		2.33 (1.86)	$t(47) = 8.96$ , $p < 0.001$ , $SD = 1.85$
	3.48 (2.18)	2.33 (1.86)	$t(47) = 4.53$ , $p < 0.001$ , $SD = 1.75$

#### Time Spent Thinking About Pain

A one way repeated measure ANOVA (using Greenhouse-Geisser) showed a significant main effect of No VR vs. passive VR vs. interactive eye tracked VR for "time spent thinking about pain,"  $F(2,94) = 28.83$ ,  $p < 0.001$ ,  $MS = 79.49$ ,  $\eta_p^2 = 0.38$ .

*Post hoc* paired comparisons (paired  $t$ -tests) are shown below. As predicted, compared to No VR, participants ratings of time spent thinking about pain during the thermal stimulus were significantly lower during passive VR. Compared to No VR, time spent thinking about pain was significantly lower during interactive eye tracked VR. But contrary to predictions, compared to passive VR, time spent thinking about pain was NOT significantly lower during interactive eye tracked VR.



## TIME Spent Thinking About PAIN

### Time spent thinking about Pain

No VR	Passive VR	Eye tracked VR	<i>Post hoc</i> paired <i>t</i> -tests
2.35 (2.40)	0.625 (1.27)		$t(47) = 5.54$ , $p < 0.001$ , $SD = 2.16$
2.35 (2.40)		0.479 (0.99)	$t(47) = 5.71$ , $p < 0.001$ , $SD = 2.28$
	0.625 (1.27)	0.479 (0.99)	$t(47) = 1.00$ , $p = 0.32$ NS, $SD = 1.01$

## Fun During the Thermal Stimulus

A one way repeated measure ANOVA showed a significant main effect of No VR vs. passive VR vs. interactive eye tracked VR for “fun,”  $F(2,94) = 107.40$ ,  $p < 0.001$ ,  $MS = 234.65$ ,  $\eta_p^2 = 0.70$ .

*Post hoc* paired comparisons (paired *t*-tests) are shown below. As predicted, compared to No VR, participants’ ratings of fun during the thermal stimulus were significantly higher during passive VR. Compared to No VR, fun was significantly higher during interactive eye tracked VR. And compared to passive VR, fun was significantly higher during interactive eye tracked VR.

### FUN

#### Fun

No VR (baseline)	Passive VR	Eye tracked VR	<i>Post hoc</i> paired <i>t</i> -tests
1.40 (2.11)	3.79 (2.21)		$t(47) = 7.38$ , $p < 0.001$ , $SD = 2.25$
1.40 (2.11)		5.81 (2.42)	$t(47) = 12.92$ , $p < 0.001$ , $SD = 2.37$
	3.79 (2.21)	5.81 (2.42)	$t(47) = 8.95$ , $p < 0.001$ , $SD = 1.56$

## NAUSEA FROM VIRTUAL REALITY

No significant difference in “nausea during VR” was found between passive VR and interactive eye tracked VR.

### Nausea

No VR	Passive VR	Eye tracked VR	<i>Post hoc</i> paired <i>t</i> -tests
	0.146 (0.51)	0.21 (0.92)	$t(47) < 1$ . NS, $p = 0.685$ , $SD = 1.06$

## Presence

Compared to their illusion of presence during passive VR, participants reported having a significantly stronger illusion of presence in virtual reality (being there), during interactive eye tracked VR.

## PRESENCE IN VIRTUAL REALITY

### Presence

No VR	Passive VR	Eye tracked VR	<i>Post hoc</i> paired <i>t</i> -tests
	4.31 (2.23)	6.14 (2.65)	$t(47) = 6.24$ , $p < 0.001$ , $SD = 2.04$

## Real

Participants rated the virtual objects as significantly more real during interactive eye tracked VR, compared to during passive VR.

## HOW REAL WERE THE OBJECTS IN VIRTUAL REALITY

### Real

No VR	Passive VR	Eye tracked VR	<i>Post hoc</i> paired <i>t</i> -tests
	3.75 (2.14)	4.75 (2.47)	$t(47) = 5.12$ , $p < 0.001$ , $SD = 1.35$

In order to test an important assumption of our thermal pain paradigm, pilot data collected from an additional 24 participants in the same context received No VR during baseline, No VR during Test 1 vs. No VR again during Test 2 (see **Table 1**). As predicted, participants who received No VR did not habituate to the thermal pain stimuli. In other words, the pain ratings from the thermal pain stimulations were stable over repeated pain stimulations for people who received one baseline pain and two test pain stimuli with No VR, using the same thermal pain paradigm as the main study. As predicted, as shown in **Table 1**, three separate Bonferroni corrected One-Way repeated measure ANOVAs indicated no significant main effect of No VR vs. passive VR vs. interactive eye-tracked VR for worst pain, pain unpleasantness, or time spent thinking about pain.

## DISCUSSION

In the current study, we measured the brief acute pain of healthy volunteers during 10 s thermal pain stimuli to test whether increasing the immersiveness of the VR system increased how effectively VR reduces acute pain during brief thermal stimulations. As predicted, compared to passive VR, interactive eye tracked VR was significantly more effective

**TABLE 1 |** Results for the control group: one-way repeated measures ANOVAs for worst pain, pain unpleasantness, and time spent thinking about pain.

<b>Worst pain ratings for the Control Group (<i>n</i> = 24 subjects)</b>			
<b>Worst pain</b>			<b>Mauchly's <math>p = 0.003</math>, sphericity not assumed, Greenhouse–Geiser used. Bonferroni corrected <math>\alpha = 0.017</math>, <math>p &lt; 0.017</math> are significant</b>
<b>No VR</b>	<b>Passive VR</b>	<b>Eye tracked VR</b>	
6.25 (1.15)	6.38 (1.44)	6.75 (1.48)	$F(1.41,46) = 4.85$ , $p = 0.024$ , not significant, $MS = 2.31$ , $\eta_p^2 = 0.17$ .
<b>PAIN UNPLEASANTNESS for the Control Group (<i>n</i> = 24 subjects)</b>			
<b>Pain unpleasantness</b>			<b>Mauchly's <math>p = 0.001</math>, sphericity not assumed, Greenhouse–Geiser used. Bonferroni corrected <math>\alpha = 0.017</math>, any <math>p &lt; 0.017</math> are significant</b>
<b>No VR</b>	<b>Passive VR</b>	<b>Eye tracked VR</b>	
4.33 (2.26)	4.29 (2.05)	4.83 (2.04)	$F(1.35,46) = 4.02$ , $p = 0.043$ , not significant, $MS = 3.23$ , $\eta_p^2 = 0.15$
<b>TIME spent thinking about PAIN for the Control Group (<i>n</i> = 24)</b>			
<b>Time spent thinking about Pain</b>			<b>Mauchly's <math>p = 0.002</math>, sphericity not assumed, Greenhouse–Geiser used. Bonferroni corrected <math>\alpha = 0.017</math>, any <math>p &lt; 0.017</math> are significant</b>
<b>No VR</b>	<b>Passive VR</b>	<b>Eye tracked VR</b>	
2.96 (2.68)	2.17 (2.71)	2.75 (3.03)	$F(1.41,46) = 5.08$ , $p = 0.021$ , not significant, $MS = 5.75$ , $\eta_p^2 = 0.18$

at reducing worst pain (sensory pain), more effective at reducing pain unpleasantness (the emotional component of pain), and interactive eye tracked VR was more fun than passive VR. Furthermore, as predicted, compared to the passive VR condition, participants rated their illusion of presence significantly higher during the interactive eye tracked VR condition, and virtual objects seemed significantly more real during interactive VR, compared to passive VR.

In addition, the current study also tested an important assumption of our thermal pain paradigm. Pilot data collected from an additional 24 participants in the same context, received No VR during baseline, No VR during Test 1, and No VR again during Test 2. As predicted, the 24 participants who received No VR reported no reduction in pain. In other words, the pain ratings from the thermal pain stimulations were stable over repeated pain stimulations for people who received one baseline pain and two test pain stimuli with No VR, using the same thermal pain paradigm as the main study.

## Limitations

The within-subjects study design used in the current study reduces noise variance and increases statistical power. However, one limitation of our current study is that researchers and subjects were not blinded to the treatment conditions (Campbell and Stanley, 1963; Schulz and Grimes, 2002). In the current study, VR was “visual only” with no sound effects. Having sounds muted may have exaggerated the benefit from the eye movement interactivity. One of the advantages of VR is the multimedia exposure. Sounds of rivers running and birds singing enhance the illusion of presence and heighten the participant's sensitivity to the 3D environment. On the other hand, the lack of sound effects in the current study may have underestimated the benefits of eye movement interactivity, because there would be more sound effects in the interactive VR condition, which could

make interactive VR more distracting and make the advantage of interactive VR over passive VR even more pronounced. Another limitation is that this study did not investigate/measure participants' attention level (e.g., Heathcote et al., 2017). In the current study, eye tracking technology was simply used as a naturalistic human computer interface. Future, more advanced versions of our new VR analgesia system may use eye blinks and duration of fixation to help define interactions, and further increase interactivity. In the current study, “hands free” eye tracked interactive VR was compared to passive non-interactive VR. Whether eye tracking increases analgesia compared to other “hands free” interactive VR systems (e.g., voice controlled, etc.) remains a possible topic for future research. In the current study, each individual selected a temperature they found “painful but tolerable for 10 s.” Clinical research is needed to determine whether the current results generalize to clinical pain settings (e.g., for 20 min wound cleaning procedures of severe burn patients that often involve severe to excruciating pain). But, encouragingly, our first pediatric burn patient pilot subject, using the same eye tracked VR analgesia system, reported unusually large reductions in pain during burn wound care in the intensive care unit, during eye-tracked VR vs. No VR.

Despite these limitations, the current study is the first PubMed indexed VR analgesia study to involve eye tracking technology embedded into the VR goggles. Although there is a large scientific literature on traditional eye tracking spanning more than 40 years of research (Kredel et al., 2017), there are very few studies combining eye tracking and immersive virtual reality, a very recent innovation/combination.

## Why Does VR Reduce Pain? Possible Mechanisms of VR Analgesia

Although there is growing evidence that VR can be effective for reducing acute pain during painful medical procedures

(Hoffman et al., 2019), the non-pharmacologic mechanism of how VR reduces pain is not fully understood and is an important research topic. According to Ballantyne (2018, p. s25) “The conscious perception of pain depends on the conversion of nociception to perception. . .”. The authors of the current study speculate that VR interferes with the conversion of nociception to conscious pain perception by inserting a powerful perceptual illusion into the painful experience. Instead of directing most of their attention toward converting nociceptive signals into pain perception during No VR, we speculate that during virtual reality, the patient’s brain is pre-occupied with converting neural signals from the visual, auditory and other sensory systems into a multisensory perceptual illusion of “presence” in virtual reality.

In the current study, we predicted that adding hands free interactive eye tracking would make VR that much more attention grabbing, and thus more effective at reducing pain, compared to passive VR. As predicted, participants in the current study also reported a significantly stronger illusion of presence, and VR was significantly more fun during the eye tracked VR. Although the current study did not directly measure attention, the pattern of results of the current study are consistent with an attentional mechanism of how VR reduces pain (see also Hoffman, 1998, 2004; Gold et al., 2006, 2007; Dahlquist et al., 2007; Birnie et al., 2017; Gold and Mahrer, 2017; Zeroth et al., 2019).

## Future Directions

Whether the current results generalize to clinical patient populations is an important topic for future research (e.g., whether eye tracking increases VR analgesia effectiveness for pediatric burn patients during burn wound care). In the current laboratory thermal pain study with healthy volunteers, eye movements were used to tell the computer what virtual objects the participant was looking at in virtual reality. In future studies on VR analgesia (e.g., virtual reality pain distraction), eye tracking technology can also be used to collect data about the patient’s current mental state. For example, pupil size, and patterns of eye movements may correlate with how much pain patients are consciously experiencing. When a burn patient’s pain becomes so extreme that the patient’s attention shifts away from VR and onto their pain, we predict large reductions in successful eye fixations on target objects in SnowCanyon.

Immersive Virtual Reality with eye tracking has wide potential clinical uses beyond pediatric burn patients. For example, VR has recently been used with spinal injury patients (Flores et al., 2018). Most paralyzed patients are able to move their eyes, and can thus use eye movements to interact with objects in the virtual world. In the future, eye-tracked virtual reality and also augmented

reality glasses may allow people to quantify and improve the efficiency of information processing/learning, etc. And there is also growing interest in using eye tracking to help improve social skills (e.g., helping autistic patients learn to make more natural patterns of eye contact with other humans). Additional research and development is recommended.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki. Both written and verbal informed consent were obtained using a protocol approved by the Effat University’s Human Subjects Review Committee.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This research was supported by the King Abdulaziz City for Science and Technology (KACST) grant no. 1425-37-ط. This research was also supported by the Effat University Research and Consultancy Institute, in Jeddah, Saudi Arabia. HH’s salary (via the University of Washington in the United States) was covered by generous support of the Mayday Fund, by NIH grant R01GM042725 to David R. Patterson, and by Shriners Hospital for Children grant (award ID #71011-GAL, PI WM).

## ACKNOWLEDGMENTS

This study was completed in partial fulfillment of the requirements for the degree of Ph.D. in Computer Science by NA-G. We would like to thank Annechien Helsdingen, Ph.D., research assistants Ahad Alhudali, Nada Aljohani, Maya Akbar, Maryam Daffa and the students at Effat University who volunteered to participate in this study.

## REFERENCES

- Atzori, B., Hoffman, H. G., Vagnoli, L., Patterson, D. R., Alhalabi, W., Messeri, A., et al. (2018a). Virtual reality analgesia during venipuncture in pediatric patients with onco-hematological diseases. *Front. Psychol.* 9:2508. doi: 10.3389/fpsyg.2018.02508
- Atzori, B., Lauro Grotto, R., Giugni, A., Calabrò, M., Alhalabi, W., and Hoffman, H. G. (2018b). Virtual reality analgesia for pediatric dental patients. *Front. Psychol.* 9:2265. doi: 10.3389/fpsyg.2018.02265
- Ballantyne, J. C. (2018). The brain on opioids. *Pain* 159(Suppl. 1), S24–S30. doi: 10.1097/j.pain.0000000000001270
- Birnie, K. A., Chambers, C. T., and Spellman, C. M. (2017). Mechanisms of distraction in acute pain perception and modulation. *Pain* 158, 1012–1013. doi: 10.1097/j.pain.0000000000000913
- Campbell, D. T., and Stanley, J. C. (1963). *Experimental And Quasi-Experimental Designs for Research*, Vol. 6. Boston, MA: Houghton Mifflin Company, 47–50.
- Cherny, N., Ripamonti, C., Pereira, J., Davis, C., Fallon, M., McQuay, H., et al. (2001). Strategies to manage the adverse effects of oral

- morphine: an evidence-based report. *J. Clin. Oncol.* 19, 2542–2554. doi: 10.1200/jco.2001.19.9.2542
- Dahlquist, L. M., McKenna, K. D., Jones, K. K., Dillinger, L., Weiss, K. E., and Ackerman, C. S. (2007). Active and passive distraction using a head-mounted display helmet: effects on cold pressor pain in children. *Health Psychol.* 26, 794–801. doi: 10.1037/0278-6133.26.6.794
- Faber, A. W., Patterson, D. R., and Bremer, M. (2013). Repeated use of immersive virtual reality therapy to control pain during wound dressing changes in pediatric and adult burn patients. *J. Burn Care Res.* 34, 563–568. doi: 10.1097/BCR.0b013e3182777904
- Fields, H. L. (2018). How expectations influence pain. *Pain* 159(Suppl. 1), S3–S10. doi: 10.1097/j.pain.0000000000001272
- Fischer, S., Vinall, J., Pavlova, M., Graham, S., Jordan, A., Chorney, J., et al. (2018). Role of anxiety in young children's pain memory development after surgery. *Pain* [Epub ahead of print].
- Flores, A., Linehan, M. M., Todd, S. R., and Hoffman, H. G. (2018). The use of virtual reality to facilitate mindfulness skills training in dialectical behavioral therapy for spinal cord injury: a case study. *Front. Psychol.* 9:531. doi: 10.3389/fpsyg.2018.00531
- Garrett, B., Taverner, T., Masinde, W., Gromala, D., Shaw, C., and Negraeff, M. (2014). A rapid evidence assessment of immersive virtual reality as an adjunct therapy in acute pain management in clinical practice. *Clin. J. Pain* 30, 1089–1098. doi: 10.2196/medinform.7271
- Gold, J. I., Belmont, K. A., and Thomas, D. A. (2007). The neurobiology of virtual reality pain attenuation. *CyberPsychol. Behav.* 10, 536–544. doi: 10.1089/cpb.2007.9993
- Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., and Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during IV placement. *CyberPsychol. Behav.* 9, 207–212. doi: 10.1089/cpb.2006.9.207
- Gold, J. I., and Mahrer, N. E. (2017). Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. *J. Pediatr. Psychol.* 43, 266–275. doi: 10.1093/jpepsy/jsx129
- Heathcote, L. C., Lau, J. Y., Mueller, S. C., Eccleston, C., Fox, E., Bosmans, M., et al. (2017). Child attention to pain and pain tolerance are dependent upon anxiety and attention control: an eye-tracking study. *Eur. J. Pain* 21, 250–263. doi: 10.1002/ejp.920
- Hemington, K. S., Cheng, J. C., Bosma, R. L., Rogachov, A., Kim, J. A., and Davis, K. D. (2017). Beyond negative pain-related psychological factors: resilience is related to lower pain affect in healthy adults. *J. Pain* 18, 1117–1128. doi: 10.1016/j.jpain.2017.04.009
- Hendrix, C., and Barfield, W. (1995). "Presence in virtual environments as a function of visual and auditory cues," in *Proceedings of VRAIS '95* (Research Triangle Park, NC: IEEE Computer Society Press), 74–82.
- Hoffman, H. G. (1998). Virtual reality: a new tool for interdisciplinary psychology research. *CyberPsychol. Behav.* 1, 195–200. doi: 10.1089/cpb.1998.1.195
- Hoffman, H. G. (2004). Virtual reality therapy. *Sci. Am.* 291, 58–65.
- Hoffman, H. G., Chambers, G. T., Meyer, W. J. III, Arceneaux, L. L., Russell, W. J., Seibel, E. J., et al. (2011). Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. *Ann. Behav. Med.* 41, 183–191. doi: 10.1007/s12160-010-9248-7
- Hoffman, H. G., Doctor, J. N., Patterson, D. R., Carrougner, G. J., and Furness, T. A. III (2000). Use of virtual reality as an adjunctive treatment of adolescent burn pain during wound care: a case report. *Pain* 85, 305–309. doi: 10.1016/s0304-3959(99)00275-4
- Hoffman, H. G., Garcia-Palacios, A., Carlin, A., Furness, T. A. I. I., and Botella-Arbona, C. (2003a). Interfaces that heal: coupling real and virtual objects to treat spider phobia. *Int. J. Hum. Comput. Interact.* 16, 283–300. doi: 10.1207/s15327590ijhc1602\_08
- Hoffman, H. G., Garcia-Palacios, A., Kapa, V. A., Beecher, J., and Sharar, S. (2003b). R. Immersive virtual reality for reducing experimental ischemic pain. *Int. J. Hum. Comput. Interact.* 15, 469–486. doi: 10.1207/s15327590ijhc1503\_10
- Hoffman, H. G., Meyer, W. J. III, Ramirez, M., Roberts, L., Seibel, E. J., Atzori, B., et al. (2014). Feasibility of articulated arm mounted oculus rift virtual reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients. *Cyberpsychol. Behav. Soc. Netw.* 17, 397–401. doi: 10.1089/cyber.2014.0058
- Hoffman, H. G., Patterson, D. R., Carrougner, G. J., and Sharar, S. R. (2001). Effectiveness of virtual reality-based pain control with multiple treatments. *Clin. J. Pain* 17, 229–235. doi: 10.1097/00002508-200109000-00007
- Hoffman, H. G., Patterson, D. R., Magula, J., Carrougner, G. J., Zeltzer, K., Dagadakis, S., et al. (2004a). Water-friendly virtual reality pain control during wound care. *J. Clin. Psychol.* 60, 189–195. doi: 10.1002/jclp.10244
- Hoffman, H. G., Richards, T. L., Coda, B., Bills, A. R., Blough, D., Richards, A. L., et al. (2004b). Modulation of thermal pain-related brain activity with virtual reality: evidence from fMRI. *Neuroreport* 15, 1245–1248. doi: 10.1097/01.wnr.0000127826.73576.91
- Hoffman, H. G., Richards, T. L., Van Oostrom, T., Coda, B. A., Jensen, M. P., Blough, D. K., et al. (2007). The analgesic effects of opioids and immersive virtual reality distraction: evidence from subjective and functional brain imaging assessments. *Anesth. Analg.* 105, 1776–1783. doi: 10.1213/01.ane.0000270205.45146.db
- Hoffman, H. G., Rodriguez, R. A., Gonzalez, M., Bernardy, M., Peña, R., Beck, W., et al. (2019). Immersive virtual reality as an adjunctive non-opioid analgesic for predominantly Latin American children with large severe burn wounds during burn wound cleaning in the Intensive Care Unit: a pilot study. *Front. Hum. Neurosci.* 13:262. doi: 10.3389/fnhum.2019.00262
- Hoffman, H. G., Seibel, E. J., Richards, T. L., Furness, T. A., Patterson, D. R., and Sharar, S. R. (2006). Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. *J. Pain* 7, 843–850. doi: 10.1016/j.jpain.2006.04.006
- Hoffman, H. G., Sharar, S. R., Coda, B., Everett, J. J., Ciol, M., Richards, T., et al. (2004c). Manipulating presence influences the magnitude of virtual reality analgesia. *Pain* 111, 162–168. doi: 10.1016/j.pain.2004.06.013
- Indovina, P., Barone, D., Gallo, L., Chirico, A., De Pietro, G., and Antonio, G. (2018). Virtual reality as a distraction intervention to relieve pain and distress during medical procedures: a comprehensive literature review. *Clin. J. Pain* 34, 858–877. doi: 10.1097/AJP.0000000000000599
- Jeffs, D., Dorman, D., Brown, S., Files, A., Graves, T., Kirk, E., et al. (2014). Effect of virtual reality on adolescent pain during burn wound care. *J. Burn Care Res.* 35, 395–408. doi: 10.1097/BCR.0000000000000019
- Jensen, M., and Karoly, P. (2001). "Self-report scales and procedures for assessing pain in adults," in *Handbook of Pain Assessment*, 2nd Edn, eds D. Turk, and R. Melzack (New York: Guilford Publications), 15–34.
- Jensen, M. P. (2003). The validity and reliability of pain measures in adults with cancer. *J. Pain* 4, 2–21. doi: 10.1054/jpai.2003.1
- Keefe, F. J., Huling, D. A., Coggins, M. J., Keefe, D. F., Zachary, R. M., et al. (2012). Virtual reality for persistent pain: a new direction for behavioral pain management. *Pain* 153, 2163–2166. doi: 10.1016/j.pain.2012.05.030
- Keefe, F. J., Main, C. J., and George, S. Z. (2018). Advancing psychologically informed practice for patients with persistent musculoskeletal pain: promise, pitfalls, and solutions. *Phys. Ther.* 98, 398–407. doi: 10.1093/ptj/pzy024
- Khadra, C., Ballard, A., Dery, J., Paquin, D., Fortin, J. S., Perreault, I., et al. (2018). Projector-based virtual reality dome environment for procedural pain and anxiety in young children with burn injuries: a pilot study. *J. Pain Res.* 14, 343–353. doi: 10.2147/JPR.S151084
- Kipping, B., Rodger, S., Miller, K., and Kimble, R. M. (2012). Virtual reality for acute pain reduction in adolescents undergoing burn wound care: a prospective randomized controlled trial. *Burns* 38, 650–657. doi: 10.1016/j.burns.2011.11.010
- Krane, E. J. (2019). The opioid debate-PRO: opioids have an important role in pain management in children. *Clin. J. Pain* [Epub ahead of print].
- Kredel, R., Vater, C., Klostermann, A., and Hossner, E. J. (2017). Eye-tracking technology and the dynamics of natural gaze behavior in sports: a systematic review of 40 years of research. *Front. Psychol.* 8:1845. doi: 10.3389/fpsyg.2017.01845
- Law, E. F., Dahlquist, L. M., Sil, S., Weiss, K. E., Herbert, L. J., Wohlheiter, K., et al. (2011). Videogame distraction using virtual reality technology for children experiencing cold pressor pain: the role of cognitive processing. *J. Pediatr. Psychol.* 36, 84–94. doi: 10.1093/jpepsy/jsq063
- Maani, C., Hoffman, H. G., DeSocio, P. A., Morrow, M., Galin, C., Magula, J., et al. (2008). Pain control during wound care for combat-related burn injuries using custom articulated arm mounted virtual reality goggles. *J. Cyberther. Rehabil.* 1, 193–198.



- Maani, C. V., Hoffman, H. G., Fowler, M., Maiers, A. J., Gaylord, K. M., and Desocio, P. A. (2011a). Combining ketamine and virtual reality pain control during severe burn wound care: one military and one civilian patient. *Pain Med.* 12, 673–678. doi: 10.1111/j.1526-4637.2011.01091
- Maani, C. V., Hoffman, H. G., Morrow, M., Maiers, A., Gaylord, K., McGhee, L. L., et al. (2011b). Virtual reality pain control during burn wound debridement of combat-related burn injuries using robot-like arm mounted VR goggles. *J. Trauma* 71 (Suppl. 1), S125–S130. doi: 10.1097/TA.0b013e31822192e2
- Malchow, R. J., and Black, I. H. (2008). The evolution of pain management in the critically ill trauma patient: emerging concepts from the global war on terrorism. *Crit. Care Med.* 36, S346–S357. doi: 10.1097/CCM.0b013e31817e2fc9
- McIntyre, M. K., Clifford, J. L., Maani, C. V., and Burmeister, D. M. (2016). Progress of clinical practice on the management of burn-associated pain: lessons from animal models. *Burns* 42, 1161–1172. doi: 10.1016/j.burns.2016.01.023
- McSherry, T., Atterbury, M., Gartner, S., Helmold, E., Searles, D. M., and Schulman, C. (2018). Randomized, crossover study of immersive virtual reality to decrease opioid use during painful wound care procedures in adults. *J. Burn Care Res.* 39, 278–285. doi: 10.1097/BCR.0000000000000589
- Melzack, R., and Wall, P. D. (1965). Pain mechanisms: a new theory. *Science* 150, 971–979. doi: 10.1126/science.150.3699.971
- Mendez-Romero, D., Clark, A. T., Christie, A., and Wolf, S. E. (2018). Weight changes and patterns of weight measurements in hospitalized burn patients: a contemporary analysis. *Burns Trauma* 6:30. doi: 10.1186/s41038-018-0131-2
- Noel, M., Chambers, C. T., McGrath, P. J., Klein, R. M., and Stewart, S. H. (2012). The influence of children's pain memories on subsequent pain experience. *Pain* 153, 1563–1572. doi: 10.1016/j.pain.2012.02.020
- Noel, M., Palermo, T. M., Chambers, C. T., Taddio, A., and Hermann, C. (2015a). Remembering the pain of childhood: applying a developmental perspective to the study of pain memories. *Pain* 156, 31–34. doi: 10.1016/j.pain.0000000000000001
- Noel, M., Rabbitts, J. A., Tai, G. G., and Palermo, T. M. (2015b). Remembering pain after surgery: a longitudinal examination of the role of pain catastrophizing in children's and parents' recall. *Pain* 156, 800–808. doi: 10.1097/j.pain.0000000000000102
- Schulz, K. F., and Grimes, D. A. (2002). Blinding in randomised trials: hiding who got what. *Lancet* 359, 696–700. doi: 10.1016/s0140-6736(02)07816-9
- Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in immersive virtual environments. *Presence Teleoper. Virtual Environ.* 3, 130–144.
- Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence Teleoper. Virtual Environ.* 6, 603–616. doi: 10.1162/pres.1997.6.6.603
- Tesler, M. D., Savedra, M. C., Holzemer, W. L., Wilkie, D. J., Ward, J. A., and Paul, S. M. (1991). The word-graphic rating scale as a measure of children's and adolescents' pain intensity. *Res. Nurs. Health* 14, 361–371. doi: 10.1002/nur.4770140507
- Vijayan, R. (2011). Managing acute pain the developing world. *Pain Clin. Updates* 19, 1–7. doi: 10.1002/ijgo.12802
- Wender, R., Hoffman, H. G., Hunner, H. H., Seibel, E. J., Patterson, D. R., and Sharar, S. R. (2009). Interactivity Influences the Magnitude of Virtual Reality Analgesia. *J. Cyber Ther Rehabil.* 2, 27–33.
- Yaster, M., Reid, A. L., Cohen, M. N., and Monitto, C. L. (2019). Opioids in the management of acute pediatric pain: an update in a time of crisis. *Clin. J. Pain* doi: 10.1097/AJP.0000000000000709 [Epub ahead of print].
- Zeroth, J. A., Dahlquist, L. M., and Foxen-Craft, E. C. (2019). The effects of auditory background noise and virtual reality technology on video game distraction analgesia. *Scand. J. Pain* 19, 207–217. doi: 10.1515/sjpain-2018-0123

**Conflict of Interest:** After the current study was completed, HH joined the Scientific Advisory Board of BehaVR.com, but no products or funding from this source were involved in the current study.

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

Copyright © 2020 Al-Ghamdi, Meyer, Atzori, Alhalabi, Seibel, Ullman and Hoffman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Adoption of New Treatment Modalities by Health Professionals and the Relative Weight of Empirical Evidence in Favor of Virtual Reality Exposure Versus Mindfulness in the Treatment of Anxiety Disorders

Kevin Nolet<sup>1</sup>, Giulia Corno<sup>1,2</sup> and Stéphane Bouchard<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Pietro Cipresso,  
Italian Auxological Institute (IRCCS),  
Italy

### Reviewed by:

Daniel David,  
Babeş-Bolyai University, Romania  
Julia Elisabeth Diemer,  
kbo Inn-Salzach-Klinikum, Germany

### \*Correspondence:

Stéphane Bouchard  
stephane.bouchard@uqo.ca

### Specialty section:

This article was submitted to  
Health,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 16 September 2019

**Accepted:** 27 February 2020

**Published:** 25 March 2020

### Citation:

Nolet K, Corno G and Bouchard S  
(2020) The Adoption of New  
Treatment Modalities by Health  
Professionals and the Relative Weight  
of Empirical Evidence in Favor  
of Virtual Reality Exposure Versus  
Mindfulness in the Treatment  
of Anxiety Disorders.  
*Front. Hum. Neurosci.* 14:86.  
doi: 10.3389/fnhum.2020.00086

<sup>1</sup> Cyberpsychology Laboratory of UQO, Department of Psychoeducation and Psychology, Université du Québec en Outaouais, Gatineau, QC, Canada, <sup>2</sup> LabPsiTec, Departamento de Personalidad, Evaluación y Tratamientos Psicológicos, Universitat de València, Valencia, Spain

Anxiety disorders are among the most prevalent mental disorders, and cognitive-behavioral therapy (CBT) with exposure exercises is considered as the gold-standard psychological intervention. New psychotherapeutic modalities have emerged in the last decade and, among them, mindfulness has been rapidly adopted by therapists. The adoption rate is slower for the use of virtual reality (VR) to conduct exposure. The goal of the present position paper is to contrast, for the treatment of anxiety disorders, the weight of empirical evidences supporting the use of exposure in VR with the use of mindfulness-based therapy (MBT). Based on the most recent meta-analyses, we found that CBT with exposure conducted in VR was more thoroughly researched and supported than MBT, receiving validation from roughly twice as many studies with high control (i.e., randomized, active controls with clinical samples). However, this conclusion is nuanced by reviewing gaps in the literature for both therapies. Potential factors influencing clinicians' choice of treatment and suggestions for future research directions are proposed.

**Keywords:** anxiety disorders, exposure therapy, cognitive behavioral therapy, virtual reality, mindfulness

## INTRODUCTION

Anxiety disorders are highly prevalent (Bandelow and Michaelis, 2015). They involve dysfunctional information processing from the limbic system. As such, cognitive-behavioral therapy (CBT) is recognized as the treatment of choice (Katzman et al., 2014; Nathan and Gorman, 2015; David et al., 2018). CBT is based on the premise that hyperactivation of the amygdala is maintained

by the interplay between environmental, biological, cognitive, and behavioral factors and that psychological interventions lead to changes in information processing of threat-related cues from the limbic system through active cognitive and behavioral changes. Techniques based on exposure and its variations, such as behavioral experiments or response prevention, are considered as the key strategies leading to significant clinical improvement (Craske et al., 2014). However, although CBT with exposure has been considered as the gold-standard psychosocial intervention for treating anxiety disorders (Hofmann and Smits, 2008; Otte, 2011), it is not without limitations.

Exposure is usually combined with other cognitive behavior techniques, including anxiety management and cognitive restructuring of dysfunctional cognitions (Abramowitz et al., 2011). A successful exposure allows the client to learn to tolerate her/his fear and anxiety while developing a new behavioral repertoire rather than relying on threat avoidance, leading to new mental associations in the limbic system with lack of threat and with stronger perceived self-efficacy for managing emotions and previously avoided situations. While exposure-based treatments are listed as one of the evidenced-based treatments (EBT) for anxiety disorders by the Division 12 of the American Psychological Association (Chambless and Ollendick, 2001) and the National Institute for Health and Care Excellence (2014) guidelines, its dissemination is confronted with numerous barriers (Hembree and Cahill, 2007). For example, when surveyed about their practice with patients suffering from anxiety disorders, the practitioners report opting for alternative therapies with less empirical support than CBT with exposure. The negative beliefs about exposure, namely, in terms of its safety and tolerability for the patient, and the impracticability of its implementation have been found predictive of the lack of use among therapists (Pittig et al., 2019).

Several therapeutic alternatives to standard exposure have emerged, notably conducting exposure in virtual reality (VR; Wiederhold and Bouchard, 2014). CBT with exposure conducted in VR (CBT-VRexp) has been developed to counter some of the limitations of *in vivo* exposure. By offering a standardized, controlled, and replicable environment that can elicit emotions for therapeutic purposes, VR is a medium that could be more practical and effective for exposure therapy.

Although extensively studied in the scientific community, the use of CBT-VRexp is not widespread among clinicians who tend to favor interventions from other paradigms, such as mindfulness-based therapy (MBT).

Documenting which forms and which psychotherapy and their variations are adopted by therapists and at which rate is challenging. Not only those data are scarce in the literature, they can also vary between mental disorders and countries, inducing biases that further limit the comparisons. In addition to the difficulty in recruiting a representative sample of therapists using probabilistic techniques, generalizing the results on the therapists' adoption of various forms of psychotherapy could be problematic. Keeping the above limitations in mind, in a survey of German behavioral therapists working in the healthcare system, exposure was used in only 46.8% of treatments focusing on anxiety disorders

(Pittig and Hoyer, 2018). Although not specifically focusing on anxiety disorders, Michalak et al. (2020) found, again in a German sample, that up to 82% of licensed therapists integrate MBT in their clinical practice, most of them (80%) using it at least occasionally (fewer than one out of two sessions). However, only 10% of those used a manualized group-based MBT. In their samples, the therapists preferred to integrate in their treatment plans stand-alone interventions such as body scan, breathing meditation, self-soothing with the five senses, or other informal practices. In a sample of practicing CBTs attending a European clinical conference, with the majority working with anxiety disorders, only 13.67% reported using CBT-VRexp occasionally or frequently with their patients (Lindner et al., 2019).

Informal observations rapidly show that the number of training, workshops, and classes on MBT clearly outweigh those on VR or CBT-VRexp. A search on Google<sup>1</sup> with the keywords "anxiety" and "VR workshops" yielded 230 results and 23,000 results with the keywords "anxiety" and "mindfulness workshops". The specific numbers vary when other keywords are used, but the ratios remain in the order of 1 to 100 in favor of MBT. Finally, although MBT has experienced a marked increase in scientific and popular interest in the past two decades, recent commentaries (e.g., Farias et al., 2016) have raised questions regarding the evidence base for this family of therapies.

The current paper was motivated by the apparent enthusiasm of mental health professionals to embrace some variations of CBT for anxiety disorders more than others. The aim of this review is to contrast the bulk of evidences supporting the efficacy of treatment of anxiety disorders using CBT-VRexp versus using MBT. The goal is *not* to compare the relative efficacy of both forms of CBT but specifically to compare the amount, or weight, of empirical evidences supporting each of them and contrast it with the therapists' enthusiasm to adopt each of them. The weight was defined here as the number, and relative efficacy, of randomized controlled trials (RCTs) conducted with clinical samples comparing a treatment with at least another treatment, ideally a treatment considered as an established standard.

## METHODS

The general methodology will follow three steps: (a) define CBT-VRexp and MBT, (b) provide a brief overview of their relevance for the treatment of anxiety and its disorders [i.e., as defined in DSM-5, with the addition of obsessive-compulsive disorder and post-traumatic stress disorder (PTSD)], and (c) review and contrast the relative weight of empirical support for both techniques based on already published meta-analyses.

## Defining CBT-VRexp and Mindful-Based Interventions

VR has been defined in different ways, but the practical definition from Schultheis and Rizzo (2001) will be used here: VR is an advanced form of human-computer interface that

<sup>1</sup>On July 5, 2019.

allows the user to “interact” with and become “immersed” in a computer-generated environment in a naturalistic fashion. Three main features differentiate VR systems from other technologies: immersion, the impression of really being in the environment, and interaction with that environment (e.g., Biocca, 1997; Lombard and Ditton, 1997; Slater, 2009; Fuchs, 2011; Wiederhold and Bouchard, 2014; Cipresso et al., 2018). Computer-generated virtual environments allow clinical assessment, treatment, and rehabilitation, providing interactive ecologically valid scenarios designed to target specific needs.

Cognitive-behavioral therapy-VRexp refers to the use of VR to conduct exposure (Bouchard and Rizzo, 2019). CBT rarely relies only on exposure, although it clearly is the main component of CBT for phobias. Additional therapeutic ingredients include working alliance, case conceptualization, psychoeducation, cognitive reframing, and relapse prevention. For more complex anxiety disorders, the treatment always includes the aforementioned ingredients, plus a stronger involvement of cognitive techniques, and may also involve other techniques, such as problem solving or assertiveness training.

In traditional CBT, many strategies are designed to change internal experiences, such as emotional states (e.g., reducing negative moods), bodily sensations (e.g., reducing pain), and the content of thoughts (e.g., from irrational and/or distorted to rational, realistic, and/or balanced) (Harrington and Pickles, 2009). To the contrary, mindfulness-based approaches teach an alternative way of relating to such experiences. Bishop et al. (2004) identified two basic components of mindfulness: one involves self-regulation of attention and another one involves an orientation toward the present moment in a way characterized by openness, curiosity, and acceptance (Hofmann et al., 2010). In other words, the essential premise underlying mindfulness practices is that experiencing the present moment in a non-judgmental and open way can effectively counter the effects of stressors, as excessive orientation toward the past or future when dealing with stressors can be related to depressive and anxious feelings (e.g., Kabat-Zinn, 2003; Hofmann et al., 2010).

Therefore, mindfulness practice encourages cultivating a new relationship with internal experiences that involves directing attention in a way that it is maintained on immediate experience, without avoiding, over engaging, or elaborating the experience (Kumar et al., 2008). More specifically, it is believed that, by a training focused on approaching stressful situations more reflectively rather than reflexively, mindfulness-based interventions (MBI) can effectively counter the use of avoidance strategies, which attempt to alter the intensity or frequency of unwanted internal experiences (Hayes et al., 2006; Hofmann et al., 2010). These maladaptive strategies are believed to contribute to the maintenance of many, if not all, emotional disorders (Bishop et al., 2004; Hayes, 2004; Hofmann et al., 2010). An important and contrasting feature of MBT is how cognitions are handled. Instead of trying to restructure them, MBT focuses on accepting them and letting them go. However, when it comes to being in contact with feared stimuli, acceptance and orienting attention to fully experience the moment share many similarities

with exposure in terms of opportunities to build new mental associations with lack of threat.

## Relevance of VR and MBT in Psychotherapy

VR is used in a wide range of fields, such as physical and neurological rehabilitation (e.g., Schultheis et al., 2002; Holden, 2005; Lam et al., 2006), neuropsychological evaluation (e.g., Rizzo and Buckwalter, 1997; Rizzo et al., 2000), education, and cognitive neuroscience (e.g., Tarr and Warren, 2002). VR started to be used in clinical psychology in the early 1990s. The most common application of VR in clinical psychology has been the treatment of phobias and anxiety-related disorders (i.e., anxiety disorders as defined in the DSM-IV). For example, in the early 1990s, Hodges et al. (1995) reported to have been using virtual environments to provide acrophobic patients with fear-producing experiences of heights in a safe situation. Since that time, VR has been proposed as a new medium for conducting exposure. The rationale behind its use is that the exposure can be conducted with more control from the therapist. CBT-VRExp offers several other advantages over *in vivo* or imaginal exposure (see Côté and Bouchard, 2008 for a detailed list), such as increased attractiveness for patients, more cost-effective, better protection of confidentiality and patient's safety, etc.

The research in this field shows that VR is able to reduce the anxiety symptoms significantly in different anxiety disorders: social anxiety (SA) disorder (Bouchard et al., 2017), generalized anxiety disorder (GAD) (e.g., Repetto et al., 2013), phobias (e.g., Garcia-Palacios et al., 2002; Parsons and Rizzo, 2008), PTSD (e.g., Gonçalves et al., 2012), panic disorder (PD) and agoraphobia (Botella et al., 2007), and psychological stress (Gaggioli et al., 2014). Studies show that the clinical outcome is superior to waitlist control conditions and comparable to *in vivo* exposure-based interventions. During the last decade, clinicians extended this field to more complex disorders, for instance, eating disorders and body image disturbance (e.g., Ferrer-Garcia et al., 2013; Corno et al., 2018), schizophrenia (e.g., da Costa and de Carvalho, 2004; Freeman, 2008; Kim et al., 2008), and building resilience and post-traumatic growth (e.g., Corno and Bouchard, 2015).

In terms of criticisms, although it is a promising therapeutic medium, adding VR to CBT may not always provide additional benefit to exposure-based therapy (e.g., McLay et al., 2017) and adds costs and complexity to an already effective treatment. The exact role of some psychological factors involved in exposure conducted in VR also needs to be clarified. While the sense of presence, the feeling of being *inside* the virtual environment, has been considered as relevant for treatment success, studies about its actual impact on treatment outcomes have produced mixed results (Botella et al., 2017).

Mindfulness-based therapy has been defined as comprising the third wave of CBT because of its differences with the first two waves, behavior therapy and cognitive therapy (Hayes, 2004; Baer and Sauer, 2009). Having its origins in Eastern Buddhist tradition that is over 2,500 years old, MBT includes mindfulness-based cognitive therapy



(MBCT; e.g., Segal et al., 2002) and mindfulness-based stress reduction (MBSR; e.g., Kabat-Zinn, 1982). MBT has become a very popular form of treatment in contemporary psychotherapy (e.g., Kabat-Zinn, 1994; Bishop, 2002; Baer, 2003; Hayes, 2004). For instance, both MBCT and MBSR have demonstrated significant clinical efficacy in the treatment of mood disorders (e.g., Segal et al., 2010), resistant depression (e.g., Kenny and Williams, 2007; Eisendrath et al., 2008), and anxiety disorders (e.g., Evans et al., 2008; Kim et al., 2009; Hofmann et al., 2010). Other clinical applications include pain management, substance use, attention disorders, PTSD, and eating disorders [see Wielgosz et al. (2019) for a review of these applications].

Despite the popularity of MBT, a limited number of clinical trials have specifically examined this treatment in anxiety disorders. More specifically, while the empirical support for the treatment of recurrent depression seems to be strong, the same cannot be as easily said for other clinical-like anxiety disorders. Questions about the methodological qualities of the literature have also been raised, ranging from a lack of active control groups (Farias et al., 2016) to problems in operationalization and measurements (Grossman, 2019).

## Contrasting the Relative Weight of Empirical Support

Previous researchers have worked in great lengths to find all available outcome studies on CBT-VRexp and MBT in order to publish meta-analyses, and contrasting the adoption of treatment modalities by therapists based on information already available to them leads to a fairer comparison. Therefore, our source of information to balance the weight of evidences providing empirical support for both techniques is based on the most recent and comprehensive meta-analyses for each treatment modality. We searched the Scopus database for meta-analyses on clinical trials for MBT, CBT-VRexp, and also CBT with *in vivo* exposure as a gold-standard comparison. We complemented this search with Google Scholar to ensure that all relevant papers were found. We used the following terms in the title, abstract, and keywords: “meta-analysis” and “anxiety”, combined with “mindfulness”, “virtual reality” or “VR” or “VRET”, and “CBT” or “exposure therapy” or “cognitive behavioral therapy”. We limited our search to papers written in English and published in peer-reviewed journals. To ensure the longest possible coverage, we aimed for the most recent meta-analysis, thus limiting our search to papers published between 2018 and July 15, 2019. The papers were reviewed by the first author following specific inclusion and exclusion criteria.

The following criteria were used to identify the meta-analyses that responded to our needs: (1) the longest coverage possible (i.e., the most recent papers going as far back possible in publication history), (2) information available on the randomization procedure used for the studies included, (3) effect sizes (ES) for each control type separately (i.e., inactive, active, and evidence-based), (4) ES calculated on an anxiety measure, and (5) preferably with patients diagnosed with an

anxiety- or stress-related disorder or with a score above the cutoff on a clinical measure. When available, we also used results pertaining to attrition and deterioration rates within these meta-analyses.

To make sure that the meta-analyses included were representative of a larger part of the literature, we excluded papers if (1) they were limited to a specific population (e.g., youth or elderly), (2) they were limited to one particular therapy or modality (e.g., self-compassion for MBT and online interventions), and (3) they were limited to one diagnostic category (e.g., phobias for CBT-VRexp) or did not provide information about each included category individually; to make sure that the ES were observed specifically for anxiety, we excluded meta-analyses if (4) they aggregated heterogeneous outcomes (e.g., psychological distress); and papers were also excluded if (5) the treatment modality was not objectively isolated (e.g., by adding a new modality to the basic treatment; see **Supplementary Material** for a complete list of the papers reviewed).

## RESULTS

For MBT, the search yielded 76 hits. An initial screening eliminated 39 papers because they were limited in their population scope (e.g., children, youth, and cancer survivors), 15 were not meta-analyses, seven were not about anxiety disorders, three were not about MBT, two were limited to online interventions, and four were limited to self-compassion, self-help, or stand-alone interventions. Of the remaining six papers, two were further discarded because they agglomerated heterogeneous outcomes (e.g., “internalizing symptoms”, de Abreu Costa et al., 2019; “negative affectivity”, Schumer et al., 2018), two did not report ES separately for each intervention and/or each control category (Bandelow et al., 2018; Hedman-Lagerlof et al., 2018), and one which had raised many methodological concerns<sup>2</sup> (Singh and Gorey, 2018). Thus, the meta-analysis from Goldberg et al. (2018) was retained for our study (see **Supplementary Material 1**).

For CBT-VRexp, the search yielded 12 hits. Three papers were removed because they were not meta-analyses, three were not about CBT-VRexp or specific to this intervention, one was not specific to anxiety disorders, and one was limited in population scope (children). Of the four remaining papers, the meta-analysis from Carl et al. (2019) was retained for our study, supplemented with the one by Benbow and Anderson (2019) for attrition data.

<sup>2</sup>Singh and Gorey (2018) paper is of special note. It directly addresses our subject matter, yet we have doubts on the robustness of their methodology. While they reported 9 RCT studies directly comparing MBI and CBT, they included 2 studies with the same sample, 2 studies on MBI with added exposure, 1 with CBT as a control group without clear definition (TAU with or without medication, not on the whole sample) and with depressive and/or anxious participants (again without stats on each diagnostic category included), 1 study with a 1-h intervention, and 1 with an error on the ES reported (advantage for MBI that was originally found for CBT). This leaves a total of 3 valid studies, which is more in line with Goldberg et al. (2018) and Hedman-Lagerlof et al. (2018). For these reasons, with opted to exclude this paper and report results from the former.

The remaining two publications were rejected because they were either limited to a specific diagnostic category (SA; Chesham et al., 2018) or focusing on deterioration data (Fernández-Álvarez et al., 2019) (see **Supplementary Material 2**).

For CBT, a total of 96 hits were obtained from our search. Of these, eight papers were rejected because they were not about CBT, 11 were not meta-analyses, 25 were not about anxiety disorders, 31 were limited in their population scope, two were not about treatment efficacy, five were limited to Internet or computer-based intervention, and six were about CBT-VRexp or MBT. Of the eight remaining papers, one was rejected because it was limited in scope (group therapy for PTSD; Schwartz et al., 2019), one was too restrictive on the measure of outcome to allow comparisons (remission rate; Springer et al., 2018), two studied the effect of added interventions to CBT (Bernard et al., 2018; Marker and Norton, 2018), and two were limited to primary care settings without information about specific anxiety disorders (Zhang et al., 2019a,b). Since Barry et al. (2018) did not provide information about RCTs, we favored Carpenter et al. (2018) for our study (see **Supplementary Material 3**).

A summary of the information extracted from published meta-analyses documenting the efficacy of CBT-VRexp and MBT is reported in **Table 1**, with the meta-analysis comparing CBT with *in vivo* exposure to other active control treatments reported as a reference for comparison. For manualized MBT, Goldberg et al. (2018) identified 22 RCTs with clinical samples of anxiety disorders and PTSD, totaling 26 ES. Of those, nine used an active control and only seven compared the efficacy over an EBT, such as CBT with *in vivo* exposure. In comparison, of the 30 RCT studies identified for CBT-VRexp (Carl et al., 2019) totaling 40 ES, twice as many (14) used CBT with *in vivo* exposure as a control intervention.

The nature of clinical samples also differs between the MBT and the CBT-VRexp studies retained in the various meta-analyses. For MBT, eight studies used SA samples, seven for PTSD, five for GAD, 2 for obsessive-compulsive disorder, and three mixed samples. For CBT-VRexp, the bulk of evidence pertains to specific phobias (SP) with 17 studies, followed by 13 for SA. Less frequent evidence was documented for PTSD and PD, with five studies each.

The number of clinical trials reporting follow-up data in the meta-analyses is much smaller and not very different between MBT and CBT-VRexp. For MBT, only two studies fitting the criteria of the meta-analyses documented the long-term efficacy compared to an EBT (CBT with *in vivo* exposure), compared to seven for CBT-VRexp. The average effect size of the comparisons with inactive or active control conditions was consistently lower in MBT compared to that in CBT-VRexp. The attrition rate reported in studies on MBT and CBT-VRexp was very similar. Odds of dropping out of MBT and CBT-VRexp were not different from other EBT.

Overall, the amount of information documenting the efficacy of using CBT-VRexp for anxiety disorders is about twice as much as for MBT. Note that our analysis is about the relative number of evidences; the comparative efficacy has not yet been empirically tested and comparing the pooled ES in **Table 1** may be hazardous. Nevertheless, these observations about available evidence from

the published literature cannot justify the disproportionately larger acceptance and enthusiasm of MBT over CBT-VRexp.

## DISCUSSION

Although CBT with exposure exercises has been considered as the gold-standard treatment for anxiety disorders, researchers and clinicians in mental health have embraced and combined different approaches to overcome some limits of CBT and exposure. In this article, we focused on two forms of CBT, CBT-VRexp, and MBT. Specifically, this study was driven by the wish to document and reflect on the apparent widespread scientific and popular interest and preference in using MBT over the use of CBT-VRexp in the treatment of anxiety disorders. Therefore, the aim of this study was to contrast the bulk of evidences supporting the efficacy, specifically for the treatment of anxiety disorders, of using CBT-VRexp versus using MBT. Faced with the growing hype around mindfulness, among both the general population and the clinicians, our question was: is this hype empirically supported? Reviewing studies gathered in meta-analyses, we found twice as many studies supporting CBT-VRexp over MBT. When looking at comparisons with CBT plus *in vivo* exposure, twice as many studies were published in support of CBT-VRexp (14) compared to MBT (seven). Strength is in the numbers: with more studies, the pooled ES are more robust and less likely to be artificially inflated by publication bias. We can also note that these ES are higher in favor of CBT-VRexp compared to MBT. The available information in meta-analyses reported large pooled ES favoring CBT-VRexp over active control conditions at post-treatment and over inactive control conditions (e.g., waiting list or no treatment) at follow-up. However, the pooled ES favoring MBT over active controls were small at post-treatment and even smaller when compared with inactive controls at follow-up. Overall, while expected changes are still clinically significant for MBT, stronger effects with more empirical support are found for CBT-VRexp for treating anxiety disorders.

However, nuances exist when looking at specific diagnostic categories. SA is by far the most studied diagnosis, both for CBT-VRexp and MBT, with respectively six and four published RCT against EBT<sup>3</sup> in the meta-analyses that we used. For PD and SP, only papers for CBT-VRexp were found. For GAD and PTSD, the studies included in the meta-analyses were only for MBT. Thus, based on the evidences, the relevance of MBT or CBT-VRexp was also carefully considered given the target disorder.

In all cases, the number of studies with follow-up data is low, both in terms of numbers and duration. Not enough data are available to draw conclusions about specific disorders. Nonetheless, clinicians may choose CBT-VRexp over MBT with some confidence for its long-term effect as more studies in the meta-analyses found no difference in the long run with the participants treated with EBT.

Another important limitation in the literature that practitioners should keep in mind is that deterioration data

<sup>3</sup>Pooled effect size for MBT was recalculated from Goldberg et al. (2018) with SA studies only: MBT was found to be significantly less effective than EBT,  $-0.31$  ( $-0.60, -0.03$ ).

**TABLE 1 |** Documenting the relative weight of evidences, based on the number of clinical trials on MBT and CBT-VRexp and on *in vivo* exposure as a reference, and the pooled effect sizes.

	MBT <sup>1</sup>	CBT-VRexp <sup>2,3</sup>	CBT- <i>in vivo</i> exp <sup>4</sup>
Number of published randomized controlled trials	22	30	>40
Total number of ES	26	40	>40
Number of ES with clinical samples	26	39	>40
Number of ES with an inactive control at post-treatment	10	20	N/A
Number ES with an active control at post-treatment	9	6	40
Pooled ES for comparisons with active controls at post-treatment	0.27 (0.06, 0.67)	0.78 (0.25, 1.31)	0.56 (0.44, 0.69)
Number of ES with <i>in vivo</i> exposure	7	14	N/A
Pooled ES for comparisons with inactive controls at the follow-up	-0.07 (-0.63, 0.48)	0.90 (-0.05, 1.85)	N/A
Number of ES with a follow-up	9	11	>22
Attrition rates	17%	16%	24%

Data are based on meta-analyses on anxiety-related disorders. CBT, cognitive-behavioral therapy; EBT evidence-based treatment; *in vivo* exp, real-life exposure; ES, effect sizes; MBT, mindfulness-based therapy; VRexp, virtual reality-based exposure. Goldberg et al. (2018) did not pool PTSD studies with those of anxiety disorders in their analysis, contrary to the other meta-analyses presented here. After contacting the main author, who accepted to share his database, we performed a re-analysis of the data (Mantel-Haenszel odds ratio) with the PTSD and the anxiety disorder studies pooled together. Sources of information: <sup>1</sup>Goldberg et al., 2018, <sup>2</sup>Carl et al., 2019 for all data except for attrition rates. <sup>3</sup>Benbow and Anderson, 2019 for attrition rates. <sup>4</sup>Carpenter et al., 2018 for a meta-analysis including only comparisons with active controls.

are rarely reported for both CBT-VRexp and MBT. This is not surprising as it is a relatively new line of inquiry for clinical efficacy studies. It is also challenging as it requires monitoring of individual data compared to group-level analysis. Yet more papers in the literature report such data for CBT-VRexp (around 40%; Fernández-Álvarez et al., 2019) compared to MBT (around 15%; Wong et al., 2018). Deterioration rates reported in the literature were lower for both patients receiving CBT-VRexp (4.0%) and other forms of treatment (2.8%) compared to the wait-listed control (15%; Fernández-Álvarez et al., 2019), while deterioration was reported for only 1% of the participants in both the MBT and the control groups [although only three studies included samples with an anxiety disorder or PTSD (Wong et al., 2018)]. While practicing meditation could be thought of as relatively harmless, this might not be the case in patients suffering from a diagnosed mental disorder. Also, mindfulness practice can be unpleasant and challenging without causing harm. As suggested by Baer et al. (2019), systematic research is needed to address this question, which would require monitoring individual data like what Fernández-Álvarez et al. (2019) did in their analysis. While more studies reported such data for CBT-VRexp, there is still room for improvement. Adverse effects can come in many forms that are not constantly measured (Fernández-Álvarez et al., 2019), such as cybersickness symptoms (i.e., feelings of nausea, dizziness, and discomfort) when using VR technology. While these symptoms do not deteriorate the condition of the patients, they might hinder their capacity to profit from the intervention.

The results of the present study lead to an interesting question: given the higher frequency of support found for CBT-VRexp over MBT for treating anxiety disorders, why has MBT attracted a much wider scientific and popular interest compared to CBT-VRexp? We have formulated some tentative answers.

One possible answer to our question could be the cost and apparent complexity of using VR technologies. While high-end technologies are costly and thus more suitable for research purposes, head-mounted display systems are increasingly suitable

for the general public. Indeed 3 years ago affordable headsets became available, and now the technology is becoming even more affordable and more immersive<sup>4</sup>. Yet in order to use VR, researchers and clinicians need virtual environments (software) and some hardware, which may be seen as cumbersome and represent additional costs. Thus, applying MBT, which “only” requires training from the health professional, could be seen as more affordable and more attractive than using technologies.

A likely answer could also be that, unfortunately, some professionals do not rely on empirical data to choose their therapeutic interventions but rather on their preferences, the appeal of the model, and the current trends in clinical orientations. Adopting an intuitive thinking style is predictive of more negative attitudes toward EBT requiring exposure and more positive ones toward the adoption of alternative therapeutic interventions (Gaudiano et al., 2011). MBT is part of the current zeitgeist (Michalak and Heidenreich, 2018): psychological stress and its reduction are major concerns in modern societies, and MBT offers a solution appealing for both scientific and spiritual reasons. At the same time, government agencies start to adopt MBT as a first-line treatment for anxiety and depression, such as the National Health Service in the United Kingdom (Mindfulness All Party Parliamentary Group [MAPPG], 2015). Thus, both psychological and sociological factors would influence clinicians in their adoption of MBT in their practice. CBT-VRexp might not have the same appeal.

There are still worries over the use of CBT-VRexp. One which is frequently cited is that VR technology could hinder the therapeutic relationship. This is not the case, as shown by Ngai et al. (2015). In addition, therapists from Lindner et al. (2019) rated “making exposure less stressful” as an advantage of using VR and “patients experiencing the VR environment as too real” as a disadvantage. Endorsements of these items by therapists might be indicative of a misunderstanding of exposure and a reluctance

<sup>4</sup>Retrieved from: <https://www.wired.com/story/oculus-rift-s-vr-headset/> and <http://time.com/4169430/oculus-rift-price-release-date-2016/>



to induce anxiety in one's patients, even for their own benefit. This could lead to the adoption of alternative clinical interventions, such as MBT, not because it would benefit the patient but because it is easier on the therapist. In fact, novice therapists tend to have comparable stress levels to their patients, both subjectively and physiologically, right before engaging in exposure compared to a control therapy session (Schumacher et al., 2014). Also, novice therapists experimentally led by researchers to hold negative beliefs about exposure did deliver the treatment sub-optimally by being overly cautious compared to those in the positive beliefs condition (Farrell et al., 2013). Many therapists deliver exposure at lower doses (i.e., for a shorter period of time and in less threatening situation) and in conjunction with controlled breathing strategies without clear empirical evidence of its added clinical efficacy (Deacon et al., 2013). In sum, CBT-VRexp may be less attractive essentially because it is a form of exposure, and therapists may not like to do exposure with their patients. Achieving suboptimal results with their patient could reinforce their negative beliefs about exposure and further justify the choice of alternative interventions. Finally, if therapists avoid using exposure because of fear of patients' discomfort, they may also favor MBT in ways that reduce its usefulness to develop new associations with lack of threat and that reinforce avoidance. When facing the avoided stimuli, MBT can be used constructively to foster exposure (e.g., "Let us be fully aware that there is a live spider crawling on the table, that it is disgusting and looking at you, and embrace the situation and how you can remain in it") or less constructively to foster avoidance (e.g., "Although there is a spider here, let us focus on your respiration and your body, let go of your worries about the spider, and pace your breathing to slowly calm down"). Doing mindful breathing exercises in conjunction with CBT-VRexp could diminish its effectiveness by reducing the anxiety of the patient, thus acting as a form of avoidance. Exposure requires the patient to experience the threatening stimuli to learn through experience that it is safe. To do so, the therapists must tolerate the idea of being "responsible" of "inflicting" anxiety to their patients, which can prompt them to seek out alternative interventions or to supplement them with tempering ones. There is a thin line that is very easy to cross in favor of helping patients develop avoidance behaviors that will either be detrimental in the long term or will need to always be used by patients as safety seeking behaviors or neutralization in order to cope.

## Future Directions

First, researchers should begin gathering data on the fit between patients and treatment modalities. While efficacy studies are important, their focus is at the group level. While we can show that two treatments are equally effective in the treatment of a disorder for two randomly selected individuals, nothing tells us that they will respond equally well to both treatments. For example, some people have a hard time feeling immersed in a VE (e.g., they have cybersickness symptoms), thus not responding to the stimuli used in the exposure therapy. Others do not adhere to the mindfulness philosophy and will not meditate or practice acceptance at home. To better inform professionals about choosing therapeutic approaches, future research should

include measures of potential predictors of treatment efficacy. With such information, clinicians could take the best decision for their patient based not only on what science tells them that is effective but also on why it works that way for some people and not for others.

To date, research on CBT-VRexp has mostly been about replicating *in virtuo* what can normally be done *in vivo*. Doing so, researchers were able to demonstrate that VR is useful to do exposure with hard-to-access stimuli in a controlled and secure environment. Thus, their line of scientific inquiry was mostly focused on phobias. As a result, VR might have been less attractive to professionals. The field of VR is now addressing the more complex anxiety disorders to provide solutions for patients that may be more frequent in therapists' caseloads (e.g., OCD, PTSD, GAD). Still we feel that this is a missed opportunity. VR could be used to improve exposure therapy by pushing its limits way further than what can be done *in vivo*, thus allowing to build stronger associations with lack of threat than what can be done *in vivo*. For example, VR is not limited to visual and auditory stimuli. Studies have successfully integrated olfactory (Baus and Bouchard, 2014), haptic (Tremblay et al., 2016), and thermal (Shaw et al., 2019) stimuli to VE. This could potentially lead to a multisensory exposure therapy for PTSD, bringing recollection of the traumatic experience and reprocessing to a new level. VR can also be used to expose patients to situations hard or impossible to do *in vivo*. For acrophobia, a therapist using CBT-VRexp could ask his/her patient to dance on the edge of a virtual cliff, test his/her balance, and actually jump at will over the cliff to confront his/her fear of falling. For social phobia, it is possible in VR to ask people on dates or set up social blunders that would be delicate to do with actual people.

Stand-alone MBI, as opposed to full MBT treatment programs, are of special note here as they are the most commonly reported form of mindfulness intervention used by clinicians (Michalak et al., 2020). This is not surprising and probably not limited to MBT, with less than 2% of psychotherapists reporting adopting only one practice orientation (Cook et al., 2010). For our review on studies with anxiety disorders, our comparisons were with studies using some form of MBI. The meta-analysis from Blanck et al. (2018), which was not used in our study because it had less RCT than that of Goldberg et al. (2018), focused on studies where only MBI were used as a stand-alone treatment. Of the 21 studies identified by Blanck et al. (2018), only five were RCT. None compared the efficacy of MBI as a stand-alone treatment to an EBT and, most importantly, none used clinical samples. No data were available for long-term effect nor adverse effects. Moreover, no study tested the impact of integrating MBI to other validated intervention protocols. This lack of empirical support could be problematic if the intervention in the integrated stand-alone treatment does not fully address the therapeutic goals of a full MBT program and does not include adequate exposure strategies. The question remains: does the professional choose a treatment strategy to avoid discomfort in their patients? Given that the therapist's experiential avoidance is a significant negative predictor of choosing exposure as a treatment option (Scherr et al., 2015), the question deserves an empirical answer.



Among the limits of the current paper, the first that comes to mind is the reliance on meta-analyses. Meta-analyses are imprecise and limited by design. Variations in inclusion criteria, search terms, and search engines can yield important differences in the result. Given the publication process, papers published in 2018 or 2019 has a coverage running up to 2017, thus important articles could have been left out of this review. The goal of this review was not to be exhaustive but to contrast the state of the literature. It would be quite surprising if the gap in the number of RCT between MBT and CBT-VRexp had been filled in the last year. Conducting our own systematic search of the literature, including unpublished reports and theses, may have provided with more precise numbers, but the ratio of evidences would have remained in the same range.

Another problem of reviewing and comparing meta-analyses is that we had no control on how the results were reported, namely, which parameters were used. For example, we were unable to report on the methodological quality evaluation of the studies as different indexes were used across meta-analyses or simply not reported. Most biases toward CBT-VRexp and MBT are the same and those found in the literature on clinical efficacy: selective reporting, small sample sizes, no intent-to-threat analyses, no deterioration analyses, or adverse effects reporting.

## CONCLUSION

The objective of this paper was to contrast, specifically for the treatment of anxiety disorders, the weight of evidences supporting the use of exposure in VR versus the use of MBT. Overall, the results of the comparisons have shown that CBT with exposure conducted in VR was more thoroughly researched and

supported than MBT. Nevertheless, this conclusion is nuanced by reviewing several gaps in the literature for both therapies, and much more research is required to establish which therapies for the treatment of anxiety disorders are suitable, how they should be carried out, and for whom.

## AUTHOR CONTRIBUTIONS

KN, GC, and SB designed the study. KN and GC collected and analyzed the data and drafted the manuscript. SB facilitated the study execution and aided in the interpretation of findings. All the authors critically reviewed the draft and made significant contributions to the final version.

## FUNDING

This work was supported by the Canada Research Chairs (#950-231039) awarded to SB and a FRQSC postdoctoral research grant (#255365) awarded to KN.

## ACKNOWLEDGMENTS

The authors wish to thank the reviewers for their thoughtful comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2020.00086/full#supplementary-material>

## REFERENCES

- Abramowitz, J. S., Deacon, B. J., and Whiteside, S. P. H. (2011). *Exposure Therapy for Anxiety: Principles and Practice*. New York, NY: Guilford Press.
- Baer, R., Crane, C., Miller, E., and Kuyken, W. (2019). Doing no harm in mindfulness-based programs: conceptual issues and empirical findings. *Clin. Psychol. Rev.* 71, 101–114. doi: 10.1016/j.cpr.2019.01.001
- Baer, R. A. (2003). Mindfulness training as a clinical intervention: a conceptual and empirical review. *Clin. Psychol. Sci. Pr.* 10, 125–143. doi: 10.1093/clipsy.bpg015
- Baer, R. A., and Sauer, S. (2009). Mindfulness and cognitive behavioral therapy: a commentary on harrington and pickles. *J. Cogn. Psychother.* 23, 324–332. doi: 10.1891/0889-8391.23.4.324
- Bandelow, B., and Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialog. Clin. Neuro.* 17:327.
- Bandelow, B., Sagebiel, A., Belz, M., Gorlich, Y., Michaelis, S., and Wedekind, D. (2018). Enduring effects of psychological treatments for anxiety disorders: meta-analysis of follow-up studies. *Br. J. Psychiatry* 212, 333–338. doi: 10.1192/bjp.2018.49
- Barry, T. J., Yeung, S. P., and Lau, J. Y. F. (2018). Meta-analysis of the influence of age on symptom change following cognitive-behavioral treatment for anxiety disorders. *J. Adolesc.* 68, 232–241. doi: 10.1016/j.adolescence.2018.08.008
- Baus, O., and Bouchard, S. (2014). Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. *Front. Hum. Neurosci.* 8:112. doi: 10.3389/fnhum.2014.00112
- Benbow, A. A., and Anderson, P. L. (2019). A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders. *J. Anxiety Disord.* 61, 18–26. doi: 10.1016/j.janxdis.2018.06.006
- Bernard, P., Romain, A. J., Caudroit, J., Chevance, G., Carayol, M., Gourlan, M., et al. (2018). Cognitive behavior therapy combined with exercise for adults with chronic diseases: systematic review and meta-analysis. *Health Psychol.* 37, 433–450. doi: 10.1037/hea0000578
- Bioeca, F. (1997). The cyborg's dilemma: progressive embodiment in virtual environments. *J. Comput. Mediat. Commun.* 3:JCMC324. doi: 10.1111/j.1083-6101.1997.tb00070.x
- Bishop, S. R. (2002). What do we really know about mindfulness-based stress reduction? *Psychosom. Med.* 64, 71–83. doi: 10.1097/00006842-200201000-00010
- Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, K., et al. (2004). Mindfulness: a proposed operational definition. *Clin. Psychol. Sci. Pr.* 11, 230–241. doi: 10.1093/clipsy.bph077
- Blanck, P., Perleth, S., Heidenreich, T., Kroger, P., Ditzgen, B., Bents, H., et al. (2018). Effects of mindfulness exercises as stand-alone intervention on symptoms of anxiety and depression: systematic review and meta-analysis. *Behav. Res. Ther.* 102, 25–35. doi: 10.1016/j.brat.2017.12.002
- Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., and Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Curr. Psychiatry Rep.* 19:42. doi: 10.1007/s11920-017-0788-4

- Botella, C., García-Palacios, A., Villa, H., Baños, R. M., Quero, S., Alcañiz, M., et al. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: a controlled study. *Clin. Psychol. Psychother.* 14, 164–175. doi: 10.1002/cpp.524
- Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., et al. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *Br. J. Psychiatry* 210, 276–283. doi: 10.1192/bjp.bp.116.184234
- Bouchard, S., and Rizzo, A. A. (2019). *Virtual Reality for Psychological and Neurocognitive Interventions*. Berlin: Springer.
- Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., et al. (2019). Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. *J. Anxiety Disord.* 61, 27–36. doi: 10.1016/j.janxdis.2018.08.003
- Carpenter, J. K., Andrews, L. A., Witcraft, S. M., Powers, M. B., Smits, J. A. J., and Hofmann, S. G. (2018). Cognitive behavioral therapy for anxiety and related disorders: a meta-analysis of randomized placebo-controlled trials. *Depress. Anxiety* 35, 502–514. doi: 10.1002/da.22728
- Chambliss, D. L., and Ollendick, T. H. (2001). Empirically supported psychological interventions: controversies and evidence. *Annu. Rev. Psychol.* 52, 685–716. doi: 10.1146/annurev.psych.52.1.685
- Chesham, R. K., Malouf, J. M., and Schutte, N. S. (2018). Meta-analysis of the efficacy of virtual reality exposure therapy for social anxiety. *Behav. Chang.* 35, 152–166. doi: 10.1017/bec.2018.15
- Cipresso, P., Chicci Giglioli, I. A., Alcañiz Raya, M., and Riva, G. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front. Psychol.* 9:2086. doi: 10.3389/fpsyg.2018.02086
- Cook, J. M., Biyanova, T., Elhai, J., Schnurr, P. P., and Coyne, J. C. (2010). What do psychotherapists really do in practice? An Internet study of over 2,000 practitioners. *Psychotherapy* 47, 260–267. doi: 10.1037/a0019788
- Corno, G., and Bouchard, S. (2015). “An innovative positive psychology VR application for victims of sexual violence: a qualitative study,” in *Psychology resEarch Progress. Virtual Reality: Technologies, Medical Applications And Challenges*, eds P. Cipresso and S. Serino (Hauppauge, NY: Nova Science Publishers), 229–267.
- Corno, G., Serino, S., Cipresso, P., Baños, R. M., and Riva, G. (2018). Assessing the relationship between attitudinal and perceptual component of body image disturbance using virtual reality. *Cyberpsychol. Behav. Soc. Netw.* 21, 679–686. doi: 10.1089/cyber.2018.0340
- Côté, S., and Bouchard, S. (2008). Virtual reality exposure for phobias: a critical review. *J. Cyber Ther. Rehabil.* 1, 75–91.
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., and Vervliet, B. (2014). Maximizing exposure therapy: an inhibitory learning approach. *Behav. Res. Ther.* 58, 10–23. doi: 10.1016/j.brat.2014.04.006
- da Costa, R. M., and de Carvalho, L. A. V. (2004). The acceptance of virtual reality devices for cognitive rehabilitation: a report of positive results with schizophrenia. *Comput. Methods Program. Biomed.* 73, 173–182. doi: 10.1016/S0169-2607(03)00066-X
- David, D., Cristea, I., and Hofmann, S. G. (2018). Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Front Psychiatry* 9:4. doi: 10.3389/fpsyg.2018.00004
- de Abreu Costa, M., de Oliveira, G. S. D. A., Tatton-Ramos, T., Manfro, G. G., and Salum, G. A. (2019). Anxiety and stress-related disorders and mindfulness-based interventions: a systematic review and multilevel meta-analysis and meta-regression of multiple outcomes. *Mindfulness* 10, 996–1005. doi: 10.1007/s12671-018-1058-1
- Deacon, B. J., Lickel, J. J., Farrell, N. R., Kemp, J. J., and Hipol, L. J. (2013). Therapist perceptions and delivery of interoceptive exposure for panic disorder. *J. Anxiety Disord.* 27, 259–264. doi: 10.1016/j.janxdis.2013.02.004
- Eisendrath, S. J., Delucchi, K., Bitner, R., Fenimore, P., Smit, M., and McLane, M. (2008). Mindfulness-based cognitive therapy for treatment-resistant depression: a pilot study. *Psychother. Psychosom.* 77, 319–320. doi: 10.1159/000142525
- Evans, S., Ferrando, S., Findler, M., Stowell, C., Smart, C., and Haglin, D. (2008). Mindfulness-based cognitive therapy for generalized anxiety disorder. *J. Anxiety Disord.* 22, 716–721. doi: 10.1016/j.janxdis.2007.07.005
- Farias, M., Wikholm, C., and Delmonte, R. (2016). What is mindfulness-based therapy good for? *Lancet Psychiatry* 3, 1012–1013. doi: 10.1016/s2215-0366(16)30211-5
- Farrell, N. R., Deacon, B. J., Kemp, J. J., Dixon, L. J., and Sy, J. T. (2013). Do negative beliefs about exposure therapy cause its suboptimal delivery? An experimental investigation. *J. Anxiety Disord.* 27, 763–771. doi: 10.1016/j.janxdis.2013.03.007
- Fernández-Álvarez, J., Rozental, A., Carlbring, P., Colombo, D., Riva, G., Anderson, P. L., et al. (2019). Deterioration rates in virtual reality therapy: an individual patient data level meta-analysis. *J. Anxiety Disord.* 61, 3–17. doi: 10.1016/j.janxdis.2018.06.005
- Ferrer-Garcia, M., Gutiérrez-Maldonado, J., and Riva, G. (2013). Virtual reality based treatments in eating disorders and obesity: a review. *J. Contemp. Psychother.* 43, 207–221. doi: 10.1007/s10879-013-9240-1
- Freeman, D. (2008). Studying and treating schizophrenia using virtual reality: a new paradigm. *Schizophr. Bull.* 34, 605–610. doi: 10.1093/schbul/sbn020
- Fuchs, C. (2011). “Cognitive capitalism or informational capitalism? The role of class in the information economy,” in *Cognitive Capitalism, Education And Digital Labor*, eds M. Peters and E. Bulut (New York, NY: Peter Lang), 75–119.
- Gaggioli, A., Pallavicini, F., Morganti, L., Serino, S., Scaratti, C., Briguglio, M., et al. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: a block randomized controlled trial. *J. Med. Internet. Res.* 16:e167. doi: 10.2196/jmir.3235
- García-Palacios, A., Hoffman, H., Carlin, A., Furness Iii, T. A., and Botella, C. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behav. Res. Ther.* 40, 983–993. doi: 10.1016/S0005-7967(01)00068-7
- Gaudiano, B. A., Brown, L. A., and Miller, I. W. (2011). Let your intuition be your guide? Individual differences in the evidence-based practice attitudes of psychotherapists. *J. Eval. Clin. Pract.* 17, 628–634. doi: 10.1111/j.1365-2753.2010.01508.x
- Goldberg, S. B., Tucker, R. P., Greene, P. A., Davidson, R. J., Wampold, B. E., Kearney, D. J., et al. (2018). Mindfulness-based interventions for psychiatric disorders: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 59, 52–60. doi: 10.1016/j.cpr.2017.10.011
- Gonçalves, R., Pedrozo, A. L., Coutinho, E. S. F., Figueira, I., and Ventura, P. (2012). Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. *PLoS One* 7:e48469. doi: 10.1371/journal.pone.0048469
- Grossman, P. (2019). On the porosity of subject and object in “mindfulness” scientific study: challenges to “scientific” construction, operationalization and measurement of mindfulness. *Curr. Opin. Psychol.* 28, 102–107. doi: 10.1016/j.copsyc.2018.11.008
- Harrington, N., and Pickles, C. (2009). Mindfulness and cognitive behavioral therapy: are they compatible concepts? *J. Cogn. Psychother.* 23, 315–323. doi: 10.1891/0889-8391.23.4.315
- Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behav. Ther.* 35, 639–665. doi: 10.1016/S0005-7894(04)80013-3
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., and Lillis, J. (2006). Acceptance and commitment therapy: model, processes and outcomes. *Behav. Res. Ther.* 44, 1–25. doi: 10.1016/j.brat.2005.06.006
- Hedman-Lagerlof, M., Hedman-Lagerlof, E., and Ost, L. G. (2018). The empirical support for mindfulness-based interventions for common psychiatric disorders: a systematic review and meta-analysis. *Psychol. Med.* 48, 2116–2129. doi: 10.1017/S0033291718000259
- Hembree, E. A., and Cahill, S. P. (2007). “Obstacles to successful implementation of exposure therapy,” in *Handbook of Exposure Therapies*, eds D. C. David, R. Lauterbach, and D. Lauterbach (Burlington, MA: Academic Press), 389–408. doi: 10.1016/B978-012587421-2/50018-1
- Hodges, L. F., Kooper, R., Meyer, T. C., Rothbaum, B. O., Opdyke, D., De Graaff, J. J., et al. (1995). Virtual environments for treating the fear of heights. *Computer* 28, 27–34. doi: 10.1109/2.391038
- Hofmann, S. G., Sawyer, A. T., Witt, A. A., and Oh, D. (2010). The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. *J. Consult. Clin. Psychol.* 78, 169–183. doi: 10.1037/a0018555
- Hofmann, S. G., and Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. *J. Clin. Psychiatry* 69, 621–632. doi: 10.4088/jcp.v69n0415

- Holden, M. K. (2005). Virtual environments for motor rehabilitation. *Cyberpsychol. Behav.* 8, 187–211. doi: 10.1089/cpb.2005.8.187
- Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. *Gen. Hosp. Psychiatry* 4, 33–47. doi: 10.1016/0163-8343(82)90026-3
- Kabat-Zinn, J. (1994). *Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life*. New York, NY: Hyperion.
- Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: past, present, and future. *Clin. Psychol. Sci. Pr.* 10, 144–156. doi: 10.1093/clipsy.bpg016
- Katzman, M. A., Bleau, P., Blier, P., Chokka, P., Kjernisted, K., and Van Ameringen, M. (2014). Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. *BMC Psychiatry* 14:S1. doi: 10.1186/1471-244X-14-S1-S1
- Kenny, M. A., and Williams, J. M. G. (2007). Treatment-resistant depressed patients show a good response to mindfulness-based cognitive therapy. *Behav. Res. Ther.* 45, 617–625. doi: 10.1016/j.brat.2006.04.008
- Kim, S. I., Ku, J., Han, K., Lee, H., Park, J., Kim, J. J., et al. (2008). Virtual reality applications for patients with schizophrenia. *J. Cyber. Ther. Rehabil.* 1, 101–112.
- Kim, Y. W., Lee, S. H., Choi, T. K., Suh, S. Y., Kim, B., Kim, C. M., et al. (2009). Effectiveness of mindfulness-based cognitive therapy as an adjuvant to pharmacotherapy in patients with panic disorder or generalized anxiety disorder. *Depress. Anxiety* 26, 601–606. doi: 10.1002/da.20552
- Kumar, S., Feldman, G., and Hayes, A. (2008). Changes in mindfulness and emotion regulation in an exposure-based cognitive therapy for depression. *Cognit. Ther. Res.* 32:734. doi: 10.1007/s10608-008-9190-1
- Lam, Y. S., Man, D. W., Tam, S. F., and Weiss, P. L. (2006). Virtual reality training for stroke rehabilitation. *Neuro Rehabil.* 21, 245–253.
- Lindner, P., Miloff, A., Zetterlund, E., Reuterskiöld, L., Andersson, G., and Carlbring, P. (2019). Attitudes toward and familiarity with virtual reality therapy among practicing cognitive behavior therapists: a cross-sectional survey study in the era of consumer VR platforms. *Front. Psychol.* 10:176. doi: 10.3389/fpsyg.2019.00176
- Lombard, M., and Ditton, T. (1997). At the heart of it all: the concept of presence. *J. Comput. Mediat. Commun.* 3:JCMC321. doi: 10.1111/j.1083-6101.1997.tb00072.x
- Marker, I., and Norton, P. J. (2018). The efficacy of incorporating motivational interviewing to cognitive behavior therapy for anxiety disorders: a review and meta-analysis. *Clin. Psychol. Rev.* 62, 1–10. doi: 10.1016/j.cpr.2018.04.004
- McLay, R. N., Baird, A., Webb-Murphy, J., Deal, W., Tran, L., Anson, H., et al. (2017). A randomized, head-to-head study of virtual reality exposure therapy for posttraumatic stress disorder. *Cyberpsychol. Behav. Soc. Netw.* 20, 218–224. doi: 10.1089/cyber.2016.0554
- Michalak, J., and Heidenreich, T. (2018). Dissemination before evidence? What are the driving forces behind the dissemination of mindfulness-based interventions? *Clin. Psychol. Sci. Pr.* 25:e12254. doi: 10.1111/p.12254
- Michalak, J., Steinhaus, K., and Heidenreich, T. (2020). (How) Do therapists use mindfulness in their clinical work? A study on the implementation of mindfulness interventions. *Mindfulness* 11, 401–410. doi: 10.1007/s12671-018-0929-9
- Mindfulness All Party Parliamentary Group [MAPPG] (2015). *Mindful Nation UK. The Mindfulness Initiative*. Available online at: <https://www.themindfulnessinitiative.org/mindful-nation-report> (accessed April 7, 2019).
- Nathan, P. E., and Gorman, J. M. (2015). *A Guide to Treatments That Work*. New York, NY: Oxford University Press.
- National Institute for Health and Care Excellence (2014). *Anxiety Disorders (NICE Quality Standard No. 53)*. Available online at: <https://www.nice.org.uk/guidance/qs53>
- Ngai, I., Tully, E. C., and Anderson, P. L. (2015). The course of the working alliance during virtual reality and exposure group therapy for social anxiety disorder. *Behav. Cogn. Psychother.* 43, 167–181. doi: 10.1017/S135246581300088X
- Otte, C. (2011). Cognitive behavioral therapy in anxiety disorders: current state of the evidence. *Dialog. Clin. Neurosci.* 13, 413–421.
- Parsons, T. D., and Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *J. Behav. Ther. Exp. Psychiatry* 39, 250–261. doi: 10.1016/j.jbtep.2007.07.007
- Pittig, A., and Hoyer, J. (2018). Exposition aus sicht niedergelassener verhaltenstherapeutinnen und verhaltenstherapeuten. *Z. Klin. Psychol. Psychopathol. Psychother.* 46, 223–235. doi: 10.1026/1616-3443/a000441
- Pittig, A., Kotter, R., and Hoyer, J. (2019). The struggle of behavioral therapists with exposure: self-reported practicability, negative beliefs, and therapist distress about exposure-based interventions. *Behav. Ther.* 50, 353–366. doi: 10.1016/j.beth.2018.07.003
- Repetto, C., Gaggioli, A., Pallavicini, F., Cipresso, P., Raspelli, S., and Riva, G. (2013). Virtual reality and mobile phones in the treatment of generalized anxiety disorders: a phase-2 clinical trial. *Pers. Ubiquit. Comput.* 17, 253–260. doi: 10.1007/s00779-011-0467-0
- Rizzo, A. A., and Buckwalter, J. G. (1997). “Virtual reality and cognitive assessment,” in *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation*, ed. G. Riva (Amsterdam: IOS Press), 123.
- Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., et al. (2000). The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *Cyberpsychol. Behav.* 3, 483–499. doi: 10.1089/10949310050078940
- Scherr, S. R., Herbert, J. D., and Forman, E. M. (2015). The role of therapist experiential avoidance in predicting therapist preference for exposure treatment for OCD. *J. Contextual Behav. Sci.* 4, 21–29. doi: 10.1016/j.jcbs.2014.12.002
- Schultheis, M. T., Himmelstein, J., and Rizzo, A. A. (2002). Virtual reality and neuropsychology: upgrading the current tools. *J. Head Trauma Rehabil.* 17, 378–394. doi: 10.1097/00001199-200210000-00002
- Schultheis, M. T., and Rizzo, A. A. (2001). The application of virtual reality technology in rehabilitation. *Rehabil. Psychol.* 46, 296–311. doi: 10.1037/0090-5550.46.3.296
- Schumacher, M., Gaudlitz, K., Plag, J., Miller, R., Kirschbaum, C., Fehm, L., et al. (2014). Who is stressed? A pilot study of salivary cortisol and alpha-amylase concentrations in agoraphobic patients and their novice therapists undergoing in vivo exposure. *Psychoneuroendocrinology* 49, 280–289. doi: 10.1016/j.psyneuen.2014.07.016
- Schumer, M. C., Lindsay, E. K., and Creswell, J. D. (2018). Brief mindfulness training for negative affectivity: a systematic review and meta-analysis. *J. Consult. Clin. Psychol.* 86, 569–583. doi: 10.1037/ccp0000324
- Schwartz, D., Barkowski, S., Strauss, B., Knaevelsrud, C., and Rosendahl, J. (2019). Efficacy of group psychotherapy for post-traumatic stress disorder: systematic review and meta-analysis of randomized controlled trials. *Psychother. Res.* 29, 415–431. doi: 10.1080/10503307.2017.1405168
- Segal, Z. V., Bieling, P., Young, T., MacQueen, G., Cooke, R., Martin, L., et al. (2010). Antidepressant monotherapy vs sequential pharmacotherapy and mindfulness-based cognitive therapy, or placebo, for relapse prophylaxis in recurrent depression. *Arch. Gen. Psychiatry* 67, 1256–1264. doi: 10.1001/archgenpsychiatry.2010.168
- Segal, Z. V., Williams, J. M. G., and Teasdale, J. D. (2002). *Mindfulness-Based Cognitive Therapy for Depression: A New Approach To Preventing Relapse*. New York, NY: Guilford Press.
- Shaw, E., Roper, T., Nilsson, T., Lawson, G., Cobb, S. V., and Miller, D. (2019). “The heat is on: exploring user behavior in a multisensory virtual environment for fire evacuation,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, (Glasgow, NY: ACM).
- Singh, S. K., and Gorey, K. M. (2018). Relative effectiveness of mindfulness and cognitive behavioral interventions for anxiety disorders: meta-analytic review. *Soc. Work Ment. Health* 16, 238–251. doi: 10.1080/15332985.2017.1373266
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philos. Trans. Royal Soc. B.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138
- Springer, K. S., Levy, H. C., and Tolin, D. F. (2018). Remission in CBT for adult anxiety disorders: a meta-analysis. *Clin. Psychol. Rev.* 61, 1–8. doi: 10.1016/j.cpr.2018.03.002
- Tarr, M. J., and Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nat. Neurosci.* 5, 1089–1092. doi: 10.1038/nn948
- Tremblay, L., Roy-Vaillancourt, M., Chebbi, B., Bouchard, S., Daoust, M., Dénommée, J., et al. (2016). Body image and anti-fat attitudes: an experimental study using a haptic virtual reality environment to replicate human touch. *Cyberpsychol. Behav. Soc. Netw.* 19, 100–106. doi: 10.1089/cyber.2015.0226

- Wiederhold, B. K., and Bouchard, S. (2014). *Advances in Virtual Reality And Anxiety Disorders*. Boston: Springer, doi: 10.1007/978-1-4899-8023-6
- Wielgosz, J., Goldberg, S. B., Kral, T. R., Dunne, J. D., and Davidson, R. J. (2019). Mindfulness meditation and psychopathology. *Annu. Rev. Clin. Psychol.* 15, 285–316. doi: 10.1146/annurev-clinpsy-021815-093423
- Wong, S. Y. S., Chan, J. Y. C., Zhang, D., Lee, E. K. P., and Tsoi, K. K. F. (2018). The safety of mindfulness-based interventions: a systematic review of randomized controlled trials. *Mindfulness* 9, 1344–1357. doi: 10.1007/s12671-018-0897-0
- Zhang, A., Bornheimer, L. A., Weaver, A., Franklin, C., Hai, A. H., Guz, S., et al. (2019a). Cognitive behavioral therapy for primary care depression and anxiety: a secondary meta-analytic review using robust variance estimation in meta-regression. *J. Behav. Med.* 42, 1117–1141. doi: 10.1007/s10865-019-00046-z
- Zhang, A., Franklin, C., Jing, S., Bornheimer, L. A., Hai, A. H., Himle, J. A., et al. (2019b). The effectiveness of four empirically supported psychotherapies for primary care depression and anxiety: a systematic review and meta-analysis. *J. Affect. Disord.* 245, 1168–1186. doi: 10.1016/j.jad.2018.12.008

**Conflict of Interest:** SB is the President of, and own equity in, Cliniques et Développement In Virtuo, a spin-off company from the university that uses virtual reality and distributes virtual environments designed for the treatment of mental disorders. The terms of these arrangements have been reviewed and approved by the Université du Québec en Outaouais in accordance with its conflict of interest policies.

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

Copyright © 2020 Nolet, Corno and Bouchard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis

Dimitrios Saredakis<sup>1\*</sup>, Ancret Szpak<sup>1†</sup>, Brandon Birkhead<sup>2</sup>, Hannah A. D. Keage<sup>1</sup>, Albert Rizzo<sup>3</sup> and Tobias Loetscher<sup>1</sup>

<sup>1</sup> Cognitive Ageing and Impairment Neurosciences Laboratory, School of Psychology, Social Work and Social Policy, University of South Australia, Adelaide, SA, Australia, <sup>2</sup> Division of Health Services Research, Department of Medicine, Cedars-Sinai Health System, Los Angeles, CA, United States, <sup>3</sup> Institute for Creative Technologies, University of Southern California, Los Angeles, CA, United States

## OPEN ACCESS

### Edited by:

Pietro Cipresso,  
Italian Auxological Institute  
(IRCCS), Italy

### Reviewed by:

Bernhard E. Riecke,  
Simon Fraser University, Canada  
Hai-Ning Liang,  
Xi'an Jiaotong-Liverpool  
University, China  
Frederic Merienne,  
ParisTech École Nationale Supérieure  
d'Arts et Métiers, France

### \*Correspondence:

Dimitrios Saredakis  
dimitrios.saredakis@  
mymail.unisa.edu.au

<sup>†</sup> These authors share first authorship

### Specialty section:

This article was submitted to  
Health,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 18 November 2019

**Accepted:** 02 March 2020

**Published:** 31 March 2020

### Citation:

Saredakis D, Szpak A, Birkhead B, Keage HAD, Rizzo A and Loetscher T (2020) Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Front. Hum. Neurosci.* 14:96. doi: 10.3389/fnhum.2020.00096

The use of head-mounted displays (HMD) for virtual reality (VR) application-based purposes including therapy, rehabilitation, and training is increasing. Despite advancements in VR technologies, many users still experience sickness symptoms. VR sickness may be influenced by technological differences within HMDs such as resolution and refresh rate, however, VR content also plays a significant role. The primary objective of this systematic review and meta-analysis was to examine the literature on HMDs that report Simulator Sickness Questionnaire (SSQ) scores to determine the impact of content. User factors associated with VR sickness were also examined. A systematic search was conducted according to PRISMA guidelines. Fifty-five articles met inclusion criteria, representing 3,016 participants (mean age range 19.5–80; 41% female). Findings show gaming content recorded the highest total SSQ mean 34.26 (95%CI 29.57–38.95). VR sickness profiles were also influenced by visual stimulation, locomotion and exposure times. Older samples (mean age  $\geq 35$  years) scored significantly lower total SSQ means than younger samples, however, these findings are based on a small evidence base as a limited number of studies included older users. No sex differences were found. Across all types of content, the pooled total SSQ mean was relatively high 28.00 (95%CI 24.66–31.35) compared with recommended SSQ cut-off scores. These findings are of relevance for informing future research and the application of VR in different contexts.

**Keywords:** cybersickness, simulator sickness, head-mounted display, virtual reality, virtual environment

## INTRODUCTION

Despite advancements in virtual reality (VR) technology, many people still report experiencing simulator sickness symptoms from its use (Rebenitsch and Owen, 2016; Gavvani et al., 2017; Duzmanska et al., 2018; Guna et al., 2019). Characterizing and quantifying these symptoms is challenging, as several factors are at play including a diverse range of technologies; the use of inconsistent terminology for sickness from using virtual environments; little consensus on the biological mechanisms of symptoms; the diverse range of VR content; along with user characteristics such as age and sex (Hale and Stanney, 2014). Identifying factors that increase the

occurrence of simulator sickness becomes necessary with the increased use of VR for rehabilitation, industry training and gaming/entertainment consumers (Gallagher and Ferrè, 2018; Powell et al., 2018; Wang et al., 2018).

Side effects from virtual environment usage has been referred to by many terms including simulator sickness (Kennedy et al., 1993), cybersickness (LaViola, 2000) and VR sickness (Kim et al., 2018). The term simulator sickness originated from the early use of flight simulators in the military (Kennedy et al., 1993), and is still currently used in research using modern HMD technology (Tyrrell et al., 2018; Ziegler et al., 2018). Cybersickness, originally used to describe side effects from use of virtual environments (McCauley and Sharkey, 1992), has often been mentioned in studies using a variety of technologies including flat screen displays and head-mounted displays (HMD) (Rebenitsch and Owen, 2016). The term VR sickness has typically been used in studies using HMDs (Cobb et al., 1999; Kim et al., 2018). Thus, diverse terminology is often used interchangeably across the virtual environments literature.

This current review focuses on adverse symptoms from HMD use, hence the term “VR Sickness” will be referred to as the symptoms (and their severity) typically reported in the literature from HMD use. The term “motion sickness” will be used to refer to more general reporting of symptoms from motion environments (e.g., air, land, or sea travel), not specific to HMDs, where symptoms can differ. For example, nausea can be more severe in seasickness, compared with simulator use (Kennedy et al., 2010). Symptomatology of sickness also differs between technologies. Compared with simulators, HMDs have been reported to produce higher symptoms related to nausea, dizziness and blurred vision (Kennedy et al., 2003).

Measures of VR sickness are a fundamental part of establishing prevalence and symptomatology in virtual environments. The Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993), originally developed for measuring motion sickness in simulators, is the most commonly used measure of sickness in virtual environments (Rebenitsch and Owen, 2016). Alternate measures, such as the *Virtual Reality Symptom Questionnaire*, which was specifically developed for HMDs (Ames et al., 2005) or the *Virtual Reality Sickness Questionnaire* (Kim et al., 2018) have yet to be widely adopted. Single item assessments that are easy to administer and monitor symptoms during VR exposure (Bos et al., 2005; Keshavarz and Hecht, 2011) are commonly used, but do not provide comprehensive measurements of the symptoms of VR sickness. Very few studies report on the use of objective physiological measures (e.g., heart rate, skin conductance, electroencephalograms, eye blink rate, and electrogastrogram) that do not rely on individual self-report data (Kim et al., 2005; Dennison et al., 2016).

Recent advances in HMD technology (field of view, resolution, framerate, and ergonomic factors) have increased the levels of immersion and realism that may have an influence on the occurrence of VR sickness (Nichols, 1999; Lee et al., 2017; Kourtesis et al., 2019). For example, if an image is clear and tracking of movement is accurate, there may be fewer sensory conflicts, and that could lead to a reduction in VR sickness

symptoms (White et al., 2015; Shin et al., 2016; Ray et al., 2018). However, an increase in the field of view may also increase risk of VR sickness (Fernandes and Feiner, 2016). Despite the improvements in HMD technology, a recent review suggests that the prevalence of VR sickness is still problematic (Rebenitsch and Owen, 2016). In addition to this, Kourtesis et al. (2019), in their review found that although recent hardware features have been an important factor in reducing VR sickness, software features also need to be taken into consideration.

The VR content delivered to users can induce or even reduce VR sickness. A rollercoaster ride may be more likely to induce VR sickness to the level of severity where users will request to discontinue the experience. For example, almost 67% of participants in a study using a rollercoaster virtual environment were unable to complete an exposure time of 14 min (Nesbitt et al., 2017). In contrast, content consisting of low amounts of motion may be less likely to induce VR sickness (Guna et al., 2019), as well as in cases where head movement in a fixed position is concordant with what the user would experience in the real world (Rizzo and Koenig, 2017).

Length of time exposed to a virtual environment may also influence likelihood and severity of VR sickness (Duzmanska et al., 2018). Significant correlations have been found between exposure time and VR sickness, with longer exposure times increasing risk of VR sickness (Stanney et al., 2003). For example, research measuring VR sickness at multiple time points found symptoms increased at 2-min increments, with the highest VR sickness scores measured in the final trial at 10 min (Moss and Muth, 2011). In contrast, a recent review has found that some people may build up a resistance or adapt over time to VR sickness, particularly over multiple sessions (Duzmanska et al., 2018). Although content and duration are significant contributing factors that may increase the likelihood of sickness symptoms, the user also needs to be taken into consideration.

User characteristics adds another layer of complexity in understanding the relationship between hardware, content and VR sickness. Research on sex and age, have generated mixed findings when it comes to the likelihood of sickness from VR (Cheung and Hofer, 2002; Benoit et al., 2015; Munafo et al., 2017; Arcioni et al., 2018). In reference to age, physiological differences over the lifespan (i.e., visual, vestibular senses) (Bermúdez Rey et al., 2016) may influence the occurrence of VR sickness and symptom profiles. For example, hormonal differences in females have been reported to influence and likely to be a factor in increased rates of VR sickness (Clemes and Howarth, 2005). Moreover, females can have a smaller interpupillary distance (Fulvio et al., 2019) and some HMDs may not be able to be adjusted accordingly therefore creating eye strain and general discomfort. Thus, it is important to increase the understanding of the relationship between these user characteristics and VR sickness.

Previous reviews (Rebenitsch and Owen, 2016; Duzmanska et al., 2018; Kourtesis et al., 2019) have focused on temporal or technological aspects of VR sickness. To date, none of the reviews on VR sickness have systematically evaluated VR content and user characteristics in a meta-analysis. The primary aim of this systematic review is to examine if VR sickness symptoms

measured with the SSQ using HMDs are influenced by different factors. More specifically, factors that will be examined in this review are content, the amount of visual stimulation (motion of virtual environment), whether a person is stationary or moving in the virtual environment and time. As the SSQ consists of three grouped factors (nausea, oculomotor, and disorientation), a summary of the most common symptoms using HMDs will be provided. Studies with the intention of inducing or not inducing VR sickness will also be compared. A secondary aim is to examine the influence of user characteristics (i.e., age and sex) on SSQ scores and dropout rates.

## METHODS

### Search Strategy

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement (Liberati et al., 2009), a systematic literature search was conducted to reveal journal and conference papers related to VR sickness from using HMDs. This review included the following search terms: virtual reality OR virtual environment\* OR VR OR VR headset OR virtual reality headset OR head-mounted display OR HMD OR helmet mounted display AND cybersickness OR motion sickness OR simulator sickness OR visually induced motion sickness OR virtual reality induced motion sickness OR virtual reality induced symptoms and effects OR virtual reality sickness OR visual-vestibular OR nausea OR aftereffect\* OR after effect\* OR VIMS. No limiters were inserted in the database searches.

This search was carried out on the 10th October 2018 in the six databases: Cochrane Library, IEEEExplore, Medline, PsycINFO, Scopus, and Web of Science. Terms were mapped to subject headings. Both journal and conference articles were *included* in this review if: participants used a head-mounted display (HMD); VR sickness was measured using the SSQ; articles were peer-reviewed and complete (i.e., includes a full paper, not just an abstract or poster presentation); the text was in English or had been translated for publication. Papers were *excluded* if they: used augmented reality (AR) or see-through displays; were reviews, dissertations, abstracts or poster presentations; used prototype HMD devices; and were case studies. Papers that included clinical samples were also excluded, however, if the study included a healthy control group, this data was included. Eligibility of studies was assessed by two independent reviewers (DS and AS).

Papers were included if they supplied mean data for the SSQ (either subscales or total scores), if no mean data was supplied they were still included in the dropout analysis if they indicated drop out rates. If papers supplied mean scores without standard deviations, authors were contacted to supply the standard deviations. Current contact details were searched for online in each case. A follow-up email was sent to authors that did not respond to the initial email. If the authors did not respond to the second email the paper was excluded. The calculation of subscale and total SSQ scores required weighting. Subscales are weighted as follows; nausea 9.54; oculomotor 7.58; and disorientation 13.92, while total scores can be calculated by multiplying unweighted subscale scores by 3.74 (Kennedy et al., 1993). This can create some confusion at times, and there

were instances where researchers calculated the scores differently. For example, multiplying the weighted subscale scores by 3.74 thereby producing inflated total scores. There were also instances where the total SSQ scores did not match the subscale scores, the same contact procedure was followed for these papers as per the missing standard deviations.

**Figure 1** shows the results of the electronic search and article selection as per PRISMA guidelines (Liberati et al., 2009).

### Statistical Approach

Comprehensive Meta-Analysis (CMA) Version 3 (Borenstein et al., 2013) was used to conduct meta-analyses. A random effects model was used to calculate pooled effect estimates with 95% confidence intervals. In studies reporting multiple experiments within groups, these means were merged in CMA to produce one mean per study. In studies reporting multiple experiments between groups, these means were calculated separately for each experiment. Pooled means were calculated for all factors separately on each subscale of the SSQ. Pooled means were also calculated for all factors separately for the total SSQ score. Differences between sub-factors within each factor were assessed using the Q-test based on analysis of variance (Borenstein et al., 2011). The Q-value for the between group analyses corresponded to the weighted sum of squared deviations of the subgroup means about the grand mean. P-values were obtained by comparing the Q-values with a chi-squared distribution with degrees of freedom equal to the number of subgroups minus one (Borenstein et al., 2013). A p-value lower than 0.05 was assumed to indicate a significant statistical difference of SSQ scores between the sub-factors. A correlation was performed between the percentage of females in studies and total SSQ scores as breakdowns for sex of means for the SSQ scores were not supplied in most studies.

### Operationalisation of Factors Being Examined

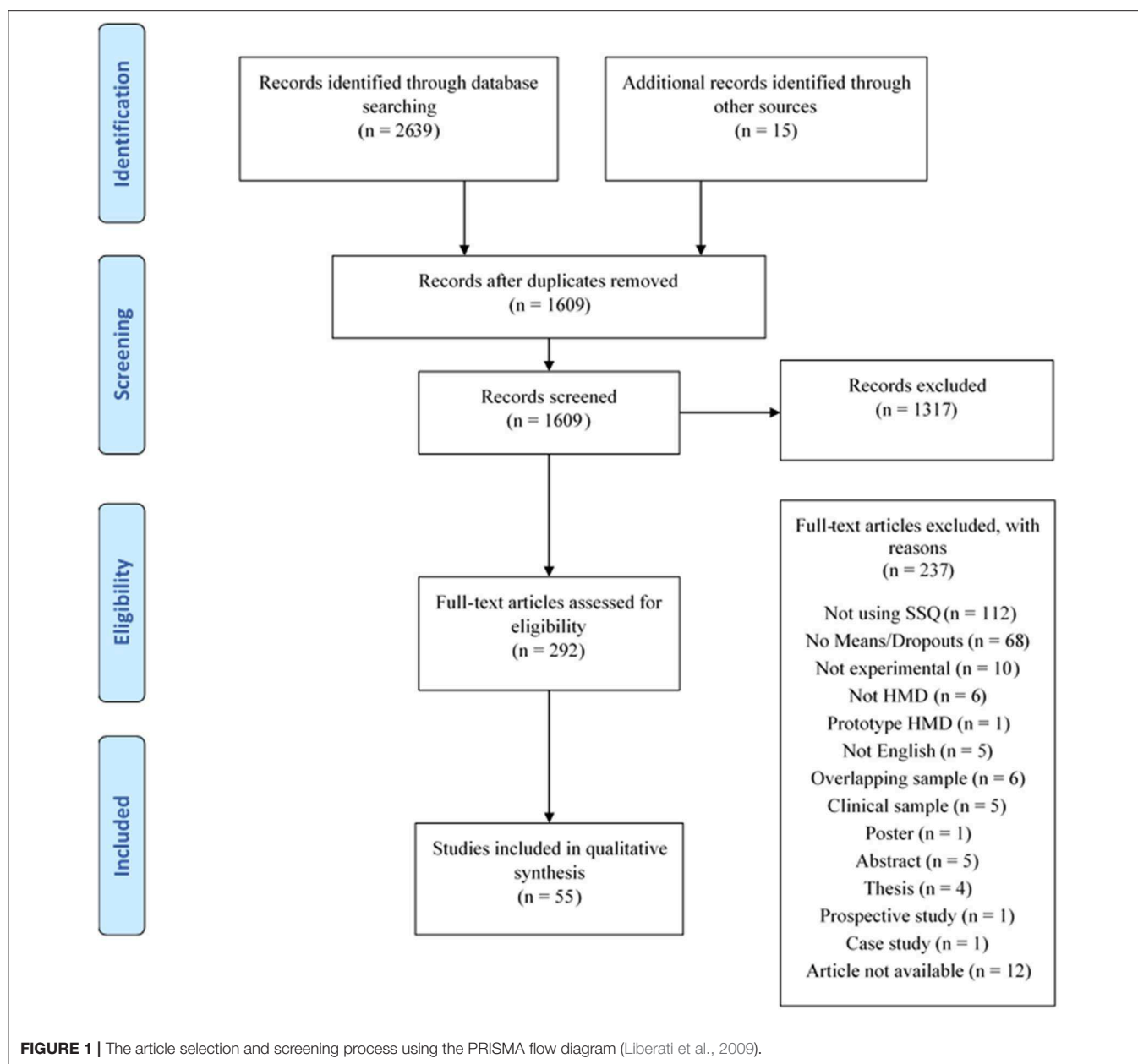
All factors were operationalised and independently reviewed by DS and AS. Any disagreements were resolved by discussion.

### Content

Four types of content were categorized in studies included for analysis; 360 videos; gaming content; minimalist content; and scenic content. User interaction and environmental features differed for each category. The 360 videos included content captured with a 360 camera or video taken that allowed a 360 view of the virtual environment. Gaming included high detailed content where the user could actively interact and perform tasks in the virtual environment including off-the-shelf games and content developed by researchers. Minimalist content consisted of basic shapes or minimal textures, with typically simple interactions. Scenic content included detailed environments, for example, a landscape or cityscape with no or simple interaction by the user. See **Figure 2** for a summary of content characteristics.

### Visual Stimulation

All studies were categorized based on the amount of visual movement within the content regardless of user-directed



movement, such as locomotion and head movement. Low visual stimulation included content with slow visual changes, while high visual stimulation included content with fast visual changes.

## Locomotion

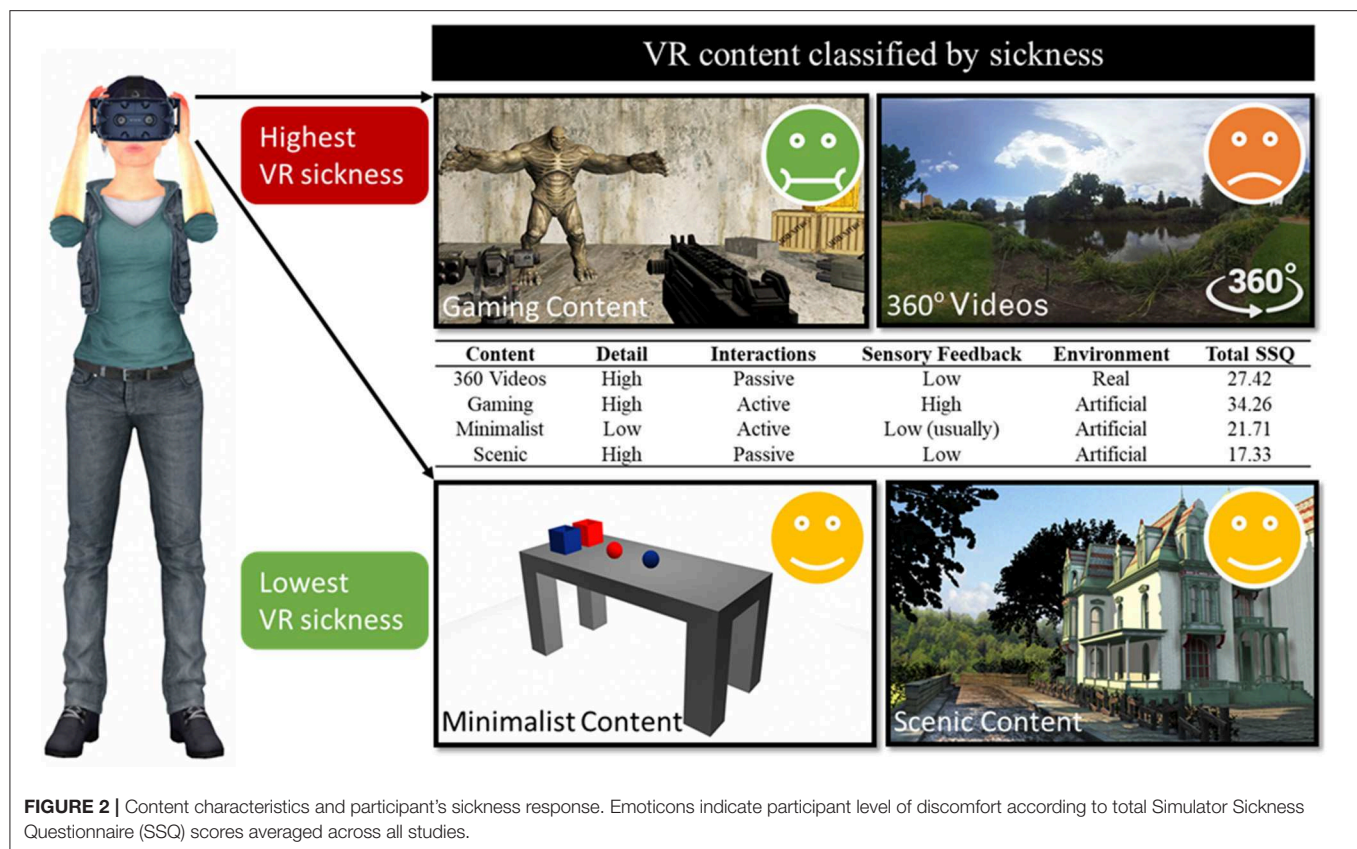
Locomotion refers to how a user navigates in the virtual environment. For the analysis in this review, locomotion was classified as either stationary, controller-based movement, or physically walking. With stationary content, the user does not move in the virtual environment. Two moving categories were included; controller and walking. Controller-based movement included the following navigation methods; flying; controller-based walking; teleporting and driving, therefore

any movement for navigation by the user. Walking included the following physical movements; walking; walking in place and walking on a treadmill. The two categories of moving were used as physically walking has been found to reduce the incidence of VR sickness compared to controller-based navigation (Chance et al., 1998).

## Time

Sickness in virtual environments has been found to increase after 10 min in HMDs and simulator studies (Min et al., 2004; Moss and Muth, 2011). Thus, time was categorized into three intervals of 10 min: <10 min, ≥10 min, or ≥20 min.





## VR Sickness Condition

Studies that explicitly set out to increase/decrease the occurrence of VR sickness or measured VR sickness as a secondary aim, were categorized into two conditions: induce, and not induce.

## User Characteristics

The user characteristic of age was categorized into a mean age of <35 years old and ≥35 years old. This cut-off was used to correspond with theories of both sensory conflict and postural instability. For example, vestibular function involved in the sensory conflict theory starts to decline around the age of 40 (Bermúdez Rey et al., 2016). With relevance to the postural instability theory, changes in altered postural balance have been reported to commence at the ages of 30–39 (Era et al., 2006).

Mean breakdowns by sex were not supplied in most SSQ studies. Therefore, a correlational analysis was performed looking at the proportion of sex (females) in studies with total SSQ mean scores. This approach aimed to give an approximation due to the lack of available data, a positive correlation in this analysis will indicate higher susceptibility of VR sickness in females.

## Dropouts

Dropouts in this review refer to participants that exited an experiment due to VR sickness.

## RESULTS

A total of 2,654 publications were identified through the search. A snowballing strategy was used to identify an additional 15 articles for inclusion. These publications were imported into EndNote where 1,045 duplicates were removed. The remaining 1,609 articles were sent to Covidence systematic review management software (Covidence, 2019) for title and abstract screening, which identified 292 articles for full-text screening. A further 237 articles were excluded as outlined in **Figure 1**. Authors were contacted for 15 papers as per the procedure described in the methods section if mean scores were supplied without standard deviations (10), or if scores did not appear to be weighted correctly (5). A total of 54% of authors replied with 20% supplying raw data to enable calculation of SSQ scores. Hence, 55 publications were identified through the systematic review process and listed in **Table 1**.

## Dropouts

The mean dropout rate reported across 46 experiments due to VR sickness was 15.6%. If studies did not report dropouts, they were not included in this analysis as it was unknown whether there were no instances of dropouts or whether they were just not reported.

**TABLE 1 |** Summary of included articles.

Authors	<i>n</i>	Age mean (SD)	Content	Visual stimulation	Locomotion	Time (minutes)	Dropout %	Condition/ Group	Nausea	(SD)	Oculomotor	(SD)	Disorientation	(SD)	Total	(SD)
Arcioni et al. (2018)	21 (14F, 6M)	28.4 (10.1)	Minimalist	High	Stationary	<10	n/a	Compensated	19.00	(20.00)	19.00	(14.00)	28.00	(37.00)		
							n/a	Inverse	20.00	(18.00)	22.00	(19.00)	31.00	(30.00)		
Bessa et al. (2016)	63 (32F, 31M)	21.49 (3.85)	360	High	Controller	<10	n/a	2D	2.73	(4.42)	10.83	(12.59)	17.90	(19.76)	11.22	(11.41)
							n/a	3D	2.94	(4.50)	9.62	(9.24)	12.31	(17.31)	9.21	(9.82)
Brooks et al. (2017)	26 (3F, 19M)	31.59 (7.72)	360	High	Controller	≥20	4								22.91	(2.31)
Budhiraja et al. (2017)	15 (3F, 12M)	18–26 (range)	Game	High	Controller	≥10	12	Rotation blurring							42.14	(27.61)
	15 (3F, 12M)						n/a	No rotation blurring							51.36	(34.67)
Carnegie and Rhee (2015)	20 (6F, 14M)	18–50 (range)	Scenic	Low	Controller	≥20	30	DoF Disabled	8.00	(6.19)	5.64	(5.02)	8.83	(6.64)		
							n/a	DoF Enabled	4.98	(5.78)	3.74	(4.19)	4.60	(4.28)		
Christou and Aristidou (2017)	18 (7F, 11M)	24.00	Game	Low	Controller	<10	17	Pointing	36.00	(37.20)	23.80	(25.90)	39.00	(46.80)	36.40	(37.80)
							n/a	Gaze-directed	21.00	(21.10)	18.70	(20.60)	29.70	(33.20)	25.40	(25.40)
							n/a	Teleport	10.80	(14.40)	13.60	(17.50)	24.10	(31.30)	17.50	(19.70)
Deb et al. (2017)	21 (11F, 10M)	27.84	Game	Low	Walking	≥20	15		8.63	(11.65)	15.16	(16.95)	11.93	(19.33)	14.07	(16.03)
Dennison and D'Zmura (2017)	15 (4F, 11M)	n/a	Minimalist	High	Stationary	<10	7	Seated	15.90	(17.70)	19.20	(18.31)	24.13	(27.01)	22.19	(20.46)
							n/a	Standing	34.34	(39.79)	17.18	(19.53)	22.27	(37.13)	27.93	(33.27)
Dennison and D'Zmura (2018)	20 (5F, 15M)	18–60 (range)	Game	Low	Controller	≥10	10	Push	72.61	(43.92)	50.95	(37.84)	77.33	(73.07)	74.38	(53.23)
							n/a	No Push	48.65	(47.90)	29.56	(29.35)	34.80	(57.30)	43.01	(46.75)
Dennison et al. (2016)	20 (6F, 14M)	n/a	Game	Low	Controller	≥10	55									
Dorado and Figueroa (2014)	44 (8F, 36M)	22	Game	Low	Controller	<10	n/a	Constant Speed Ramp							17.03	(15.87)
							n/a	Constant Speed Stairs							28.42	(22.25)
Farmani and Teather (2018)	14 (5F, 9M)	26.4	Game	High	Controller	≥20	7	Viewpoint Snapping								
	14 (6F, 8M)	26.4					14	No Viewpoint Snapping								
Frommel et al. (2017)	24 (7F, 17M)	27.04 (4.02)	Game	Low	Controller	≥20	n/a	Fixpoint	19.08	(27.27)	18.00	(24.77)	26.10	(36.79)	23.38	(29.95)
							n/a	Free	15.90	(19.22)	16.11	(23.26)	29.00	(35.53)	21.97	(26.36)
							n/a	Guided	27.03	(22.27)	28.43	(26.38)	41.76	(36.48)	35.84	(29.22)
							n/a	Touchpad	28.62	(27.85)	29.06	(26.98)	47.56	(49.07)	38.34	(34.48)
Fujikake et al. (2009)	10	23.6 (2.2)	Minimalist	High	Stationary	<10	n/a	Conventional 3D	11.40	(3.70)	18.20	(4.10)	23.70	(8.80)	19.80	(5.30)
	10	23.6 (2.2)					n/a	New 3D	10.50	(4.40)	17.40	(4.90)	19.50	(6.60)	18.00	(4.90)
Guna et al. (2019)	26 (3F, 23M)	24.75 (5.69)	360	High	Stationary	<10	n/a	Oculus Rift DK1 Action	22.38	(22.08)	26.82	(19.91)	44.44	(41.30)	33.95	(25.28)
							n/a	Oculus Rift DK2 Action	29.72	(29.97)	30.32	(27.79)	54.61	(50.72)	41.28	(35.91)
							n/a	Oculus Rift CV1 Action	29.72	(33.63)	24.49	(24.94)	44.97	(54.46)	35.82	(38.47)
							n/a	Samsung Gear VR Action	26.79	(28.30)	31.20	(25.76)	44.44	(46.92)	37.83	(33.39)
Hutton et al. (2018)	20 (6F, 13M)	26	Minimalist	Low	Stationary		25									
Iskenderova et al. (2017)	31 (6F, 25M)	25.4 (3.3)	Game	Low	Controller	≥10	19	Placebo (1st trial)	86.65	(25.55)	37.26	(20.3)	77.72	(42.16)	73.24	(27.59)
							n/a	Placebo (2nd trial)	81.88	(42.94)	46.11	(26.54)	96.28	(67.29)	80.72	(45.23)
							n/a	Alcohol (1st trial)	82.92	(40.73)	54.22	(33.87)	97.44	(68.66)	85.44	(44.62)
							n/a	Alcohol (2nd trial)	40.36	(35.93)	35.56	(28.08)	64.24	(61.34)	50.63	(42.33)

(Continued)

TABLE 1 | Continued

Authors	<i>n</i>	Age mean (SD)	Content	Visual stimulation	Locomotion	Time (minutes)	Dropout %	Condition/ Group	Nausea	(SD)	Oculomotor	(SD)	Disorientation	(SD)	Total	(SD)
Kang et al. (2008)	20 (10F, 10M)	48.7 (10.7)	Game	Low	Controller	<10	n/a		24.50	(31.20)	37.90	(26.10)	76.60	(73.60)	35.80	(31.40)
Karl et al. (2013)	44 (15F, 29M)	29 (10)	Game	High	Controller	≥10	4									
Kesztyues et al. (2000)	22 (12F, 10M)	22–50 (range)	Scenic	Low	Controller		45		12.40	(14.20)	11.70	(12.10)	3.60	(5.40)	14.20	(15.40)
Kim et al. (2017)	11 (6F, 5M)	28 (7)	Scenic	Low	Walking	≥20	0	Healthy young							8.30	(10.50)
	11 (8F, 3M)	66 (3)	Scenic	Low	Walking	≥20	0	Healthy old							6.50	(13.00)
Kinsella et al. (2016)	120	n/a	360	Low	Stationary	≥10	21									
Kruse et al. (2018)	20 (7F, 13M)	25.75	Minimalist	Low	Walking	≥20	n/a								31.23	(31.30)
Kuiper et al. (2018)	18 (8F, 10M)	25.2 (3.6)	Minimalist	High	Controller	≥10	0		10.60	(12.60)	9.68	(8.14)	21.70	(18.60)		
Lee et al. (2017)	20 (10F, 10M)	25.5 (M) 23.4 (F)	Scenic	Low	Controller	<10	n/a	Gamepad	13.36	(14.90)	26.90	(24.15)	50.11	(51.22)	31.98	(28.67)
					Controller			Hand interface	12.40	(13.86)	25.39	(19.95)	50.81	(47.87)	31.04	(25.23)
					Walking			Walking Simulator	12.89	(13.92)	24.64	(17.76)	43.15	(37.33)	28.80	(22.03)
Ling et al. (2012)	88 (35F, 53M)	28 (6.3)	Game	Low	Stationary	≥10	n/a								2.40	(13.04)
Llorach et al. (2014)	55 (24F, 31M)		Scenic	Low	Controller	≥10	13	Game controller	38.85	(36.49)	15.16	(15.16)	38.48	(36.00)	32.27	(29.26)
	61 (23F, 38M)		Scenic	Low	Walking	≥10	0	Position Estimation	15.37	(17.16)	9.97	(10.60)	18.56	(22.25)	15.93	(14.81)
McGill et al. (2017)	18 (18M)	25.1 (4.7)	Game	High	Stationary	≥10	n/a	VR Motion	39.20	(29.80)	35.00	(28.00)	57.20	(63.40)	47.90	(41.10)
							n/a	VR	53.50	(52.40)	37.90	(33.70)	62.60	(71.50)	56.50	(54.70)
							n/a	VR with Compensation	58.80	(49.90)	43.00	(37.30)	71.90	(72.10)	63.60	(55.90)
							n/a	VR with Peripheral	60.40	(49.70)	43.40	(35.30)	72.70	(71.40)	64.60	(53.20)
Merhi et al. (2007)	24 (11F, 13M) 9 (2F, 7M)	22 20	Game	High	Controller	≥20	100	Experiment 1:							63.60	(49.80)
							59	Standing							58.10	(45.80)
							89	Experiment 1: Sitting							79.40	(24.70)
Mittelstaedt et al. (2018)	60 (40F, 20M)	25.62 (9.34)	Game	High	Controller		7	Experiment 2: Sitting								
Moss et al. (2008)	10 (8F, 2M)	20.6	360	Low	Stationary	≥10	n/a								31.48	(27.37)
Moss et al. (2011)	22 (11F, 11M)	22.6	360	Low	Stationary	≥20	9									
Moss and Muth (2011)	80 (50F, 30M)	19.5 (18–24 range)	360	Low	Stationary	≥10	9								32.33	(4.35)
Munafo et al. (2017)	36 (18F, 18M)	20.7 (.85)	Game	High	Stationary	≥10	17	Experiment 1								
	36 (18F, 18M)	22.7 (3.56)					44	Experiment 2								
Neth et al. (2011)	32 (17F, 15M)	27.26	Minimalist	Low	Walking		9									
Papachristos et al. (2017)	30 (26F, 4M)	24.83 (8.9)	Game	High	Controller	<10	n/a	Mobile Oculus	17.17	(25.80)	24.76	(22.82)	51.97	(65.60)	32.91	(37.43)
Parijat and Lockhart (2011)	16 (8F, 8M)	74.18 (5.82)	Scenic	Low	Walking	≥20	n/a		14.63	(11.89)	32.85	(23.39)	47.33	(43.96)	34.66	(25.66)
Pettijohn et al. (2018)	17 (4F, 13M)	37.4 (3.5)	Scenic	High	Stationary	≥20	18	Profile 1 Active	14.31	(4.33)	8.12	(1.68)	4.97	(2.77)	10.95	(2.46)
							n/a	Profile 1 Passive	11.58	(2.86)	10.29	(2.59)	7.95	(4.05)	11.75	(2.63)
							n/a	Profile 2 Active	20.44	(4.99)	13.54	(3.39)	5.97	(4.05)	16.30	(3.56)
							n/a	Profile 2 Passive	19.76	(4.63)	19.49	(5.37)	9.94	(4.71)	20.04	(5.14)
							n/a	Profile 3 Active	25.89	(5.14)	20.03	(5.43)	8.95	(3.31)	22.44	(4.83)
							n/a	Profile 3 Passive	23.85	(5.08)	22.20	(5.42)	9.94	(5.54)	22.97	(5.14)
							n/a	Posttest Active	12.95	(3.41)	9.75	(2.69)	3.98	(1.74)	10.95	(2.40)
							n/a	Posttest Passive	8.18	(3.30)	10.83	(3.05)	3.98	(2.27)	9.62	(2.62)

(Continued)

TABLE 1 | Continued

Authors	<i>n</i>	Age <i>mean</i> ( <i>SD</i> )	Content	Visual stimulation	Locomotion	Time (minutes)	Dropout %	Condition/ Group	Nausea	( <i>SD</i> )	Oculomotor	( <i>SD</i> )	Disorientation	( <i>SD</i> )	Total	( <i>SD</i> )
Pot-Kolder et al. (2018)	95 (47F, 48M)	25.4 (4.6)	Game	Low	Controller	≥20	n/a	Control (Excluded Clinical)	40.60	(37.40)	29.30	(29.30)	54.20	(52.10)	44.90	(39.30)
Pouke et al. (2018)	13 13	n/a	Scenic	Low	Controller	≥10	8 0	High realism No detail	29.35 16.88	(37.25) (11.78)	25.66 15.74	(25.62) (12.17)	40.69 28.91	(49.37) (25.07)	35.10 22.15	(39.21) (15.30)
Ragan et al. (2017)	40 (10F, 30M)	18–34 (range)	Game	Low	Controller	≥10	n/a								51.43	(40.21)
Rupp et al. (2019)	136 (66F, 70M)	19.82 (2.44)	360	Low	Controller	<10	n/a n/a n/a	Cardboard Oculus DK2 Oculus CV1	11.5 6.17 6.17	(0.24) (0.33) (0.31)	14.05 10.26 12.48	(0.38) (0.34) (0.37)	16.38 9.42 7.37	(0.57) (0.67) (0.60)		
Saldana et al. (2017)	13 (10F, 3M)	8–81.4 (6.25) 5–78.4 (9.37)	Game	Low	Stationary	<10	8	Visit 1								
Schmitz et al. (2018)	35 (19F, 16M) 10 (2F, 8M)	24.97 (3.71) 23.40 (3.2)	Game	Low	Walking		23 0	Exp 2 Post 1 Exp 3 Post 1							48.92 8.31	(45.92) (12.51)
Serge and Fragomeni (2017)	24 (12F, 12M)	19.75 (2.21)	Game	High	Walking	≥10	n/a n/a n/a n/a n/a n/a n/a n/a n/a	S1—Head Tilt S1—Head-Turn S1—Controller S2—Head Tilt S2—Head-Turn S2—Controller S3—Head Tilt S3—Head-Turn S3—Controller					6.55 8.23 5.61 13.56 13.84 15.58 21.04 20.57 10.60	(11.61) (9.13) (8.77) (18.32) (13.66) (13.26) (25.21) (18.51) (3.68)		
Sharples et al. (2008)	19 (9F, 10M)	n/a	Game	Low	Controller	≥20	n/a	Experiment 1: HMD	28.62	(24.63)	27.53	(22.35)	32.97	(31.54)		
Singla et al. (2017)	30 (15F, 15M) 30 (15F, 15M)	25.62 25.62	360	Low	Stationary	≥10	0 0	HTC Vive Oculus Rift								
Song (2017)	14		360	High	Stationary	<10	n/a n/a	Control Experiment	63.70 35.40	(16.00) (12.20)	48.20 41.20	(11.20) (7.90)	83.50 37.80	(16.40) (14.30)		
Stanney et al. (2003)	240 240 240 240 (396F, 564M)	15–53 (range)	n/a	n/a	Controller	≥10 ≥20 ≥20 ≥20	13	15 min 30 min 45 min 60 min	15.86 21.50 23.49 24.80	(23.94) (24.85) (26.50) (25.04)	14.75 19.30 22.55 27.13	(20.19) (19.32) (20.18) (21.86)	24.07 28.94 32.54 33.47	(34.04) (35.67) (35.49) (34.43)	19.96 25.73 29.08 32.10	(25.92) (25.73) (26.63) (26.56)
St. Pierre et al. (2015)	120 (64F, 56M)	n/a	360	Low	Stationary	≥10	0 7 7 23	Baseline condition Constant condition Fixed amplitude condition Varying amplitude condition							24.19 27.43 34.53 60.84	(24.55) (32.87) (33.52) (41.22)
Stauffert et al. (2018)	45 (36F, 9M)	21.18 (2.58) 21.18 (2.58)	Min.	Low	Walking	<10	n/a n/a	No latency With latency jitter	13.83 14.99	(17.91) (15.84)	17.43 16.96	(14.15) (16.93)	13.22 20.55	(15.95) (17.94)	17.58 19.77	(15.78) (16.69)
Tyrrell et al. (2018)	20 (9F, 11M)	46.5 (12)	Game	Low	Controller	<10	n/a	Control	24.33	(30.90)	12.13	(17.80)	23.66	(37.00)	21.88	(29.70)
Walch et al. (2017)	20 (5F, 14M)	25 (3.2)	Game	High	Controller	≥20	0								29.09	(27.65)
Weidner et al. (2017)	94 (24F, 70M)	24.8 (4.7)	Game	High	Controller	≥10	n/a								30.91	(28.24)
Young et al. (2014)	13 (7F, 6M)	18–23 (range)	Scenic	Low	Walking	<10	15									

F, Female; M, Male; HMD, Head-mounted display; DoF, Depth of field; VR, Virtual reality.



**TABLE 2 |** Summary of results.

	Total SSQ score	Nausea	Oculomotor	Disorientation
Content	Q-value = 18.745 $p < 0.001$	Q-value = 47.425 $p < 0.001$	Q-value = 59.106 $p < 0.001$	Q-value = 41.835 $p < 0.001$
Visual stimulation	Q-value = 0.491 $p = 0.483$	Q-value = 0.768 $p = 0.381$	Q-value = 7.314 $p = 0.007$	Q-value = 3.484 $p = 0.062$
Locomotion	Q-value = 15.987 $p < 0.001$	Q-value = 13.141 $p = 0.001$	Q-value = 18.893 $p < 0.001$	Q-value = 5.918 $p = 0.052$
Time	Q-value = 5.433 $p = 0.066$	Q-value = 30.362 $p < 0.001$	Q-value = 2.912 $p = 0.233$	Q-value = 7.126 $p = 0.028$
VR sickness condition	Q-value = 12.236 $p < 0.001$	Q-value = 29.059 $p < 0.001$	Q-value = 24.206 $p < 0.001$	Q-value = 6.562 $p = 0.010$
Age	Q-value = 7.430 $p = 0.006$	Q-value = 0.290 $p = 0.590$	Q-value = 0.010 $p = 0.919$	Q-value = 4.426 $p = 0.035$

SSQ, Simulator Sickness Questionnaire; VR, Virtual reality.

## Description of Studies

Out of the 55 papers included in this review, 20 papers reported both subscale scores and total SSQ scores, 7 papers reported subscale SSQ scores only, and 16 papers reported total SSQ scores only. Twenty papers that reported SSQ scores also reported dropout rates. A further 12 papers that used the SSQ but only reported dropout rates were also included. The total number of experiments from these papers included 54 that reported the total SSQ scores and 38 that reported the subscale SSQ scores (these numbers include between group studies from the same paper). The number of participants included in all experiments represented 3,016 participants. Heterogeneity was consistently high for all analyses ( $I^2 > 90$ ).

Studies came from: Australia ( $n = 3$ ), Canada ( $n = 1$ ), Columbia ( $n = 1$ ), Cyprus ( $n = 1$ ), Finland ( $n = 1$ ), Germany ( $n = 11$ ), Greece ( $n = 1$ ), Japan ( $n = 1$ ), Korea ( $n = 4$ ), Netherlands ( $n = 3$ ), New Zealand ( $n = 1$ ), Portugal ( $n = 1$ ), Slovenia ( $n = 1$ ), Spain ( $n = 1$ ), United Kingdom ( $n = 2$ ), United States of America ( $n = 22$ ).

The pooled mean age of participants was 24 years (of 45 studies that included mean age), with the youngest sample having a mean age of 19.5 years and the oldest having a mean age of 80 years. Fifty-one studies included both female and male participants, 4 studies did not report sex distributions, and 41% of participants were female. Bivariate correlations between the SSQ and percentage of females in studies were not significant ( $r = -0.172$ ,  $p = 0.170$ ).

See **Table 2** for a summary of results showing factors associated with both total and subscale SSQ scores.

**TABLE 3 |** Simulator sickness questionnaire total scores.

		Total SSQ	
	<i>n</i>	<i>M (SE)</i>	95% CI
<b>Content</b>			
360 Videos	10	27.418 (3.434)	[20.688, 34.148]
Game	25	34.259 (2.392)	[29.571, 38.948]
Minimalist	5	21.709 (4.768)	[12.364, 31.055]
Scenic	10	17.329 (3.338)	[10.787, 23.871]
<b>VS</b>			
High	17	30.837 (3.597)	[23.787, 37.888]
Low	33	27.763 (2.511)	[22.841, 32.685]
<b>Locomotion</b>			
Controller	31	32.545 (2.174)	[28.284, 36.806]
Stationary	12	28.036 (3.417)	[21.338, 34.733]
Walking	12	16.993 (3.227)	[10.669, 23.317]
<b>Time</b>			
<10min	15	23.466 (3.226)	[17.144, 29.788]
≥10min	20	33.417 (2.905)	[27.723, 39.111]
≥20min	16	27.354 (3.121)	[21.238, 33.470]
<b>VR sickness condition</b>			
Induce	19	35.274 (2.700)	[29.983, 40.565]
Not induce	35	23.763 (1.882)	[20.075, 27.451]
<b>Age</b>			
Mean < 35	50	28.438 (1.521)	[25.457, 31.420]
Mean ≥ 35	4	14.299 (4.959)	[4.579, 24.019]
<b>All studies</b>	54	28.001 (1.706)	[24.656, 31.345]

VS, Visual stimulation; VR, Virtual reality.

## DISCUSSION

The aim of the review was to synthesize the literature on VR sickness symptoms using HMDs measured using the SSQ. The primary aim was to examine if VR sickness symptoms are influenced by content (four categories), the amount of visual stimulation, how a person moves in the virtual environment and exposure times. With a secondary aim of examining the influence of user characteristics (i.e., age and sex).

## SSQ Scores Interpretation

In this review, total SSQ mean scores ranged from 14.30 to 35.27. Pooled total SSQ scores were relatively high across all studies and content type ( $M = 28.00$ ) with high levels of heterogeneity. Historically the SSQ was intended for military personnel using simulators, however, the different applications and interpretation of the scores have changed with increased use of VR and advancements in technology. When interpreting

**TABLE 4 |** Simulator sickness questionnaire subscale scores.

		Nausea		Oculomotor		Disorientation	
	<i>n</i>	<i>M (SE)</i>	95% CI	<i>M (SE)</i>	95% CI	<i>M (SE)</i>	95% CI
<b>Content</b>							
360 Videos	7	10.460 (1.437)	[7.644, 13.276]	15.186 (0.974)	[13.278, 17.095]	19.493 (2.246)	[15.091, 23.895]
Game	13	26.261 (1.832)	[22.671, 29.852]	26.336 (1.540)	[23.318, 29.353]	38.999 (2.980)	[33.158, 44.840]
Minimalist	6	13.944 (1.874)	[10.271, 17.617]	16.118 (1.306)	[13.559, 18.677]	20.270 (2.686)	[15.005, 25.534]
Scenic	8	15.513 (1.597)	[12.383, 18.642]	12.106 (1.066)	[10.017, 14.196]	15.610 (2.312)	[11.078, 20.142]
<b>VS</b>							
High	12	14.830 (1.359)	[12.167, 17.493]	18.339 (1.017)	[16.346, 20.333]	25.473 (2.212)	[21.138, 29.808]
Low	22	16.338 (1.055)	[14.271, 18.404]	14.968 (0.721)	[13.555, 16.380]	20.435 (1.547)	[17.402, 23.467]
<b>Locomotion</b>							
Controller	27	16.040 (0.907)	[14.262, 17.817]	16.303 (0.643)	[15.043, 17.562]	23.875 (1.424)	[21.084, 26.666]
Stationary	7	22.609 (1.936)	[18.815, 26.403]	22.470 (1.345)	[19.835, 25.106]	28.552 (2.883)	[22.903, 34.202]
Walking	5	13.208 (1.920)	[9.446, 16.970]	15.336 (1.481)	[12.433, 18.238]	18.599 (2.909)	[12.898, 24.300]
<b>Time</b>							
<10min	18	12.884 (1.066)	[10.795, 14.973]	17.030 (0.791)	[15.480, 18.579]	22.217 (1.713)	[18.858, 25.575]
≥10min	10	23.281 (1.836)	[19.683, 26.880]	15.819 (1.206)	[13.455, 18.183]	30.827 (2.745)	[25.446, 36.208]
≥20min	9	19.748 (1.404)	[16.995, 22.500]	18.442 (0.996)	[16.490, 20.393]	24.078 (2.103)	[19.955, 28.200]
<b>VR sickness condition</b>							
Induce	13	24.211 (1.590)	[21.093, 27.328]	21.859 (1.116)	[19.671, 24.047]	28.662 (2.326)	[24.102, 33.222]
Not Induce	25	14.436 (0.871)	[12.730, 16.143]	15.574 (0.621)	[14.357, 16.791]	21.755 (1.363)	[19.084, 24.427]
<b>Age</b>							
Mean < 35	36	16.613 (0.789)	[15.067, 18.158]	17.078 (0.565)	[15.971, 18.185]	24.095 (1.207)	[21.729, 26.461]
Mean ≥ 35	2	18.423 (3.269)	[12.015, 24.831]	17.313 (2.234)	[12.935, 21.691]	13.049 (5.110)	[3.034, 23.065]
<b>All studies</b>	38	16.723 (0.768)	[15.218, 18.229]	17.092 (0.547)	[16.019, 18.165]	23.499 (1.173)	[21.200, 25.797]

VS, Visual stimulation; VR, Virtual reality.

total SSQ scores, according to Kennedy et al. (2003); scores between 10 and 15 indicate significant symptoms; between 15 and 20 are a concern; and scores over 20 indicate a problem simulator. These cut-off scores were established from military personnel using flight simulators, these scores may differ in the general population, additionally, SSQ scores do tend to be higher in other virtual environments compared to flight simulators (Stanney and Kennedy, 1997; Kennedy et al., 2003). According to the Kennedy et al. (2003) categories, even the lowest total SSQ mean score of 14.30 found in studies including older adults in this current review would be regarded as significant symptoms. All remaining classifications displayed higher means with the highest total SSQ score displayed in studies that set out to induce motion sickness.

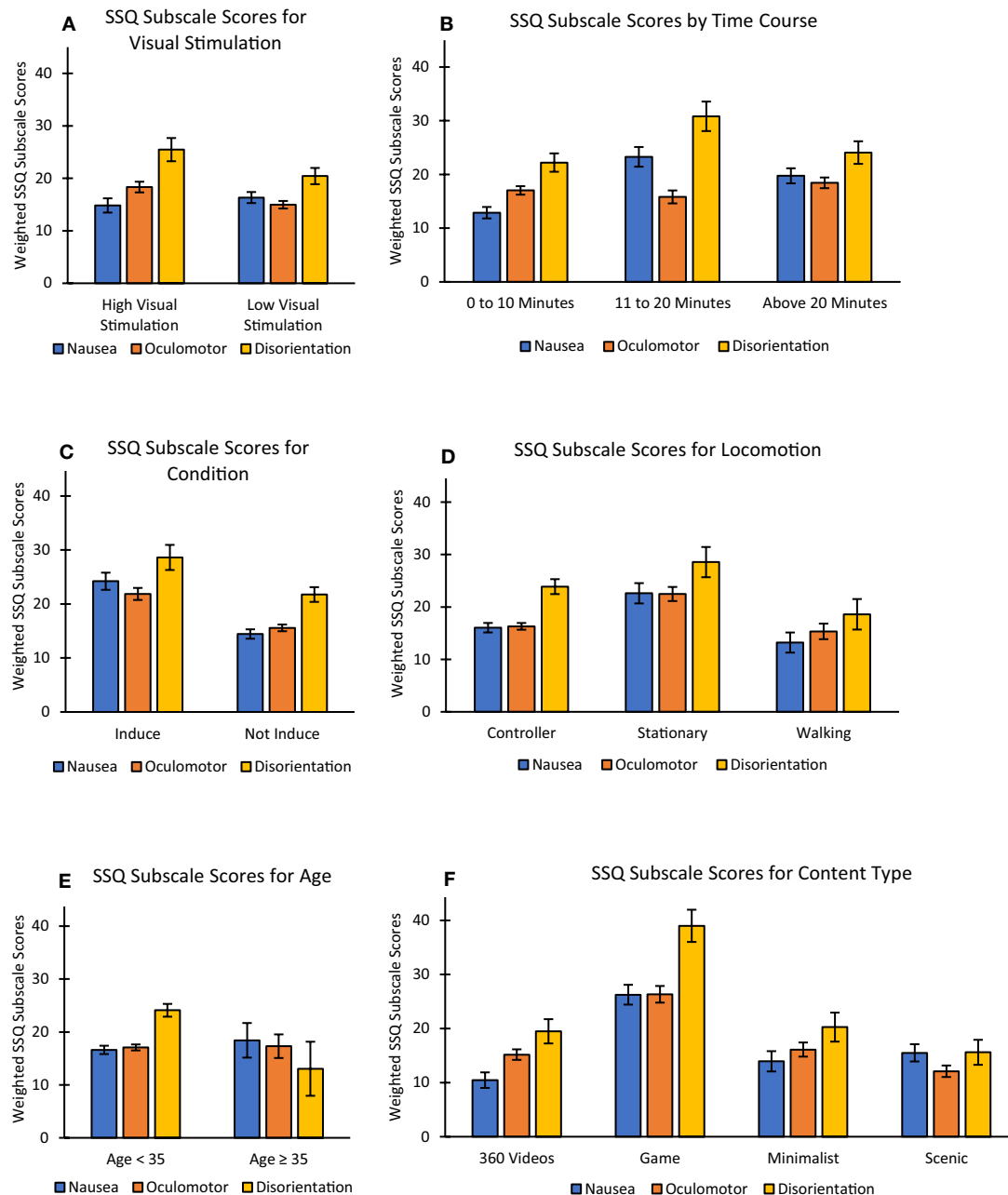
## VR Sickness Symptom Profiles

Across all studies, this review found the highest pooled SSQ subscale scores for disorientation ( $M = 23.50$ ), followed by oculomotor ( $M = 17.09$ ) and nausea ( $M = 16.72$ ). This subscale distribution demonstrates the difference with the symptom profile of motion sickness where nausea typically has the highest rating, followed by oculomotor and disorientation (Rebenitsch and Owen, 2016). These findings increase awareness of symptoms that may be more likely to develop when using HMDs (e.g., dizziness, blurred vision and difficulty focusing).

However, the weighting of these subscales makes it unclear as to what degree these symptoms differ.

## VR Content

The content characteristics in **Figure 2** highlight the distinguishing features of the four content types that may account for the distribution of SSQ scores in this review. SSQ scores were significantly influenced by content type with gaming content displaying the highest total SSQ mean ( $M = 34.26$ ). This effect was also seen for subscale SSQ scores with all measured subscale symptoms of nausea, oculomotor and disorientation highest for gaming content compared to other types of content (see **Table 4**). Consistent with these results, previous studies using gaming content reported the highest dropout rates, ranging from 44 to 100% (Merhi et al., 2007; Dennison et al., 2016; Munafo et al., 2017). The second highest total SSQ means were found in studies using 360 videos. This was followed by minimalist content, with scenic content producing the lowest total SSQ mean. The total SSQ means did not always correspond with dropout rates, for example higher dropout rates were found in scenic content than 360 videos. This discrepancy highlights the variability in how users tolerate HMD use that could be due to other factors. Exposure time, user characteristics or the amount of visual stimulation are all other factors that may have contributed to the high heterogeneity found in this review. Thus, a limitation of this



**FIGURE 3 |** VR sickness symptoms across the different SSQ subscales of all measured factors including **(A)** Visual Stimulation, **(B)** Time, **(C)** VR Sickness Condition, **(D)** Locomotion, **(E)** Age, and **(F)** Content. SSQ, Simulator Sickness Questionnaire; Error bars represent standard error.

current meta-analysis and meta-analyses in general is that methodological differences between studies are collapsed when pooling results.

### Influence of Visual Stimulation on Sickness

Content varies not only by type but also by the amount of visual stimulation offered. For example, all four types of

content examined in this review may provide varying degrees of visual movement to the user. Oculomotor subscale SSQ mean scores were significantly higher for high visual stimulation compared with low visual stimulation. Some of the symptoms in the oculomotor subscale relate to eyestrain, difficulty focusing, difficulty concentrating and blurred vision. Despite recent improvements in display technology, stereoscopic HMDs may

produce more side effects due to the vergence-accommodation conflict. Vergence refers to the way the eyes move laterally to adjust to items moving toward and away from the eyes combined with the process of focusing (accommodation). These visual processes do not occur naturally in a stereoscopic display as accommodation occurs at a fixed screen depth (Terzić and Hansard, 2016). This conflict may be a reason for the higher SSQ means for high visual stimulation in the oculomotor SSQ subscale. When there is a high level of visual stimulation there are more changes in the stimulus distance compared to content with low visual stimulation. The level of visual stimulation is meaningful, as research examining rapid vs. slow changes in stimulus distance found rapid changes to increase visual discomfort (Kim et al., 2014). In a virtual environment, a conflict may be created due to the differences in what a person sees and what their body experiences. With the emergence of new VR technologies, high-quality stereoscopic HMDs are now capable of simulating the visual and spatial properties of the real-world. Despite improvements, current technology still falls short of replicating how humans see and perceive depth under natural viewing conditions (Howarth and Costello, 1997). There are software solutions that can help to reduce discomfort by introducing blurring during motion (Budhiraja et al., 2017), however, this technique may not be effective for everyone. The shortcomings of current HMDs can produce unnatural visual conflicts, which have been shown to play a role in VR sickness (Carnegie and Rhee, 2015), especially when they are combined with visually stimulating VR environments (Kim et al., 2014).

## Locomotion Type in Virtual Environment

SSQ scores were significantly influenced by locomotion type with controller-based movement displaying the highest total SSQ mean ( $M = 32.55$ ). Both nausea and oculomotor subscale SSQ scores means were also significantly influenced by locomotion type with higher scores when stationary as opposed to both controller-based moving and walking (see **Table 4**), high heterogeneity between studies has contributed to these differences. There are several other factors that can account for differences between total and subscale SSQ scores for locomotion between controller-based and stationary content. This includes differences in the number of studies, with seven stationary and five walking studies that reported subscale SSQ data, compared with 12 studies that reported total SSQ data for these locomotion categories. Additionally, relatively high total SSQ scores were reported for controller-based studies (Merhi et al., 2007; Budhiraja et al., 2017; Ragan et al., 2017) that did not report any subscale scores. Finally, these differences between SSQ totals and subscales may result from certain methods of locomotion having a greater impact on specific symptoms in the subscale SSQ scores depending on locomotion type that would not be reflected in the total SSQ scores. For example, being stationary in the real world may induce a greater conflict in a virtual environment where there is movement and hence may increase nausea symptoms. This is consistent with research that indicates a reduction in symptoms when user-initiated movement is matched to the environment (Lee et al., 2017; Misha

et al., 2018), these findings also support the sensory conflict theory relating to a visual-vestibular conflict (Reason and Brand, 1975). Thus, the visual-vestibular conflict may be exacerbated by the type of content (moving vs. static) being viewed combined with the locomotion method. A reduction in visual-vestibular conflict may be the reason that the lowest total and subscale SSQ scores for locomotion were consistently reported in studies that included physically walking content. More research is needed to increase the understanding of how the type of locomotion can influence specific symptoms of VR sickness.

## VR Exposure Time on VR Sickness

Both nausea and disorientation SSQ subscale scores in studies for exposure times of  $<10$  min were lower than those that were equal to or  $>10$  min. Interestingly scores were lower for studies that were equal to or  $>20$  min than those equal to or  $>10$  min (see **Figure 3**). This contradicts a recent summary in a review suggesting that longer exposure times are more likely to increase VR sickness (Duzmanska et al., 2018). Content may have been a factor contributing to this pattern of results within each of the time categories. In examining the distribution of content among the time breakdowns  $\geq 10$  min studies did have the highest percentage of gaming content (62%), compared to studies with the shortest ( $<10$ ) and longest exposure times ( $\geq 20$ ). In addition to this 50% of studies with the longest exposure times ( $\geq 20$ ) consisted of minimalist or scenic content. More research is needed to determine the relationship between content and exposure time. Within-subject designs with different exposure times and controlled content may assist with answering questions around safe exposure times as this information is important when planning clinical trials to avoid VR sickness and dropouts and establish safe use procedures.

## Age and VR Sickness

Four studies included older samples (studies with a mean age range  $\geq 35$  years;  $n = 64$ ) that reported total SSQ scores. Not only did these studies report lower total SSQ scores for older samples ( $M = 14.30$ ) compared to younger samples ( $M = 28.44$ ), these studies reported the lowest SSQ scores when compared with all other examined factors (see **Table 3**). Two of the four studies with older samples also included subscale SSQ scores with 37 participants in total. The disorientation subscale recorded significantly lower SSQ scores for the older samples compared with the younger samples. While scores for nausea and oculomotor subscales were higher for the older adult samples compared with younger samples, they were not statistically significant. Previous research has found inconsistent findings when looking at older samples (Kennedy et al., 2010; Benoit et al., 2015). Even though age has been reported as a user characteristic likely to predict motion sickness (Golding, 2006), the results from this review support previous research that there may be a decline in susceptibility to VR sickness as a person ages (Paillard et al., 2013). However, as there are a limited number of studies including older samples, these results should be interpreted with caution. Additionally, three of the studies used scenic content and one study used gaming content. What also needs to be considered is that the VR content for the studies including older adults may



be assessing specific symptoms, and the virtual environments may be designed to reduce the likelihood of side effects. For example, two of the studies (Parijat and Lockhart, 2011; Kim et al., 2017) involved walking on a treadmill to assess gait or balance and consisted of content with the lowest total SSQ mean scores in this review (scenic content). It is also possible that older adults may experience symptoms that differ to younger adults as indicated with lower disorientation subscale SSQ scores found in the older samples (symptoms related to dizziness, vertigo, blurred vision, nausea and difficulty focusing). With many companies offering VR services to aged care facilities (Aged Care Virtual Reality, 2018; Reminiscence, 2018; Rendeever, 2018), the use by older adults will continue to increase. Moreover, VR delivered in HMDs is being widely used for rehabilitation, assessment and even prediction of cognitive impairments in older adults (Optale et al., 2010; Corriveau Lecavalier et al., 2018; Howett et al., 2019). Therefore, more research is needed to evaluate safety aspects of using HMD-delivered VR with older adults having cognitive decline or other age-related health conditions.

## Sex and VR Sickness

An analysis of sex differences was performed with a correlation between the percentage of females in studies and total SSQ scores. Sex breakdown was not supplied in studies when reporting total SSQ scores, therefore, this was the only way that sex could be analyzed and therefore a limitation of this analysis. The results indicated no difference. This is not consistent with research indicating that females are at higher risk of VR sickness (Lawson et al., 2004). Finding evidence in studies that females are more susceptible than males to VR sickness depends on what study is examined with many confounding variables not taken into account (Lawson, 2015). The importance of this topic suggests that more research is needed to better understand the incidence of VR sickness based on sex differences. Age and sex have been stated as being the most common user characteristics likely to predict motion sickness (Golding, 2006) highlighting a need for further research. Other user characteristics including ethnicity; motion sickness susceptibility; fitness; and prior experience of VR may provide a deeper insight into symptomatology of user characteristics and assist to develop a more targeted approach to dealing with VR sickness.

## Strengths and Limitations

This is the first study to pool estimates of VR sickness symptoms measured with the SSQ using HMDs with a pooled sample size of 3,016, however, the study is not without limitations. Although the most commonly used measure of VR sickness was used (SSQ), there were also many studies excluded (112) that did not use the SSQ. As the SSQ is self-report participants may under or over-report symptoms. Physiological measures can assist with overcoming this limitation however, a consensus is yet to be reached on the best physiological response for assessing VR sickness (Duzmanska et al., 2018). The scoring system for the SSQ can create some confusion and this was seen in this review with some authors incorrectly calculating total scores. Another limitation of the SSQ is the relevance of symptoms for HMD use. For example, the *Virtual Reality Symptom Questionnaire* (Ames

et al., 2005), increased the focus on oculomotor symptoms, while Kim et al. (2018) removed the symptom of nausea in the *Virtual Reality Sickness Questionnaire*, due to not contributing to motion sickness compared with other symptoms, both of these studies were HMD specific. For a more detailed discussion of alternative measures see (Hale and Stanney, 2014).

Additionally, all analyses had high heterogeneity demonstrating large variation across the included studies. As well as individual differences of age and sex, susceptibility to VR sickness can also vary between individuals and therefore influence results. Gaming or VR experience is another individual difference that can influence the likelihood of side effects and needs to be both reported and taken into account during analysis of results. The small number of studies including older adults and lack of reporting of sex differences and dropouts are also limitations and areas requiring further research or improved reporting in future VR studies including HMDs. As 22 studies did not report dropout rates, the rate of 15.6% may be inflated if many of these studies did not have dropouts, however, we cannot assume there were no dropouts if they were not reported. This highlights the need to make reporting of dropout rates a standard in VR research.

Finally, another limitation involves the varied nature of the HMDs used across these studies. HMDs can differ in terms of field of view, use of stereo, resolution, framerate, availability of inter-pupillary distance controls/adjustment, and other technical display factors. Modern HMDs from the last 5 years differ fundamentally from the more limited display technology that was available before these recent advances (Kourtesis et al., 2019), and since 35% of papers included in this analysis used these older HMDs, it is difficult to predict how those findings would predict the occurrence of symptoms with use of currently available HMDs. Moving forward, there is an obvious need for more controlled laboratory research with standard reference VR environments that are adjustable in terms of content, movement, user interaction, etc. With such specifically created environments, one would be able to test out the incidence of side effects across different display types with varied hardware capabilities. This will be essential for promoting parametric research that creates a database of known properties for different types of virtual environments delivered across varied hardware types and would serve to produce the baseline normative data needed to enable better research in how to mitigate or eliminate the incidence of these use-limiting side effects.

## Conclusion

Previous research has focused on the influence of technological aspects on VR sickness. This review advances this knowledge by examining content as a major contributing factor to VR sickness, which will remain a problem despite future technological advances. Our findings show that content significantly influences VR sickness symptoms. Recent HMD technology can provide a better experience (Kourtesis et al., 2019) and if this is combined with careful selection of content the risk of VR sickness can be reduced and those symptoms that do occur can be easily managed. In this review, we compared our total SSQ scores with the cut-off scores suggested by Kennedy et al. (2003), what

these scores mean in relation to HMDs and how these scores relate to the general population remains unclear. Nevertheless, comparing total scores between studies shows that content is a major contributing factor. This review also highlights the need for a further understanding of the influence of user characteristics such as age and sex as there is a lack of studies including older samples, and sex differences that are often not reported. Increasing our understanding of VR sickness could be particularly valuable to researchers and practitioners, as there may be ethical and liability implications in research, training and clinical applications.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## REFERENCES

- Aged Care Virtual Reality (2018). *VR Programs for Health Benefits*. Retrieved from <http://www.agedcarevirtualreality.com/> (accessed December 17, 2018).
- Ames, L. S., Wolffsohn, S. J., and McBrien, A. N. (2005). The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optom. Vis. Sci.* 82, 168–176. doi: 10.1097/01.OPX.0000156307.95086.6
- Arcioni, B., Palmisano, S., Apthorp, D., and Kim, J. (2018). Postural stability predicts the likelihood of cybersickness in active HMD-based virtual reality. *Displays* 58, 3–11. doi: 10.1016/j.displa.2018.07.001
- Benoit, M., Guerchouche, R., Petit, P.-D., Chapoulie, E., Manera, V., Chaurasia, G., et al. (2015). Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering. *Neuropsychiatr. Dis. Treat.* 11, 557–563. doi: 10.2147/NDT.S73179
- Bermúdez Rey, M. C., Clark, T. K., Wang, W., Leeder, T., Bian, Y., and Merfeld, D. M. (2016). Vestibular perceptual thresholds increase above the age of 40. *Front. Neurol.* 7:162. doi: 10.3389/fneur.2016.00162
- Bessa, M., Melo, M., Narciso, D., Barbosa, L., and Vasconcelos-Raposo, J. (2016). “Does 3D 360 video enhance user’s VR experience? An evaluation study,” in *Proceedings of the XVII International Conference on Human Computer* (New York, NY), 1–4. doi: 10.1145/2998626.2998669
- Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2013). *Comprehensive Meta-Analysis Version 3*. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.
- Bos, J. E., MacKinnon, S. N., and Patterson, A. (2005). Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. *Aviat. Space Environ. Med.* 76, 1111–1118.
- Brooks, J., Lodge, R., and White, D. (2017). Comparison of a head-mounted display and flat screen display during a micro-UAV target detection Task. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 61, 1514–1518. doi: 10.1177/1541931213601863
- Budhiraja, P., Miller, M. R., Modi, A. K., and Forsyth, D. (2017). Rotation blurring: use of artificial blurring to reduce cybersickness in virtual reality first person shooters. *arXiv [Preprint]*. arXiv:1710.02599.
- Carnegie, K., and Rhee, T. (2015). Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE Comput. Graph. Appl.* 35, 34–41. doi: 10.1109/MCG.2015.98
- Chance, S. S., Gaunet, F., Beall, A. C., and Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: the contribution of vestibular and proprioceptive inputs to path integration. *Presence Teleop. Virt. Environ.* 7, 168–178. doi: 10.1162/105474698565659
- Cheung, B., and Hofer, K. (2002). Lack of gender difference in motion sickness induced by vestibular Coriolis cross-coupling. *J. Vestibul. Res.* 12, 191–200.

## AUTHOR CONTRIBUTIONS

DS, AS, and TL conception of the work and analyzed and interpreted the results. DS and AS article selection and screening. DS wrote the manuscript. All authors revised the work critically for important intellectual content and have read and approved the manuscript. AS and DS created **Figure 2**, certain 3D models for this figure sourced from cadnav.com and modified.

## FUNDING

DS was supported by the Australian Government Research Training Program Scholarship. TL and HK were funded by National Health and Medical Research Council (NHMRC) Dementia Research Leadership Fellowships (GNT1136269 & GNT1135676 respectively).

- Christou, C. G., and Aristidou, P. (2017). *Steering Versus Teleport Locomotion for Head Mounted Displays*. Cham: Springer. doi: 10.1007/978-3-319-60928-7\_37
- Clemes, S. A., and Howarth, P. A. (2005). The menstrual cycle and susceptibility to virtual simulation sickness. *J. Biol. Rhythms.* 20, 71–82. doi: 10.1177/0748730404272567
- Cobb, S. V. G., Nichols, S., Ramsey, A., and Wilson, J. R. (1999). Virtual Reality-Induced Symptoms and Effects (VRISE). *Presence Teleop. Virt. Environ.* 8, 169–186. doi: 10.1162/105474699566152
- Corriveau Lecavalier, N., Ouellet, É., Boller, B., and Belleville, S. (2018). Use of immersive virtual reality to assess episodic memory: a validation study in older adults. *Neuropsychol. Rehabil.* 1–19. doi: 10.1080/09602011.2018.1477684
- Covidence (2019). *Covidence Systematic Review Software*. Melbourne, Australia: Veritas Health Innovation. Retrieved from: <https://www.covidence.org/home> (accessed February 4, 2019).
- Deb, S., Carruth, D. W., Sween, R., Strawderman, L., and Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. *Appl. Ergon.* 65, 449–460. doi: 10.1016/j.apergo.2017.03.007
- Dennison, M. S., and D’Zmura, M. (2017). Cybersickness without the wobble: experimental results speak against postural instability theory. *Appl. Ergon.* 58, 215–223. doi: 10.1016/j.apergo.2016.06.014
- Dennison, M. S., and D’Zmura, M. (2018). Effects of unexpected visual motion on postural sway and motion sickness. *Appl. Ergon.* 71, 9–16. doi: 10.1016/j.apergo.2018.03.015
- Dennison, M. S., Wisti, A. Z., and D’Zmura, M. (2016). Use of physiological signals to predict cybersickness. *Displays* 44, 42–52. doi: 10.1016/j.displa.2016.07.002
- Dorado, J. L., and Figueroa, P. A. (2014). “Ramps are better than stairs to reduce cybersickness in applications based on a HMD and a Gamepad,” in *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (Minneapolis, MN), 47–50. doi: 10.1109/3DUI.2014.6798841
- Duzmanska, N., Strojny, P., and Strojny, A. (2018). Can simulator sickness be avoided? a review on temporal aspects of simulator sickness. *Front. Psychol.* 9:2132. doi: 10.3389/fpsyg.2018.02132
- Era, P., Sainio, P., Koskinen, S., Haavisto, P., Vaara, M., and Aromaa, A. (2006). Postural balance in a random sample of 7,979 subjects aged 30 years and over. *Gerontology* 52, 204–213. doi: 10.1159/000093652
- Farmani, Y., and Teather, R. J. (2018). “Viewpoint snapping to reduce cybersickness in virtual reality,” in *Proceedings of Graphics Interface* (Toronto, ON). doi: 10.20380/GI2018.21
- Fernandes, A. S., and Feiner, S. K. (2016). “Combating VR sickness through subtle dynamic field-of-view modification,” in *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (Greenville, SC). doi: 10.1109/3DUI.2016.7460053
- Frommel, J., Sonntag, S., and Weber, M. (2017). “Effects of controller-based locomotion on player experience in a virtual reality exploration game,” in *Proceedings of the 12th International Conference on the Foundations of Digital Games* (Hyannis, MA), 1–6. doi: 10.1145/3102071.3102082

- Fujikake, K., Miyao, M., Watanabe, T., Hasegawa, S., Omori, M., and Takada, H. (2009). "Evaluation of body sway and the relevant dynamics while viewing a three-dimensional movie on a head-mounted display by using stabilograms," in *International Conference on Virtual and Mixed Reality*, ed R. Shumaker (Berlin: Heidelberg), 41–50. doi: 10.1007/978-3-642-02771-0\_5
- Fulvio, J. M., Ji, M., and Rokers, B. (2019). Variability in sensory sensitivity predicts motion sickness in virtual reality. *bioRxiv [Preprint]*. 488817. doi: 10.1101/488817
- Gallagher, M., and Ferré, E. R. (2018). Cybersickness: a multisensory integration perspective. *Multisens. Res.* 31, 645–674. doi: 10.1163/22134808-20181293
- Gavani, A. M., Nesbitt, K. V., Blackmore, K. L., and Nalivaiko, E. (2017). Profiling subjective symptoms and autonomic changes associated with cybersickness. *Auton. Neurosci.* 203, 41–50. doi: 10.1016/j.autneu.2016.12.004
- Golding, J. F. (2006). Predicting individual differences in motion sickness susceptibility by questionnaire. *Pers. Individ. Dif.* 41, 237–248. doi: 10.1016/j.paid.2006.01.012
- Guna, J., Geršak, G., Humar, I., Song, J., Drnovšek, J., and Pogačnik, M. (2019). Influence of video content type on users' virtual reality sickness perception and physiological response. *Future Gener. Comput. Syst.* 91, 263–276. doi: 10.1016/j.future.2018.08.049
- Hale, K. S., and Stanney, K. M. (2014). *Handbook of Virtual Environments: Design, Implementation, and Applications, 2nd Edn.* Boca Raton, FL: CRC Press. doi: 10.1201/b17360
- Howarth, P. A., and Costello, P. J. (1997). The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18, 107–116. doi: 10.1016/S0141-9382(97)00011-5
- Howett, D., Castegnaro, A., Krzywicka, K., Hagman, J., Marchment, D., Henson, R., et al. (2019). Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain* 142, 1751–1766. doi: 10.1093/brain/awz116
- Hutton, C., Ziccardi, S., Medina, J., and Rosenbarg, E. S. (2018). "Please don't puke: early detection of severe motion sickness in VR," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen), 579–580. doi: 10.1109/VR.2018.8446382
- Iskenderova, A., Weidner, F., and Broll, W. (2017). "Drunk virtual reality gaming: exploring the influence of alcohol on cybersickness," in *Conference: Annual Symposium on Computer-Human Interaction in Play* (Amsterdam), 561–572. doi: 10.1145/3116595.3116618
- Kang, Y. J., Ku, J., Han, K., Kim, S. I., Yu, T. W., Lee, J. H., et al. (2008). Development and clinical trial of virtual reality-based cognitive assessment in people with stroke: preliminary study. *Cyberpsychol. Behav.* 11, 329–339. doi: 10.1089/cpb.2007.0116
- Karl, I., Berg, G., Ruger, F., and Farber, B. (2013). Driving behavior and simulator sickness while driving the vehicle in the loop: validation of longitudinal driving behavior. *IEEE Trans. Intell. Transp. Syst. Mag.* 5, 42–57. doi: 10.1109/ITS.2012.2217995
- Kennedy, R. S., Drexler, J., and Kennedy, R. C. (2010). Research in visually induced motion sickness. *Appl. Ergon.* 41, 494–503. doi: 10.1016/j.apergo.2009.11.006
- Kennedy, R. S., Drexler, J. M., Compton, D. E., Stanney, K. M., Lanham, D. S., and Harm, D. L. (2003). "Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences," in *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues*, eds L. J. Hettinger and M. W. Haas (Lawrence Erlbaum Associates Publishers), 247–278. doi: 10.1201/9781410608888.ch12
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303\_3
- Keshavarz, B., and Hecht, H. (2011). Validating an efficient method to quantify motion sickness. *Hum. Factors* 53, 415–426. doi: 10.1177/0018720811403736
- Kesztyues, T. I., Mehlitz, M., Schilken, E., Weniger, G., Wolf, S., Piccolo, U., et al. (2000). Preclinical evaluation of a virtual reality neuropsychological test system: occurrence of side effects. *Cyberpsychol. Behav.* 3, 343–349. doi: 10.1089/10949310050078788
- Kim, A., Darakjian, N., and Finley, J. M. (2017). Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *J. Neuroeng. Rehabil.* 14:16. doi: 10.1186/s12984-017-0225-2
- Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. *Appl. Ergon.* 69, 66–73. doi: 10.1016/j.apergo.2017.12.016
- Kim, J., Kane, D., and Banks, M. S. (2014). The rate of change of vergence-accommodation conflict affects visual discomfort. *Vision Res.* 105, 159–165. doi: 10.1016/j.visres.2014.10.021
- Kim, Y. Y., Kim, H. J., Kim, E. N., Ko, H. D., and Kim, H. T. (2005). Characteristic changes in the physiological components of cybersickness. *Psychophysiology* 42, 616–625. doi: 10.1111/j.1469-8986.2005.00349.x
- Kinsella, A., Mattfeld, R., Muth, E., and Hoover, A. (2016). Frequency, not amplitude, of latency affects subjective sickness in a head-mounted display. *Aerosp. Med. Hum. Perform.* 87, 604–609. doi: 10.3357/AMHP.4351.2016
- Kourtesis, P., Collina, S., Doumas, L. A. A., and MacPherson, S. E. (2019). Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Front. Hum. Neurosci.* 13:342. doi: 10.3389/fnhum.2019.00342
- Kruse, L., Langbehn, E., and Steinicke, F. (2018). "I can see on my feet while walking: sensitivity to translation gains with visible feet," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen), 305–312. doi: 10.1109/VR.2018.8446216
- Kuiper, O. X., Bos, J. E., and Diels, C. (2018). Vection does not necessitate visually induced motion sickness. *Displays* 58, 82–87. doi: 10.1016/j.displa.2018.10.001
- LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* 32, 47–56. doi: 10.1145/333329.333344
- Lawson, B. D. (2015). "Motion sickness symptomatology and origins," in *Handbook of Virtual Environments*, eds K. S. Hale and K. M. Stanney (Boca Raton FL: Taylor & Francis Group), 531–600. doi: 10.1201/b17360-29
- Lawson, B. D., Kass, S., Lambert, C., and Smith, S. (2004). Survey and review concerning evidence for gender differences in motion susceptibility. *Aviat. Space Environ. Med.* 75:105.
- Lee, J., Kim, M., and Kim, J. (2017). A study on immersion and VR sickness in walking interaction for immersive virtual reality applications. *Symmetry* 9:78. doi: 10.3390/sym9050078
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6:e1000100. doi: 10.1371/journal.pmed.1000100
- Ling, Y., Brinkman, W. P., Nefs, H. T., Qu, C., and Heynderickx, I. (2012). Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking. *Presence Teleop. Virt. Environ.* 21, 254–267. doi: 10.1162/PRES\_a\_00111
- Lorach, G., Evans, A., and Blat, J. (2014). "Simulator sickness and presence using HMDs: comparing use of a game controller and a position estimation system," in *VRST '14 Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh), 137–140. doi: 10.1145/2671015.2671120
- McCauley, M. E., and Sharkey, T. J. (1992). Cybersickness: perception of self-motion in virtual environments. *Presence Teleop. Virt. Environ.* 1, 311–318. doi: 10.1162/pres.1992.1.3.311
- McGill, M., Ng, A., and Brewster, S. (2017). "I am the passenger: how visual motion cues can influence sickness for in-car VR," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO), 5655–5668. doi: 10.1145/3025453.3026046
- Merhi, O., Faugloire, E., Flanagan, M., and Stoffregen, T. A. (2007). Motion sickness, console video games, and head-mounted displays. *Hum. Factors* 49, 920–934. doi: 10.1518/001872007X230262
- Min, B.-C., Chung, S.-C., Min, Y.-K., and Sakamoto, K. (2004). Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. *Appl. Ergon.* 35, 549–556. doi: 10.1016/j.apergo.2004.06.002
- Misha, S., Xuhai, X., Aske, M., and Pattie, M. (2018). "VMotion: designing a seamless walking experience in VR," *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong).
- Mittelstaedt, J., Wacker, J., and Stelling, D. (2018). Effects of display type and motion control on cybersickness in a virtual bike simulator. *Displays* 51, 43–50. doi: 10.1016/j.displa.2018.01.002
- Moss, J. D., Austin, J., Salley, J., Coats, J., Williams, K., and Muth, E. R. (2011). The effects of display delay on simulator sickness. *Displays* 32, 159–168. doi: 10.1016/j.displa.2011.05.010



- Moss, J. D., and Muth, E. R. (2011). Characteristics of head-mounted displays and their effects on simulator sickness. *Hum. Factors* 53, 308–319. doi: 10.1177/0018720811405196
- Moss, J. D., Scisco, J., and Muth, E. (2008). “Simulator sickness during head mounted display (HMD) of real world video captured scenes,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (New York, NY), 52, 1631–1634.
- Munafò, J., Diedrick, M., and Stoffregen, T. A. (2017). The virtual reality head-mounted display occlusal rift induces motion sickness and is sexist in its effects. *Exp. Brain Res.* 235, 889–901. doi: 10.1007/s00221-016-4846-7
- Nesbitt, K., Davis, S., Blackmore, K., and Nalivaiko, E. (2017). Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays* 48, 1–8. doi: 10.1016/j.displa.2017.01.002
- Neth, C. T., Souman, J. L., Engel, D., Kloos, U., Bühlhoff, H. H., and Mohler, B. J. (2011). “Velocity-dependent dynamic curvature gain for redirected walking,” in *IEEE Transactions on Visualization and Computer Graphics* 18 (Singapore), 1041–1052. doi: 10.1109/VR.2011.5759454
- Nichols, S. (1999). Physical ergonomics of virtual environment use. *Appl. Ergon.* 30, 79–90. doi: 10.1016/S0003-6870(98)00045-3
- Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., et al. (2010). Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. *Neurorehabil. Neural Repair.* 24, 348–357. doi: 10.1177/1545968309353328
- Paillard, A., Quarcq, G., Paolino, F., Denise, P., Paolino, M., Golding, J. F., et al. (2013). Motion sickness susceptibility in healthy subjects and vestibular patients: effects of gender, age and trait-anxiety. *J. Vestib. Res.* 23, 203–209. doi: 10.3233/VES-130501
- Papachristos, N. M., Vrellis, I., and Mikropoulos, T. A. (2017). “A comparison between Oculus Rift and a low-cost smartphone VR headset: immersive user experience and learning,” in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (Timisoara), 477–481. doi: 10.1109/ICALT.2017.145
- Parijat, P., and Lockhart, T. E. (2011). “Can virtual reality be used as a gait training tool for older adults?” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55 (Las Vegas, NV), 157–161. doi: 10.1177/1071181311551033
- Pettijohn, K. A., Geyer, D., Gomez, J., Becker, W. J., and Biggs, A. T. (2018). Postural instability and simulator seasickness. *Aerosp. Med. Hum. Perform.* 89, 634–641. doi: 10.3357/AMHP.4998.2018
- Pot-Kolder, R., Veling, W., Counotte, J., and van der Gaag, M. (2018). Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. *Cyberpsychol. Behav. Soc. Netw.* 21, 187–193. doi: 10.1089/cyber.2017.0082
- Pouke, M., Tiirio, A., LaValle, S. M., and Ojala, T. (2018). “Effects of visual realism and moving detail on cybersickness,” in *25th IEEE Conference on Virtual Reality and 3D User Interfaces, (VR)* (Reutlingen), 665–666. doi: 10.1109/VR.2018.8446078
- Powell, W., Rizzo, A., Sharkey, P., and Merrick, J. (2018). Innovations and challenges in the use of virtual reality technologies for rehabilitation. *J. Altern. Med. Res.* 10, 3–4.
- Ragan, E. D., Scerbo, S., Bacim, F., and Bowman, D. A. (2017). Amplified head rotation in virtual reality and the effects on 3D search, training transfer, and spatial orientation. *IEEE Trans. Vis. Comput. Graph.* 23, 1880–1895. doi: 10.1109/TVCG.2016.2601607
- Ray, B., Jung, J., and Larabi, M. (2018). 3-7 Sept. 2018). “On the possibility to achieve 6-DoF for 360 video using divergent multi-view content,” in *2018 26th European Signal Processing Conference (EUSIPCO)* (Rome), 211–215. doi: 10.23919/EUSIPCO.2018.8553397
- Reason, J. T., and Brand, J. J. (1975). *Motion Sickness*. Oxford: Academic Press.
- Rebenitsch, L., and Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Real* 20, 101–125. doi: 10.1007/s10055-016-0285-9
- Reminiscent (2018). *Reminiscent Based Virtual Reality*. Retrieved from <https://www.reminiscent.co.uk/about-us> (accessed December 17, 2018).
- Rendever (2018). *Virtual Reality for Seniors*. Retrieved from <http://rendever.com/> (accessed December 17, 2018).
- Rizzo, A. S., and Koenig, S. T. (2017). Is clinical virtual reality ready for primetime? *Neuropsychology* 31, 877–899. doi: 10.1037/neu0000405
- Rupp, M. A., Odette, K. L., Kozachuk, J., Michaelis, J. R., Smither, J. A., and McConnell, D. S. (2019). Investigating learning outcomes and subjective experiences in 360-degree videos. *Comput. Educ.* 128, 256–268. doi: 10.1016/j.compedu.2018.09.015
- Saldana, S. J., Marsh, A. P., Rejeski, W. J., Haberl, J. K., Wu, P., Rosenthal, S., et al. (2017). Assessing balance through the use of a low-cost head-mounted display in older adults: a pilot study. *Clin. Interv. Aging.* 12, 1363–1370. doi: 10.2147/CIA.S141251
- Schmitz, P., Hildebrandt, J., Valdez, A. C., Kobbelt, L., and Ziefle, M. (2018). “You spin my head right round: threshold of limited immersion for rotation gains in redirected walking,” in *IEEE Trans. Vis. Comput. Graph.* 24, 1623–1632. doi: 10.1109/TVCG.2018.2793671
- Serge, S. R., and Fragomeni, G. (2017). “Assessing the relationship between type of head movement and simulator sickness using an immersive virtual reality head mounted display: a pilot study,” in *International Conference on Virtual, Augmented and Mixed Reality*, Vol. 10280 (Cham: Springer) 556–566. doi: 10.1007/978-3-319-57987-0\_45
- Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection display systems. *Displays* 29, 58–69. doi: 10.1016/j.displa.2007.09.005
- Shin, J., An, G., Park, J.-S., Baek, S. J., and Lee, K. (2016). Application of precise indoor position tracking to immersive virtual reality with translational movement support. *Multimed. Tools Appl.* 75, 12331–12350. doi: 10.1007/s11042-016-3520-1
- Singla, A., Fremerey, S., Robitza, W., Lebreton, P., and Raake, A. (2017). “Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution,” in *Proceedings of the Thematic Workshops of ACM Multimedia* (Mountain View, CA), 511–519. doi: 10.1145/3126686.3126768
- Song, E.-J. (2017). A Study on cyber sickness of virtual reality based on HMD. International Information Institute (Tokyo). *Information* 20, 5015–5022.
- Stanney, K. M., Hale, K. S., Nahmens, I., and Kennedy, R. S. (2003). What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Hum. Factors* 45, 504–520. doi: 10.1518/hfes.45.3.504.27254
- Stanney, K. M., and Kennedy, R. S. (1997). The psychometrics of cybersickness. *Commun. ACM* 40, 66–68. doi: 10.1145/257874.257889
- Stauffer, J. P., Niebling, F., and Latoschik, M. E. (2018). “Effects of Latency Jitter on Simulator Sickness in a Search Task,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen), 121–127. doi: 10.1109/VR.2018.8446195
- St. Pierre, M. E., Banerjee, S., Hoover, A. W., and Muth, E. R. (2015). The effects of 0.2 Hz varying latency with 20–100 ms varying amplitude on simulator sickness in a helmet mounted display. *Displays* 36, 1–8. doi: 10.1016/j.displa.2014.10.005
- Terzić, K., and Hansard, M. (2016). Methods for reducing visual discomfort in stereoscopic 3D: a review. *Signal Process. Image Commun.* 47, 402–416. doi: 10.1016/j.image.2016.08.002
- Tyrrell, R., Sarig-Bahat, H., Williams, K., Williams, G., and Treleaven, J. (2018). Simulator sickness in patients with neck pain and vestibular pathology during virtual reality tasks. *Virtual Real.* 22, 211–219. doi: 10.1007/s10055-017-0324-1
- Walch, M., Frommel, J., Rogers, K., Schüssel, F., Hock, P., Döbelstein, D., et al. (2017). “Evaluating VR driving simulation from a player experience perspective,” in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, CO), 2982–2989. doi: 10.1145/3027063.3053202
- Wang, P., Wu, P., Wang, J., Chi, H.-L., and Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. *Int. J. Environ. Res. Public Health* 15, 1204. doi: 10.3390/ijerph15061204
- Weidner, F., Hoesch, A., Poeschl, S., and Broll, W. (2017). “Comparing VR and non-VR driving simulations: an experimental user study,” in *2017 IEEE Virtual Reality (VR)* (Los Angeles, CA), 281–282. doi: 10.1109/VR.2017.7892286
- White, P. J., Byagowi, A., and Moussavi, Z. (2015). Effect of viewing mode on pathfinding in immersive virtual reality. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 4619–4622. doi: 10.1109/EMBC.2015.7319423



- Young, M. K., Gaylor, G. B., Andrus, S. M., and Bodenheimer, B. (2014). "A comparison of two cost-differentiated virtual reality systems for perception and action tasks," in *Proceedings of the ACM Symposium on Applied Perception* (Vancouver, BC), 83–90. doi: 10.1145/2628257.2628261
- Ziegler, P., Roth, D., Knots, A., Kreuzer, M., and von Mammen, S. (2018). "Simulator sick but still immersed: a comparison of head-object collision handling and their impact on fun, immersion, and simulator sickness," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Reutlingen), 743–744. doi: 10.1109/VR.2018.8446221

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

Copyright © 2020 Saredakis, Szpak, Birkhead, Keage, Rizzo and Loetscher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Application of Supervised Machine Learning for Behavioral Biomarkers of Autism Spectrum Disorder Based on Electrodermal Activity and Virtual Reality

Mariano Alcañiz Raya<sup>1\*</sup>, Irene Alice Chicchi Giglioli<sup>1</sup>, Javier Marín-Morales<sup>1</sup>, Juan L. Higuera-Trujillo<sup>1</sup>, Elena Olmos<sup>1</sup>, Maria E. Minissi<sup>1</sup>, Gonzalo Teruel Garcia<sup>1</sup>, Marian Sirera<sup>2</sup> and Luis Abad<sup>2</sup>

<sup>1</sup> Instituto de Investigación e Innovación en Bioingeniería, Universitat Politècnica de Valencia, Valencia, Spain, <sup>2</sup> Red Cenit, Centros de Desarrollo Cognitivo, Valencia, Spain

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Dayi Bian,  
Vanderbilt University, United States  
Máté Aller,  
University of Cambridge,  
United Kingdom

### \*Correspondence:

Mariano Alcañiz Raya  
malcaniz@i3b.upv.es

### Specialty section:

This article was submitted to  
Sensory Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 15 November 2019

**Accepted:** 27 February 2020

**Published:** 03 April 2020

### Citation:

Alcañiz Raya M, Chicchi Giglioli IA, Marín-Morales J, Higuera-Trujillo JL, Olmos E, Minissi ME, Teruel Garcia G, Sirera M and Abad L (2020) Application of Supervised Machine Learning for Behavioral Biomarkers of Autism Spectrum Disorder Based on Electrodermal Activity and Virtual Reality. *Front. Hum. Neurosci.* 14:90. doi: 10.3389/fnhum.2020.00090

**Objective:** Sensory processing is the ability to capture, elaborate, and integrate information through the five senses and is impaired in over 90% of children with autism spectrum disorder (ASD). The ASD population shows hyper-hypo sensitiveness to sensory stimuli that can generate alteration in information processing, affecting cognitive and social responses to daily life situations. Structured and semi-structured interviews are generally used for ASD assessment, and the evaluation relies on the examiner's subjectivity and expertise, which can lead to misleading outcomes. Recently, there has been a growing need for more objective, reliable, and valid diagnostic measures, such as biomarkers, to distinguish typical from atypical functioning and to reliably track the progression of the illness, helping to diagnose ASD. Implicit measures and ecological valid settings have been showing high accuracy on predicting outcomes and correctly classifying populations in categories.

**Methods:** Two experiments investigated whether sensory processing can discriminate between ASD and typical development (TD) populations using electrodermal activity (EDA) in two multimodal virtual environments (VE): forest VE and city VE. In the first experiment, 24 children with ASD diagnosis and 30 TDs participated in both virtual experiences, and changes in EDA have been recorded before and during the presentation of visual, auditive, and olfactive stimuli. In the second experiment, 40 children have been added to test the model of experiment 1.

**Results:** The first exploratory results on EDA comparison models showed that the integration of visual, auditive, and olfactive stimuli in the forest environment provided higher accuracy (90.3%) on sensory dysfunction discrimination than specific stimuli. In the second experiment, 92 subjects experienced the forest VE, and results on 72 subjects showed that stimuli integration achieved an accuracy of 83.33%. The final confirmatory test set ( $n = 20$ ) achieved 85% accuracy, simulating a real application

of the models. Further relevant result concerns the visual stimuli condition in the first experiment, which achieved 84.6% of accuracy in recognizing ASD sensory dysfunction.

**Conclusion:** According to our studies' results, implicit measures, such as EDA, and ecological valid settings can represent valid quantitative methods, along with traditional assessment measures, to classify ASD population, enhancing knowledge on the development of relevant specific treatments.

**Keywords:** autism spectrum disorder, sensory dysfunction, virtual reality, electrodermal activity, assessment

## INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopment disorder characterized by a wide range of impairments, ranging from social to physical and cognitive functions (Baron-Cohen, 1990), affecting one in 160 children (World Health Organization [WHO], 2019). ASD symptoms arise as early as 2 to 4 years in age, and in some cases, the signs of ASD might start as early as 6 months old (Lord et al., 2006; Anagnostou et al., 2014). Specifically, ASD is associated with social and interaction symptoms as well as stereotyped and repetitive behavior patterns (American Psychiatric Association, 2013) that have a significant impact on educational (Levy and Perry, 2011) and social life (Schmidt et al., 2015). Furthermore, sensory processing dysfunctions have been observed as a relevant aspect of ASD symptomatology; indeed it is experienced by over 90% of ASD children (Leekam et al., 2007; Tomchek and Dunn, 2007; Baron-Cohen et al., 2009). Sensory processing is the ability to capture, elaborate, and integrate information through the five senses (touch, movement, smell, taste, vision, and hearing), allowing adapting behavioral responses to the environment (Miller et al., 2007). In the ASD population, such sensory processing and integration of stimuli are experienced differently from that of the typical development (TD) population, affecting response to stimuli. In more details, they show hyper-sensitivities (over-responsiveness) and hypo-sensitivities (under-responsiveness) to a wide range of sensory stimuli. Previous studies on sensory dysfunctions showed a hypersensitivity to visual and auditive stimuli, such as bright lights or noisy sounds (Tomchek and Dunn, 2007; Baron-Cohen et al., 2009; Tomchek et al., 2014); conversely, with olfactive stimuli, they present hypo-sensitiveness in detecting odor threshold (Dudova et al., 2011; Ashwin et al., 2014). Sensory dysfunction consequently affects the information processing in ASD, and it has been suggested that it may be the cause of impairments in several psychological domains, such as in cognitive and social responses (Tomchek and Dunn, 2007; Baron-Cohen et al., 2009).

### Current Issues in ASD Diagnosis and the Need for Biomarkers in ASD

Traditionally, ASD diagnosis and assessment include a series of explicit qualitative and quantitative measures characterized by semi-structured behavioral tasks' observations in which the examiner rates and scores an individual's responses to prompted situations (e.g., the Autism Diagnostic Observation Schedule, ADOS; Lord et al., 1999) and family structured interview (e.g., the

Autism Diagnostic Interview-Revised, ADI-R; Lord et al., 1994). For example, the ADOS measure consists of various standardized activities introduced by the examiner, such as a simulation of having a snack together, that permits to observe the occurrence or non-occurrence of behaviors related to ASD. ADOS principally focuses on social behavior and communication analysis, and it is characterized by five different modules that allow tailoring of assessment to the age and communication development of the participants. Regarding sensory processing, the utmost test for its evaluation is Sensory Profile-2 (Dunn, 2014), a qualitative questionnaire in which family caregivers answer to several questions about activities at home, in school, and in the community (see the section "Materials and Methods" for test description).

Despite these instruments having been widely adopted in ASD research and clinical practice, several limitations remain (Volkmar et al., 2009), mainly regarding the absence of explicit sensory functioning assessment, the subjective evaluation and the examiner's expertise, and the ecological validity of the assessment setting.

Concerning the first limitation, traditional assessments have been designed following both ASD ICD-10 and DSM IV guidelines that do not consider sensory dysfunction as a necessary and distinct diagnostic criterion. Thus, ADI-R and ADOS do not tap sensory processing and responsiveness (Leekam et al., 2007). Second, training in administration and scoring is crucial and highly recommended (Lord et al., 2001) since test results and diagnosis rely on the examiner's subjective ability to detect ASD-related features. Examiners who not have a high level of ASD-specific previous training and expertise might lead to inappropriate task presentation and administration. This could influence the rating and the scoring, contributing to over- or under-interpretation of the outcomes and prompting a misleading assessment (Reaven et al., 2008). Another limitation that can cause ADOS' unreliable outcomes and affect the truthfulness of responses is the social desirability bias (Paulhus, 1991). Social desirability is a response bias in which individuals attempt to answer to tasks or questions in a manner that will be viewed as favorable by others (Edwards, 1957). First, part of the ASD assessment consists of reporting child information by family caregivers, who can interpret differently specific behaviors according to their personal perspective and experience (Möricke et al., 2016). Second, in ASD assessment, children may have been taught to act according to specific settings (e.g., laboratory settings) (Francis, 2005), and it might be that whether the same situation happened in the real world, examiners would obtain

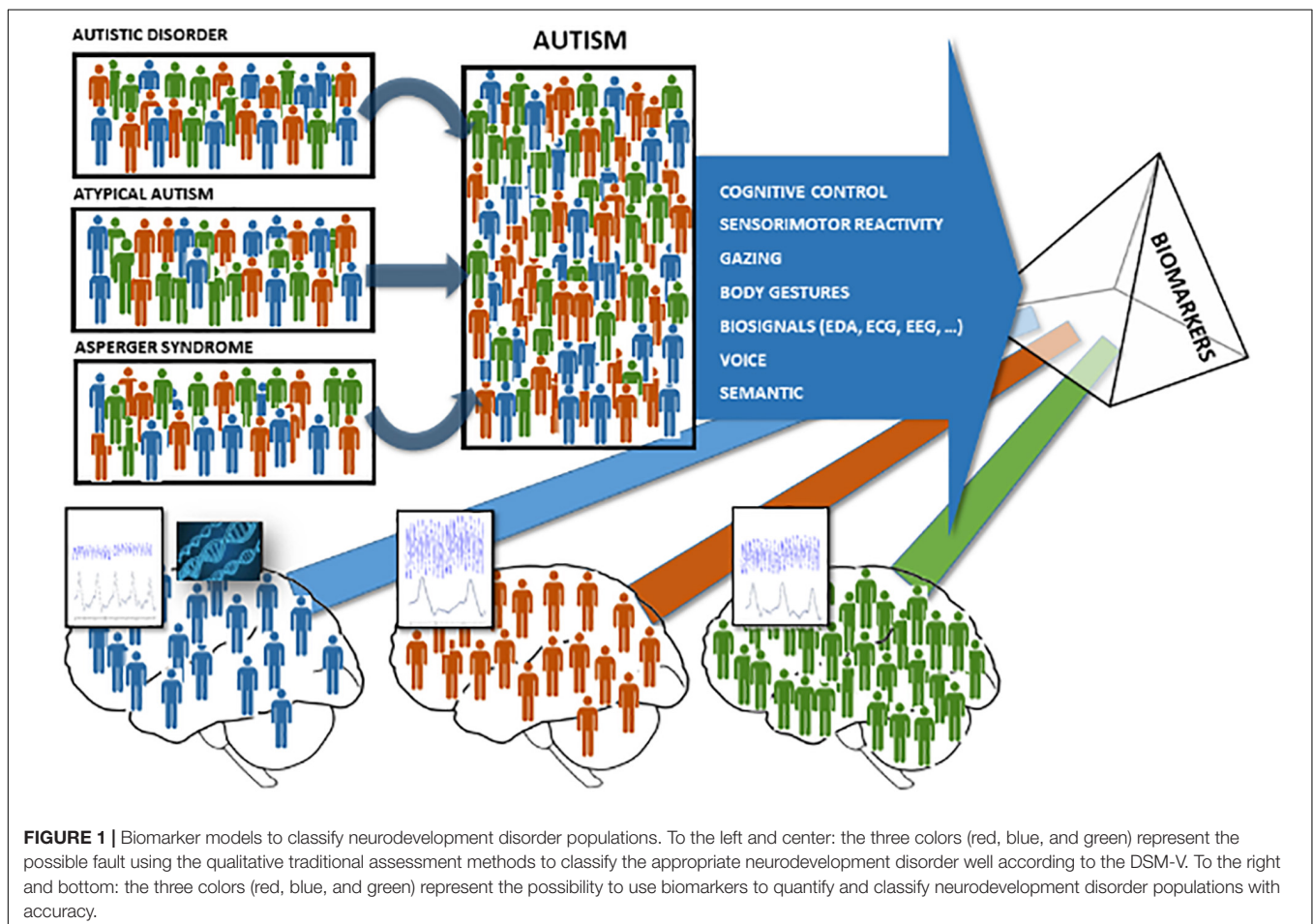
different responses (Gillberg and Rasmussen, 1994). Finally, although diagnostic structured interviews are considered as the gold standard in ASD assessment (Goldstein et al., 2009), they usually take place in the laboratory rather than in ecologically valid settings. Ecologically valid settings are environments and situations similar to real ones, able to elicit everyday experiences and behaviors related to daily functioning (Franzen and Wilhelm, 1996; Chaytor et al., 2006). The more the assessment measure is valid from an ecological point of view, the more that the results can be generalized to the real world (Brunswick, 1955; Chaytor et al., 2006). Indeed recent studies showed that traditional assessment results did not reflect performance in real-life situations and *vice versa* (Parsons S., 2016).

According to these limitations, the existing ASD diagnosis criteria (DSM, ICD, ADOS, and ADIR) do not consider quantitative variations in symptom severity in each person's measurements and do not take into account the biological bases of the disorder. Recently, there has been a growing need for more reliable and valid diagnostic measures, such as biomarkers, to distinguish typical and atypical functioning and to reliably track the progression of the illness, thus helping to diagnose ASD (Figure 1). In order to generate valid quantitative models between explicit symptoms and implicit biomarkers, the emerging field of Computational Psychiatry (CP) is seeking, first,

to mathematically model brain responses to the problems it faces and, second, to study how the “abnormal” experiences, emotions, and behaviors that are commonly used to describe disorders contribute to normal function and neural processes (Montague et al., 2012; Friston et al., 2014; Wang and Krystal, 2014; Redish and Gordon, 2016).

## Implicit Processes as Pillars for ASD Biomarkers

Currently, the EU AIMS Longitudinal European Autism Project is one of the largest multicenter, multidisciplinary studies to identify the stratification biomarkers for ASD and the biomarkers that may serve as surrogate ends (Murphy and Spooren, 2012). However, all participants are comprehensively characterized in terms of their brain structure and function [assessed using structural magnetic resonance imaging (sMRI), functional MRI (fMRI), and electroencephalogram (EEG)], biochemical biomarkers, prenatal environmental risk factors, and genomics. Nonetheless, when experiencing social situations, it is equally important to study the related behavioral outputs. Up to now, most of the information contained in the behavioral inputs do not seem to have been noticed. Studying social situations on how people process, store, and apply data about other people and





social circumstances can provide us with objective information about the ASD evaluation.

Recent progress in social cognitive neuroscience (SCN), a field of study including biological processes and cognition-based aspects (Lieberman, 2010), is confuting the majority of social cognition models that suggest that humans can analyze and correctly verbalize their beliefs, feelings, and behaviors (Nosek et al., 2011), showing that our social interactions are mostly governed by unconscious processes that happen without conscious awareness or control (Forscher et al., 2019). To study the unconscious processes, several implicit measures, including brain images, behavior, and psychophysiological tracking, have been developed as alternative research methods to explicit measures since they are able to capture implied brain processes (Ledoux et al., 2016).

In the ASD population, implicit measures can contribute, along with traditional techniques, to obtain a more objective assessment from a quantitative point of view (Alcañiz et al., 2019). Various techniques used are based on measurements linked to some implied system in effect. The adoption of implicit SCN metrics as biomarker input variables for ASD evaluation suggests a move toward a quantitative ASD diagnosis. Some previous studies proposed the use of brain activity (fMRI and EEG), physiological measures (heart variability—HR), and behavioral responses (eye tracking measures—ET—body movement recognition), with the goal of capturing the ASD patient's behavioral structure while being subjected to a stimulus (Di Martino et al., 2014; Van Hecke et al., 2015; Chita-Tegmark, 2016; Wang et al., 2016; Großekathöfer et al., 2017). For example, brain activity studies showed that ASD patients using fMRI present general brain hyperactivity and alterations in the middle and the posterior insula and in the cingulate posterior cortex (Di Martino et al., 2014). EEG studies in ASD showed greater activity in the left hemisphere in social situations (Van Hecke et al., 2015). In ASD, the study of gaze activity measured by eye tracking tools was analyzed as behavioral tests, linking the gaze patterns to the existence of nuclear deficits. Many studies have succeeded in linking this implicit measure with the affection core deficits, with the degree of social, emotional, and cognitive skill development. Even in circumstances of social participation, predictors of ASD were found based on ocular actions and facial processing (Chita-Tegmark, 2016).

## Electrodermal Activity in ASD

To date, electrodermal activity (EDA, Nikula, 1991), a marker of sympathetic nervous system arousal, is one of the main implicit measures examined in ASD (White et al., 2014; Fenning et al., 2017). Specifically, it is an implicit neurophysiological process related to electrical properties of the skin, based on variations in sweating, skin conductance, heart rate, and blood flow to muscles when individuals are facing either internal or external stimuli (Fagius and Wallin, 1980; Benedek and Kaernbach, 2010; Boucsein, 2012). Its analysis allows to discern, among others, the phasic component of the signal, with rapidly changing activity, referred to the subject's responses to discrete stimuli (being an indicator of sympathetic activity), and the tonic component,

with slowly changing activity, referred to the subject's basal conductance level (Dawson et al., 2007).

Regarding sensory dysfunction in ASD, multiple studies have investigated its relationship with EDA, comparing baseline arousal and EDA reactions to sensory stimuli among ASD individuals, neurotypical development population, and other diagnostic groups (for reviews, see Rogers and Ozonoff, 2005; White et al., 2014; Lydon et al., 2016). The evidence from these studies are controversial: some research found no differences in EDA levels in response to sensory stimuli (e.g., Zahn et al., 1987; Rogers and Ozonoff, 2005; McCormick et al., 2014), whereas other studies were successful (van Engeland et al., 1991; Miller et al., 2001; Rogers and Ozonoff, 2005; Schoen et al., 2009).

Overall regarding auditive stimulation, the enhanced EDA levels in ASD individuals have been associated to both baseline arousal and reaction to stimulus presentation (Palkovitz and Wiesenfeld, 1980; Barry and James, 1988; Chang et al., 2012); nevertheless, there are also instances about no differences between ASD and typical populations (e.g., Stevens and Gruzelier, 1984; Allen et al., 2013). Moreover, same pattern of mixed results has been found for immediate EDA of ASD individuals in visual stimulations: regarding reactions to facial expressions, autistic people exhibited weakened EDA responses compared to typical adults and children (Hirstein et al., 2001; Hubert et al., 2009; Riby et al., 2012), whereas Ben Shalom et al. (2006) found no differences; likewise, several studies related increased EDA reactivity to direct eye gaze in children with ASD (Kylliäinen and Hietanen, 2006; Joseph et al., 2008; Kylliäinen et al., 2012), but, conversely, other investigations did not (Louwerse et al., 2013). Furthermore, regarding smell processing, the ASD children seemed to be more sensitive than the TD children (Schoen et al., 2009); thus, they can detect odors at shorter distances (Ashwin et al., 2014); on the other hand, they have difficulties in detecting odor threshold (Dudova et al., 2011).

Finally, the correlation between ASD traditional assessments and EDA measures has been studied, observing that higher levels of ASD symptoms, measured by ADOS, are related to greater variability in EDA (Fenning et al., 2017).

## Use of Virtual Reality in ASD

To date, the above-described implicit measuring methodologies can be divided into two groups: studying the actions of the subject in a real scenario or conducting experiments in laboratory settings. The main problem with actual real-life scenarios is that it is not easy to study human responses in real situations because the experimenter struggles to fully monitor the stimuli involved in the encounter. Conversely, participants face controlled conditions in laboratory settings that do not include certain variables present in real-life situations, resulting in the experiment's low ecological validity.

Virtual reality (VR) emerges as a promising technology capable of overcoming the problems mentioned above. VR offers the opportunity to create different real situations, including social situations that produce body interactions in which the body, environment, and brain are closely related. VR can be described as a virtual 3D environment that can replicate real experiences where participants can interact as if they were in

the real world. Different technical tools can create a sense of presence, enabling the subjects to view their behaviors as real (Slater, 2009). Experiencing a high sense of presence enables the participants in the virtual environment (VE) to communicate and behave as if they were thinking, acting, and communicating in their real life (Alcañiz et al., 2019). Therefore, actions, attitudes, and beliefs can be transferred from nature to virtuality and *vice versa* and can occur spontaneously and unconsciously, generating circumstances of high ecological validity and maintaining high experimental control in stimuli presentation and in gathering behavioral performance. Neuroscientists are increasingly using VR to replicate natural phenomena and social interactions, developing immersive and multimodal sensory stimuli that provide advantages over real-life and traditional testing methodologies on the controlled stimuli and accuracy in data gathering (Bohil et al., 2011) and allowing also the integration of behavioral measures.

The use of VR in ASD research has been postulated as one of the methods with great potential in the treatment of the main symptomatological nucleus (Wing et al., 2011; Parsons T. D., 2016; Golestan et al., 2018). Such advantages have the theoretical basis established by Blascovich et al. (2002), who argued that interactive VEs would be able to change interaction and evaluation by offering the opportunity to study human behavior in normal, controlled, and replicable environments to produce an individual response close to that obtained in a real context.

One of the aims that are replicated for ASD and VR users throughout the research is to improve their ability to work in everyday life. Research that has built VE to learn different skills in children with ASD are not difficult to find: cognitive learning (Kandalaf et al., 2013), interaction (Bernardini et al., 2014), and emotional training (Bekele et al., 2014).

Nevertheless, there is a lack of research applied to the diagnosis in the field of VR in which an objective assessment of ASD is conducted through individualized clinical tests (behavioral biomarkers), customizing the treatment to each patient's profile.

To our knowledge, no one has investigated whether multimodal VR settings and EDA reactions might contribute to predicting ASD population versus TD children. Starting from these premises, we performed two studies (the first exploratory and the second confirmatory) to discriminate and predict sensory processing in the ASD population versus in a TD population through the combined use of implicit measure (EDA) and different sensory stimuli, involving two different VE and tasks.

To this extent, the first experiment aimed to analyze the influence of three factors in predicting ASD: (1) the VE contents, one VE including a relaxing environment and another one including an arousal environment; (2) the task, one related to the subject's greeting responses in the relaxing environment and others related to the subject's imitation in the arousal environment; and (3) the stimuli conditions (SC), including visual (V), visual and auditive (VA), and visual, auditive, and olfactive stimuli (VAO). Specifically, in the first environment, the participants have been projected into a forest wherein the visual stimulus was a girl avatar appearing, the auditive stimulus was the sound of the rain, and the olfactive stimulus was the odor of fresh-cut grass. In this relaxing environment, the subjects were

asked to complete tasks related to responding to the greetings of the avatars. In the second environment, the participants were introduced in a city street intersection in which the visual stimulus was the presence of two avatars (a girl and a boy), the auditive stimulus was a song that avatars danced to, and the olfactive stimulus was the smell of butter related to avatars that bit a muffin. In this arousal environment, the subjects were asked to complete a task related to the imitation of the actions of the avatars. In both environments and experiments, the EDA responses were recorded and introduced in a supervised machine learning classifier in order to recognize ASD.

Starting from these premises and aims, the first hypothesis in experiment 1 was that the ASD recognition is higher in the forest since the response to a greeting is one of the confirmatory symptoms in the ASD. The second hypothesis was that, by including more sensory modalities, the ASD recognition using EDA would present a better performance. After that, we performed a second experiment in order to develop a supervised learning model using the outputs of the first experiment. We increased the number of subjects used to calibrate the model and we tested it in a set of subjects not used before, simulating a real-world application.

## MATERIALS AND METHODS

### Experiment 1

#### Participants

The study included 52 children between the ages of 4 and 7 years. In detail, 23 TD children (age =  $4.87 \pm 0.92$ ; male = 13, female = 10) and 29 children with a previous diagnosis of ASD (age =  $5.20 \pm 1.34$ ; male = 26, female = 3) participated in experiment 1. The ASD group sample was recruited from the Development Neurocognitive Centre, Red Cénit, Valencia, Spain. The ASD and the TD participants presented individual assessment reports that included the results of their ADOS-2 test. A sample management company recruited the TD group through targeted mailings to families. Before participating in the study, the family caregivers received written information about the study and they were required to give written consent for inclusion in the investigation. The study obtained ethical approval from the Ethical Committee of the Polytechnic University of Valencia. Furthermore, all procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

#### Psychological Assessment

The following scales and tasks have been administered to the participants and their family caregivers:

- Autism Diagnostic Interview-Revised (ADI-R): The ADI-R (Lord et al., 1994) is a clinical semi-structured interview used to detect ASD and answered by family caregivers. The questions are linked to ICD-10 and DSM-IV criteria for autism and yield separate scores

in three domains—communication, social interaction, and restricted, repetitive, and stereotyped behaviors. The answers are scored on a 0–3-point scale, in which 0 indicates the absence of the behavior and 3 indicates the clear manifestation of the determined behavior. ADI-R presents high psychometric properties and the test–retest reliability ranged from 0.93 to 0.97.

- Autism Diagnostic Observation Schedule (ADOS-2): The ADOS-2 (Lord et al., 1999) includes structured and semi-structured tasks to assess children's development in several areas, such as communication, use of imagination, social interaction and play, and restrictive and repetitive behaviors. The measure uses five modules, tailored to the age and communication development of the participants. Concretely, module T is for young children who are between 12 and 30 months old and do not use phrase language consistently, module 1, for children who are 31 months or older and who do not use phrase language consistently, module 2 for children of any age who use phrase language but who do not have verbal fluency, module 3 for children with fluent language and young adolescents (under 16), and finally module 4 for adults and adolescents (16 years and older) with fluent language. From the observation of these behaviors, the items are scored between 0 (no evidence of abnormality related to autism) and 3 (definitive evidence), and from the sum of scores, two specific indices (social affectation and restricted and repetitive behavior) and the ASD global total index are obtained. The ADOS-2 presents excellent psychometric properties: the test–retest reliability is 0.87 for the social affectation index, 0.64 for the repetitive behavior index, and 0.88 for the total global index. In the study, the assessment was performed using module 1, corresponding to children from 31 months of age who do not use phrase language consistently.

## The Virtual Environments

The 3D models were developed in the Institute for Research and Innovation in Bioengineering (i3B) at the Polytechnic University of Valencia. The environment was developed and projected inside a three-surface Cave Assisted Virtual Environment (CAVE<sup>TM</sup>) with dimensions of 4 m × 4 m × 3 m. It was equipped with three ceiling ultra-short lens projectors, which can project a 100° image from just 55 cm. The sound system used was the Logitech Speaker System Z906 500W 5.1 THX Digital (**Figure 2**).

Two VEs were developed:

1. A virtual forest, including three controlled stimuli conditions: visual, visual–auditive, and visual–auditive–olfactive (**Figure 3**). The visual stimuli consisted of a girl's avatar appearing from the left side of the forest and walking to the central virtual scene, where she stopped and waved her hand three times to say hello to the child, and then leaving the virtual scene, walking to the right side of the forest (**Figure 4**). The auditive stimuli consisted of adding to the virtual forest a storm and rain sound. Finally, the olfactive stimuli consisted of an odor of fresh-cut grass.

2. The other VE involved a simulated city street intersection (**Figure 5**) and was divided into three experimental stimuli conditions: visual, visual–auditive, and visual–auditive–olfactive. First, in the V stimuli condition, a boy's avatar appeared from the left side of the surface CAVE<sup>TM</sup>, walking to the center of the virtual scene, where he stopped and waved his hand three times to say hello to the child, and then leaving the virtual scene, walking out of the street intersection (**Figure 6**). Successively, a girl's avatar appeared in the central of the surface CAVE<sup>TM</sup>, walking to the right of the virtual scene, where she stopped and repeated the three waves with her hand to say hello to the child, and then leaving the virtual scene, walking to the right side of the street intersection. This sequence was repeated three times. In the second VA stimuli condition, the same avatars appeared in the same order from the same directions, but instead of waving the hand to say hello, they danced over a piece of music for 10 s for three times. In the VAO stimuli condition, the same avatars appeared in the same order and from the same directions, but they bit a buttered muffin, accompanied by the same song of the previous condition and an artificial butter smell that was released during the VR experience.

To avoid transfer effects over VR experiences, the VE presentation (forest and city street intersection) was counterbalanced across participants and a 1-week rest was left between the two VR experiences. Despite counterbalancing practice is also recommended for stimuli conditions of VEs, to reduce the possibility to provoke sensory sensitiveness overload in ASD children, the same stimuli presentation order was maintained (V, VA, and VAO) for the entire sample in both VR experiences. Indeed sensory sensitiveness in ASD can suddenly emerge in different situations that require the processing capacity of sensory integration from several channels (Bogdashina, 2016); such concurrent sensory decoding of stimuli might yield ASD children distress and uncomfortable states that could affect the quality of performance and assessment in VEs.

## Physiological Assessment and Data Processing

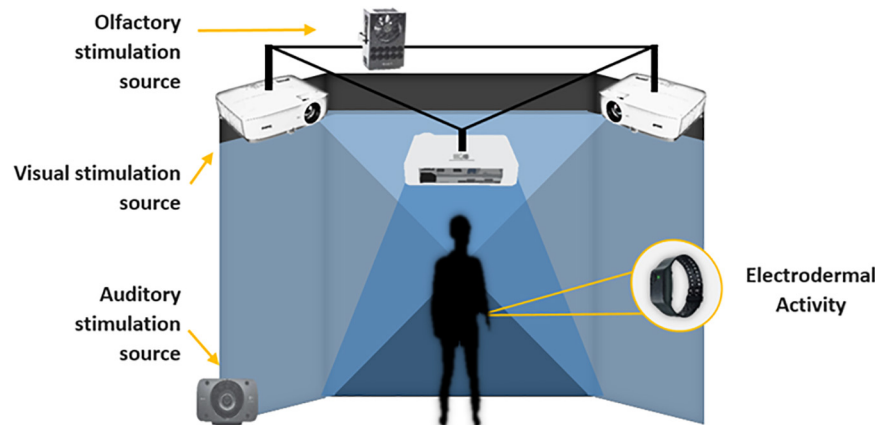
Electrodermal activity signal was recorded using an Empatica E4 wristband.<sup>1</sup> Its reliability has been found to be comparable to clinical devices in appropriate circumstances (McCarthy et al., 2016). Raw signal (recorded at 4 Hz and 0.001–100  $\mu$ S) was pre-processed and analyzed using Ledalab<sup>2</sup> (v.3.4.8) via Matlab<sup>3</sup> (v.2016a). Pre-processing consisted of two successive phases: (1) Butterworth low-pass signal filtering at 2.5 Hz (Valenza and Scilingo, 2013) and (2) visual diagnosis of artifacts and their corrections. Due to the records characteristics and the analysis chosen, it was not considered necessary to apply signal-smoothing techniques. The analysis was tackled through the continuous decomposition analysis (CDA) method. It is based on the deconvolution of the skin conductance signal by the general response shape, prior to the data decomposition in the tonic and

<sup>1</sup>www.empatica.com

<sup>2</sup>www.ledalab.de

<sup>3</sup>www.mathworks.com





**FIGURE 2 |** Experimental setting.

phasic components. As mentioned above, the tonic component generates slow changes in the conductance signal (magnitude of minutes), being considered the basal activity, and the phasic component generates rapid changes in the conductance signal (magnitude of seconds), being considered the response of the subjects to discrete stimuli. CDA has been proven to be an appropriate method for the analysis of short intervals between stimuli, especially in situations that can generate a high phasic activity (Benedek and Kaernbach, 2010). In order to reduce inter-subject differences, all values were standardized according to Venables and Christie (1980). This process was applied to the subject's whole experience record. Finally, the set of metric extracted to characterize each stimuli condition includes the mean of tonic (BL tonic) and phasic (BL phasic) component of the baseline performed previously to the stimuli condition, the mean of tonic and phasic component of the responses to the stimuli condition, and the ratio between the tonic and the phasic component of the responses to the stimuli condition.

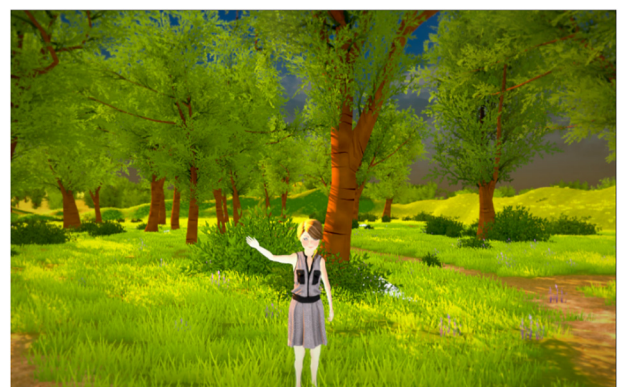
### The Olfactive System

For the olfactive stimulus, we used the Olorama<sup>4</sup> Technology™ wireless freshener. It features 12 scents arranged in 12 pre-charged channels, which can be selected and triggered by means of a UDP packet. The device encompasses a programmable fan time system that dissipates the scent. Both the intensity of the chosen scent (amount of time the scent valve was open) and

the amount of fan time were programmed. The scent valve was opened all the time during the last stimuli condition (VAO).

### Experimental Procedure

First, the family caregivers of the participants were informed about the general objectives of the research, the physiological measure and its device localization, and the VR system. Second, the Empatica E4 device was shown and placed on the participant's non-dominant hand before the virtual session. Subsequently, the child was accompanied in the CAVE by the researcher and by his or her family caregiver, according to the child's needs, and was placed in the middle of the virtual room, standing in front of the central surface at a distance of 1.5 m. Before



**FIGURE 4 |** Girl's avatar saying hello.



**FIGURE 3 |** Virtual forest.



**FIGURE 5 |** Virtual city street intersection.

<sup>4</sup>[www.olorama.com](http://www.olorama.com)





**FIGURE 6** | Boy's avatar saying hello.

each stimuli condition, 2 min of EDA baseline was recorded in rest and relaxing state, and then the VE experiences started (**Figure 7**). The total duration of the forest VE experience was 8 min and 15 s, and each stimuli condition lasted 45 s. The total duration of the city VE was 14 min, and each stimuli condition lasted for 2 min and 40 s. The participants were balanced between the two VEs, leaving a 1-week rest between the two experimental sessions.

During the three VR stimuli conditions in both virtual experiences, the EDA signals were recorded. The researcher monitored the child state during the entire experiment, and care was taken to address any indisposition derived from the use of the devices.

## Experiment 2

### Participants

The study added 40 children, between the ages of 4 and 7 years, to experiment 1. In detail, 23 TD children (age  $4.86 \pm 0.91$ ; male = 13, female = 10) and 17 ASD children (age  $5.13 \pm 1.35$ ; male = 14; female = 3) participated in experiment 2. The ASD group sample was recruited from the Development Neurocognitive Centre, Red Cenit, Valencia, Spain. The ASD and the TD participants presented an individual assessment report that included the results of their ADOS-2 test. A sample management company recruited the TD group through targeted mailings to families. Before participating in the study, the family caregivers received written information about the study, and they were required to give written consent for the inclusion in the investigation.

### Psychological Assessment

Experiment 2 utilized the same scales and tests of experiment 1.

### Physiological Assessment and Data Processing

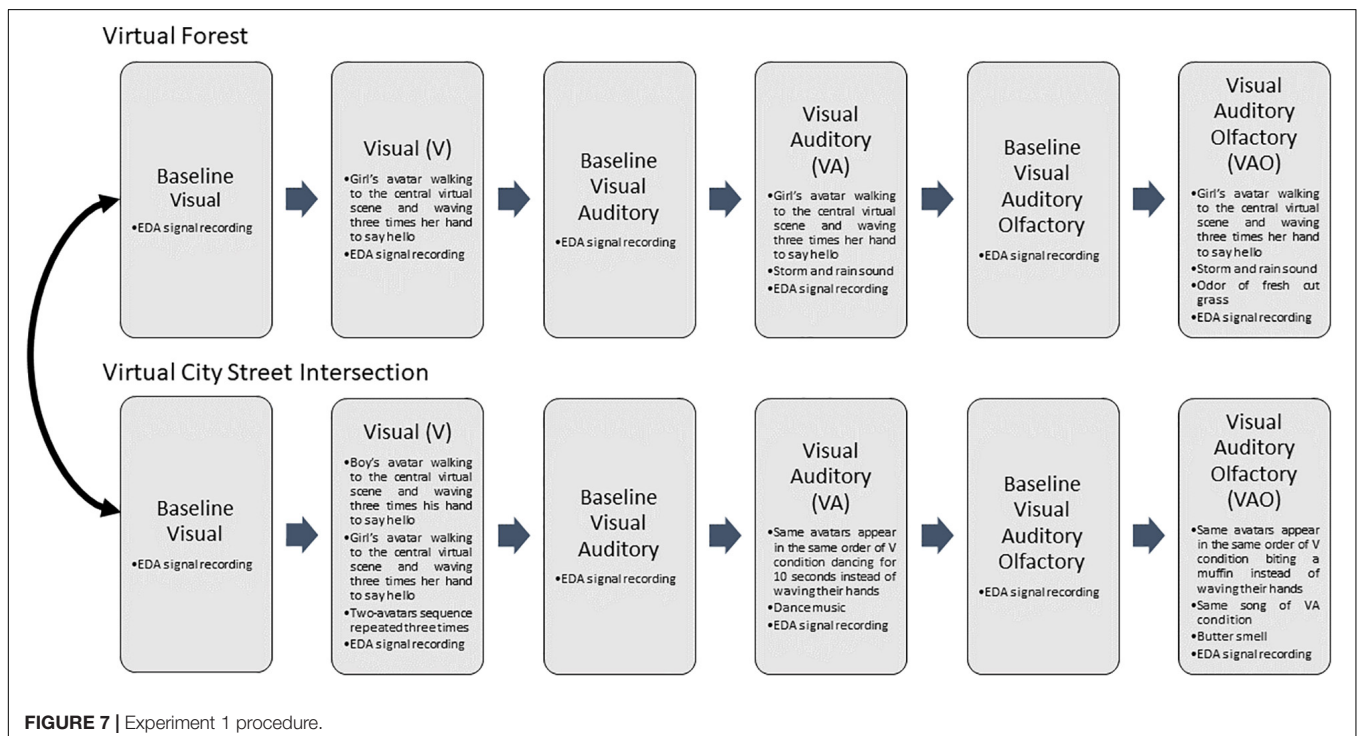
The EDA signals were recorded using Empatica E4 wristband, as in experiment 1,<sup>5</sup> and the physiological data processing and analyses were performed using the method described in the “Physiological Assessment and Data Processing” section of experiment 1.

### The Olfactive System

The system used was the same as that implemented in experiment 1: the Olorama Technology<sup>TM6</sup> wireless freshener.

<sup>5</sup>www.empatica.com

<sup>6</sup>www.olorama.com



**FIGURE 7** | Experiment 1 procedure.

## Experimental Procedure

In experiment 2, the participants only experienced the forest VE as follows: first, as in experiment 1, the family caregivers of the participants were informed about the general objectives of the research, the physiological measure and its device localization, and about the VR system. Second, the Empatica E4 device was shown and placed on the participants' arm of the non-dominant hand before the virtual session. Consequently, the child was accompanied in the CAVE by the researcher and by his or her family caregiver according to the child's needs. The participant was placed in the middle of the virtual room, standing in front of the central surface at 1.5 m. Firstly, 2 min of EDA baseline was recorded in resting and relaxing state. Next, the three stimuli VR experience conditions were presented, recording a 2-min EDA baseline before each one.

## Statistical Analysis

In experiment 1 ( $n = 54$ ), three participants were excluded from the analysis for lack of EDA data due to bad recording: two from the forest VE and one from the city VE. Consequently, the sample size included 52 children for forest VE analysis and 51 in the city VE. In this preliminary stage, we developed four models for each environment (forest VE and city VE) in order to explore the importance of the scenario and each stimuli condition (SC). The first model included all the SC, the second only the visual stimuli, the third only the VA stimuli, and the fourth only the visual, auditive, and olfactive SC. Moreover, we developed two extra models to analyze if they can achieve a performance better than chance. To this extent, we computed a permutation-based test, i.e., we developed two models with city/forest VE input data (all stimuli) with a random output class assignment. The development of the models (parameter tuning and feature selection) used cross-validation with all the samples of the experiment. To compare model performance, we used the output of the classification algorithm without bipolarization, i.e., the probability between 0 and 1 that the model as their true class classified a subject. Due to the Gaussianity of the data ( $p > 0.05$  from the Shapiro–Wilk test with null hypothesis of having a Gaussian sample), we performed a statistical model comparison using the probabilities of the models by applying a one-way ANOVA with Tukey–Kramer correction.

In experiment 2 ( $n = 92$ ): in order to calibrate and test the final model, we used the case of the forest VE and all the SC, increasing the participants to boost the final model and test it. We split the dataset into a training set ( $n = 72$ ) and a test set ( $n = 20$ ). The test set was sliced randomly using the new subjects, but keeping a balanced 50% of each class. The development of the model (parameter tuning and feature selection) used cross-validation with the training set and afterward was applied in the test set that has not been previously used.

To develop the models, we used support vector machine (SVM)-based pattern recognition (Schölkopf et al., 2000) with a leave-one-subject-out (LOSO) cross-validation procedure. For the LOSO scheme, the training set was normalized by subtracting the median value and dividing by the median absolute deviation over each dimension. In each iteration, the validation set

**TABLE 1** | Overview of the performance of the models including total accuracy, Cohen's kappa, permutation test (\*shows significant differences), true positive, and true negative.

Study	Confusion matrix				Features included in the model														
	Acc (%)	Kappa	Permutation test	Visual stimuli condition (V)				Visual and auditive stimuli condition (VA)				Visual, auditive, and olfactive stimuli condition (VAO)							
				TPR (%)	TNR (%)	BL tonic	BL phasic	Tonic	Phasic	Ratio	BL tonic	BL phasic	Tonic	Phasic	Ratio	BL tonic	BL phasic	Tonic	Phasic
Forest—all ( $n = 52$ )	90.38	0.806	*	89.66	91.30	X	X		X		X			X	X		X		
Forest—V ( $n = 52$ )	84.62	0.691	*	82.76	86.96	X	X		X										
Forest—VA ( $n = 52$ )	71.15	0.418	—	72.41	69.57	—	—		—		X			X	—		—		
Forest—VAO ( $n = 52$ )	75.00	0.496	—	75.86	73.91	—	—		—		—			—	X		—		
City—all ( $n = 51$ )	70.59	0.397	—	75.86	63.64										X	X	X	X	
City—V ( $n = 51$ )	68.63	0.323	—	89.66	40.91				X		—			—	—		—		
City—VA ( $n = 51$ )	72.55	0.415	—	89.66	50.00	—	—		—		—			X	—	—	—		
City—VAO ( $n = 51$ )	76.47	0.520	—	79.31	72.73	—	—		—		—			—	—	X	X	X	X

Features included in the models have been marked with an (x), features proposed but not included (i.e., not selected by the automatic feature selection procedure) have been marked with a blank space (), and features excluded a priori of the model have been marked with a dash (—). Acc, accuracy; TPR, true positive rate; TNR, true negative rate; BL, baseline.

consisted of one specific subject and he/she was normalized using the median and deviation of the training set. In particular, we used an optimized C-SVM using a sigmoid kernel function, changing the parameters of cost and gamma using a vector with 15 parameters logarithmically spaced between 0.1 and 1,000. Moreover, we performed a feature selection strategy to explore the relative importance of each feature. A support vector machine recursive feature elimination (SVM-RFE) procedure, in a wrapper approach, was included (RFE was performed on the training set of each fold and we computed the median rank for each feature over all folds). We specifically chose a recently developed, non-linear SVM-RFE, which includes a correlation bias reduction strategy in the feature elimination procedure (Yan and Zhang, 2015). The model was optimized to achieve best Cohen's kappa. The algorithms were implemented using Matlab® R2016a and LIBSVM toolbox (Chang and Lin, 2011).

## RESULTS

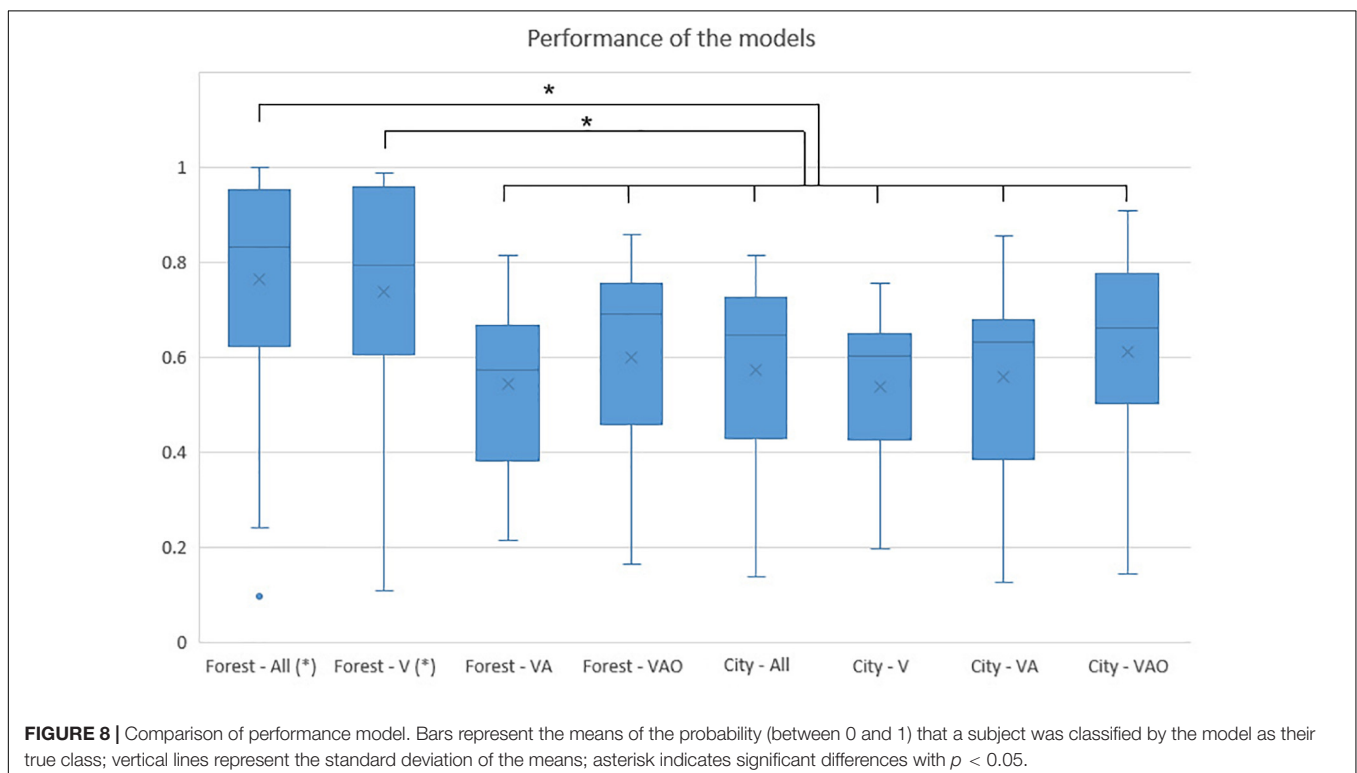
### Experiment 1: Model Comparisons

**Table 1** shows the performance of the eight models performed, considering both VEs and SC. It includes the accuracy of each model, the confusion matrix, and the features included derived to the automatic feature selection procedure. In addition, **Figure 8** shows a comparison of the performance of each model, considering the probability that the model as their true class classified a subject and the significant differences between models were derived from one-way ANOVA using a Tukey–Kramer correction. We included in the ANOVA 2 permutated models

to test if the accuracy is significantly better than chance, where the accuracy is 67.30% for the forest VE and 68.62% for the city VE. The result of the one-way ANOVA shows that there are differences between models ( $p < 0.0001$ ).

The highest accuracy (90.3%, kappa = 0.80) was achieved by experiment 1 including all SC and it presented higher performance than the rest of the models (forest—VA  $p = 0.000$ , forest—VAO  $p = 0.000$ , city—all  $p = 0.000$ , city—V  $p = 0.000$ , city—VA  $p = 0.000$ , city—VAO  $p = 0.002$ ), except with the forest—V SC where no statistical significance was found. The model included four features of the V SC (baseline tonic, baseline phasic, phasic, and ratio), one feature of the VA SC (ratio), and two features of VAO SC (baseline tonic and tonic). The second highest accuracy (84.6%, kappa = 0.69) was achieved by the forest including V SC, and it presented a higher performance than the rest of the models with lower accuracy (forest—VA  $p = 0.000$ , forest—VAO  $p = 0.008$ , city—all  $p = 0.001$ , city—V  $p = 0.000$ , city—VA  $p = 0.000$ , city—VAO  $p = 0.029$ ). The model only included two features (baseline tonic and baseline phasic). Both models have a strongly balanced confusion matrix. These two models presented a performance statistically different than chance (forest—all  $p = 0.000$  and forest—V  $p < 0.0001$ ).

The rest of the models presented accuracy between 68 and 76% and did not present statistically significant differences in terms of performance between them and the permutated models. The model, including the VA stimuli condition of the forest, showed an accuracy of 71.15% (kappa = 0.41) and included the three features (phasic baseline, phasic, and ratio). The model, including VAO SC of the forest, achieved a balanced accuracy of 75.00% (kappa = 0.49), including only the tonic responses in the baseline.



Regarding the city VE, the model included all stimuli conditions, achieved 70.5% ( $\kappa = 0.39$ ) using one feature of VA SC (phasic) and four features of VAO SC (baseline tonic, tonic, and phasic and ratio). The models, including the V and the VA SC, achieved 68.63% ( $\kappa = 0.32$ ) and 72.55% ( $\kappa = 0.41$ ) of accuracy respectively, but with a very bad balance in terms of false positives. The model, including the VAO SC, achieved a balanced 76.47% ( $\kappa = 0.52$ ) of accuracy including four features (baseline phasic, tonic, phasic, and ratio).

## Experiment 2: Development of the Final Model

**Table 2** shows the performance of the final model derived from the forest VE after the increment of the subjects. The validation set ( $n = 72$ ) shows a balanced accuracy of 83.33% ( $\kappa = 0.668$ ). The test set ( $n = 20$ ) achieved 85% of accuracy ( $\kappa = 0.700$ ), recognizing 80% of subjects with sensory dysfunction. The model included one feature of the V SC (phasic), three features of the VA SC (baseline tonic, baseline phasic, and phasic), and one feature of the VAO SC (baseline tonic). In addition, **Figure 9** shows the ROC curve of the performance of the final model, achieving an area under a curve (AUC) of 0.897 in the validation and 0.870 in the test.

## DISCUSSION

The main aim of this study was to discriminate and predict sensory processing, recognizing ASD population versus TD population through the combined use of implicit measure (EDA) and different sensory stimuli in VR. Specifically, two experiments have been run, testing two different VEs, presenting three sensory stimuli conditions each—visual, visual and auditive, and visual, auditive, and olfactive stimuli—and examining EDA changes before and during the presentation of the virtual and the sensory stimuli. The focus has been on sensory processing because there are evidences that it is relatively impaired in the ASD population (Leekam et al., 2007; Tomchek and Dunn, 2007; Baron-Cohen et al., 2009).

The results can be discussed on four levels: (1) the influence of scenarios and stimuli conditions, (2) the role of EDA and the features used, (3) the performance of ASD recognition, and (4) the limitations and further studies.

## The Influence of Scenarios and Stimuli Conditions

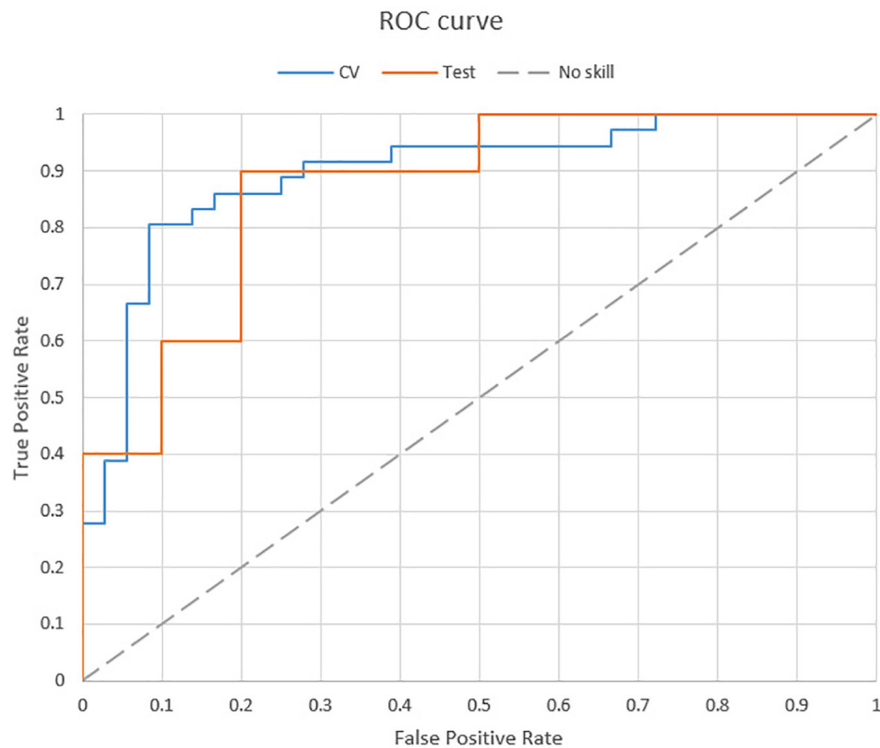
Regarding scenarios, the model developed using the forest VE presented a higher accuracy (forest VE—all, 90.3%) than the model developed using the city VE (city VE—all, 70.59%). Since we used the same set of subjects, the results are not influenced by the individuals' bias. Therefore, a model comparison validated the hypothesis that ASD recognition was higher in the forest VE (forest VE—all vs. city VE—all,  $p = 0.000$ ). Moreover, the permutated test shows that forest—all and forest—V are the models that statistically offer a performance better than chance. This outcome could be due to task characteristics since the response to a greeting is one of the confirmatory symptoms in

**TABLE 2 |** Overview of the performance of the final model including accuracy, AUC, Cohen's kappa, true positives, true negatives, and features included in the models (marked with an X), considering validation and test set.

Study	Confusion matrix					Features included in the model											
	Acc (%)	Kappa	AUC	TPR (%)	TNR (%)	Visual phase (V)				Visual and auditive phase (VA)				Visual, auditive and olfactive phase (VAO)			
						BL tonic	BL phasic	Tonic	Phasic	Ratio	BL tonic	BL phasic	Tonic	Phasic	Ratio	BL tonic	BL phasic
Forest—validation set ( $n = 72$ )	83.33	0.668	0.897	86.11	80.55						X						
Forest—test set ( $n = 20$ )	85.00	0.700	0.870	80.00	90.00							X					X

BL represents the responses recorded in the baseline.





**FIGURE 9 |** ROC curve of the final model.

ASD. In addition, several previous studies showed the influence of nature scenes in reducing arousal (Liszio et al., 2018; White et al., 2018). Therefore, the forest VE can be assumed as a more relaxed environment than the city VE. Since EDA is highly affected by the arousal (Picard et al., 2016), the results suggested that a natural and relaxed environment as a forest VE could be a better scenario to detect changes in the ASD population due to sensory processing dysfunctions. In addition to the increase of arousal derived from city VE, the avatars imitation task provoked a physical activity in the subject that could affect arousal, decreasing the recognition performance of models, due to an arousal saturating effect; hence, the results supported the use of a low-arousal natural environment and non-physical activities to increase the performance recognition of models using EDA.

Regarding stimuli conditions, the model developed in the forest VE with all the stimuli conditions achieved 5.78% of accuracy more than the forest with only visual stimulation, but it did not show statistical differences. Both models presented higher accuracy and performance than the rest of the models, including the permuted one. However, the model developed in the city VE used only one feature of V SC, three features of VA SC, and four features of VAO SC. Therefore, even though the exploratory analysis performed in forest VE suggested that VA and VAO did not play an important role in the ASD recognition in comparison with V, the feature selection for the final model showed high reliance on the multimodal sensory condition since four out of five of the features selected were from VA and VAO stimuli.

The hypothesis that increasing sensory modalities would have contributed to better ASD recognition through EDA is partially confirmed by the final model.

## The Role of EDA and the Features Used

To our knowledge, we proposed the first supervised ML model using EDA for ASD population [see Hyde et al. (2019) for a review of ASD models recognition]. Our results were in accordance with previous research that showed that ASD is associated with the autonomic nervous system and can be measured using EDA (Miller et al., 2001; Rogers and Ozonoff, 2005; Schoen et al., 2009; Bujnakova et al., 2016). However, other researches did not find differences in EDA levels in response to sensory stimuli in the ASD population (e.g., Zahn et al., 1987; Rogers and Ozonoff, 2005; McCormick et al., 2014). The level of recognition of the presented models represents a new step in the use of the autonomic nervous system as a biomarker for ASD recognition. In addition, the CDA analysis showed a valid signal processing method to extract valuable features from EDA. The phasic responses of the subjects in the two SCs (V and VA) are included in the feature selection of the final model and in many of the exploratory model comparison. A baseline is also a very important part of the stimuli since the baseline responses of VA (tonic and phasic) and VAO (phasic) SCs were included in the final model. Moreover, the baseline responses were also included in forest VE—all and forest VE—V models of forest VE. In this regard, it should be noted that variations in the phasic and tonic components, related to changes in emotional

arousal, have been reported by studies carried out in different experimental paradigms (Kreibig, 2010). The relevant role of the baseline was in accordance with previous research that suggested that the participants are likely to be hyper- or hypo-responders independent of any effects of stimuli (Braithwaite et al., 2013). Moreover, the role of baselines could be especially important in research on sensory processing disorders as in ASD.

## The Performance of ASD Recognition

Regarding the final model on ASD recognition, the validation set using 72 subjects achieved 83.33% of accuracy (kappa: 0.668, AUC: 0.897), including 86.11% of true positives. Moreover, we tested the model in a set of 20 subjects (10 ASD and 10 controls) recruited in a second phase, and the model achieved 85% accuracy (kappa: 0.700, AUC: 0.870). The results presented perform a new step in ASD recognition since, to our knowledge, we presented the first ASD-supervised ML recognition model using EDA and multimodal VR. Moreover, the methodology presented some advantages in contrast to previous research. Li et al. (2017) presented an analysis using kinematics recognizing ASD in adults achieving 86.7% accuracy ( $n = 30$ ). Liu et al. (2016) developed a model using eye tracking to recognize ASD children based on face processing, achieving 88.51% of accuracy ( $n = 58$ ). Nakai et al. (2017) presented an ASD model recognition in children using voice analysis, achieving 76% of accuracy ( $n = 30$  ASD,  $n = 51$  TD). All of them validated their models using cross-validation procedures and used ecological biomarkers to recognize ASD. Contrarily, the presented model achieved the same (or more) level of accuracy, but using a broader sample size and, moreover, applying the model to a new test set that was not used before, simulating a real application. It supposes a new step forward in order to develop scalable clinical applications of ASD recognition models. On the other hand, previous research by Chen et al. (2015) showed a very large study ( $n = 252$ ) with a very high accuracy (91%) using fMRI. In contrast to this approach, we proposed an ecological environment and instrumentation using VR and EDA wristband sensor. This ecological approach is particularly important in the field of ASD and can offer cheaper and quicker clinical diagnostic models in the future.

## Limitations and Future Studies

Although this study did a step forward in the field of ASD sensory processing assessment, it presented some limitations regarding sample characteristics, specific ASD symptoms and their related measures, and VEs.

First, the participants were from 4 to 7 years old and selected ASD children received, according to their symptomatology and age, a previous ASD diagnosis through the module 1 of the ADOS-2 questionnaire that is addressed to infants older than 31 months of age but who do not use phrase language consistently. Nonetheless, it has been decided to test only participants who pertained to this class and characteristics to control and ensure results, but these narrow criteria limit the generalization of findings.

Second, the present study mostly focused on ASD sensory processing although it is not a core ASD symptom for

diagnostic manuals, such as DSM-5 and ICD-10. Furthermore, regarding VEs, at the first time, children might experience them as astonishing and impressive (Novelty effect, Clark, 1983; Gravetter and Forzano, 2018); for this reason, in the first part of each study, there might be a common effect on EDA metrics, especially in forest-V and city-V conditions. However, this artificial activation arousal, that is due to the sense of being physically present in a VE despite the certainty of not being physically there, decreases as the familiarity with the virtual world and device increases. Third, the sample size was restricted and not matched on socio-demographics, limiting the generalization of the model outcomes. In accordance with limitations, future works are needed in order to develop an objective method for the assessment of sensory processing in the ASD population. Future studies must, first, include a broader sample size with further control and matching on sociodemographics. Socioeconomic status is recommended in order to avoid misleading model outcomes based on other metrics far from ASD presence or absence (Delobel-Ayoub et al., 2015). Second, ASD individuals should be diagnosed by the five modules of ADOS-2 questionnaire to test whether the results presented here may be generally replicated in all age range and linguistic ability clusters. Moreover, in conjunction with sensory measures, the inclusion of core symptom analyses in VR is suggested, for example, repetitive and stereotypical behaviors, and communication and social abilities. Biomarkers that could be relevant for this purpose are eye tracking, body movement analysis, and EEG (Loth et al., 2016); indeed eye tracking glasses and RGBD cameras for body movement analysis might be included in future studies on current VR experiences in order to enhance model strength and accuracy. Furthermore, to discern impaired sensory processing, the present study involved three VR conditions (visual, auditive, and olfactive) and it could be interesting to add a fourth condition about haptic processing since it enhances immersion in the VE, providing a more ecological and realistic experience (Slater and Wilbur, 1997). Finally, some adjustments of the virtual content might bring more sense of presence to the participants, such as the introduction, in both VA and VAO stimuli conditions, of auditive stimulation consistent with the avatar that is waving the hand to say hello.

## CONCLUSION

Sensory processing is a relevant ability in information processing, allowing adapting behavioral responses to the environment (Miller et al., 2007). ASD show hyper-sensitiveness (over-responsiveness) to VA stimuli and hypo-sensitiveness (under-responsiveness) to olfactive stimuli (Tomchek and Dunn, 2007; Baron-Cohen et al., 2009; Dudova et al., 2011; Ashwin et al., 2014; Tomchek et al., 2014). The hyper-hypo sensitiveness to sensory stimuli can generate an alteration in information processing, affecting cognitive and social responses in daily life situations. Traditional ASD assessment, based on semi-structured behavioral task observations on laboratory settings

and structured interviews, does not take into account the dysfunctional sensory processing in real life. According to the results, current studies have shown that it is possible to obtain biomarkers for ASD classification using a CP paradigm based on implicit brain processes, measured through psychophysiological signals and the subjects' behavior, while exposed to complex social conditions using VR interfaces. The ASD classification using biomarkers, along with traditional assessment, could enhance knowledge on the development of relevant specific treatments.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Polytechnic University of Valencia. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## REFERENCES

- Alcañiz, M. L., Olmos-raya, E., and Abad, L. (2019). Uso de entornos virtuales para trastornos del neurodesarrollo: una revisión del estado del arte y agenda futura. *Medicina* 79, 77–81.
- Allen, R., Davis, R., and Hill, E. (2013). The effects of autism and alexithymia on physiological and verbal responsiveness to music. *J. Autism Dev. Disord.* 43, 432–444. doi: 10.1007/s10803-012-1587-8
- American Psychiatric Association, (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Arlington, TX: American Psychiatric Association.
- Anagnostou, E., Zwaigenbaum, L., Szatmari, P., Fombonne, E., Fernandez, B. A., Woodbury-Smith, M., et al. (2014). Autism spectrum disorder: advances in evidence-based practice. *CMAJ* 186, 509–519.
- Ashwin, C., Chapman, E., Howells, J., Rhydderch, D., Walker, I., and Baron-Cohen, S. (2014). Enhanced olfactory sensitivity in autism spectrum conditions. *Mol. Autism* 5:53. doi: 10.1186/2040-2392-5-53
- Baron-Cohen, S. (1990). Autism: a specific cognitive disorder of 'mind-blindness'. *Int. Rev. Psychiatry* 2, 81–90.
- Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., and Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1377–1383. doi: 10.1098/rstb.2008.0337
- Barry, R. J., and James, A. L. (1988). Coding of stimulus parameters in autistic, retarded, and normal children: evidence for a two-factor theory of autism. *Int. J. Psychophysiol.* 6, 139–149. doi: 10.1016/0167-8760(88)90045-1
- Bekele, E., Crittendon, J., Zheng, Z., Swanson, A., Weitlauf, A., Warren, Z., et al. (2014). Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. *J. Autism Dev. Disord.* 44, 1641–1650. doi: 10.1007/s10803-014-2035-8
- Ben Shalom, D., Mostofsky, S. H., Hazlett, R. L., Goldberg, M. C., Landa, R. J., Faran, Y., et al. (2006). Normal physiological emotions but differences in expression of conscious feelings in children with high-functioning autism. *J. Autism Dev. Disord.* 36, 395–400. doi: 10.1007/s10803-006-0077-2
- Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 80–91. doi: 10.1016/j.jneumeth.2010.04.028

## AUTHOR CONTRIBUTIONS

All authors have contributed to the manuscript as follows: MA and LA designed the study and supervised the whole study. EO and MS recruited and assessed the eligible subjects. JM-M conducted the odor and statistical analyses. IC, GT, MM, MS, JM-M, and JH-T wrote the original manuscript and MA and LA revised the manuscript. All authors assisted in the revision process, read, and approved the final manuscript.

## FUNDING

This work was supported by the Spanish Ministry of Economy, Industry, and Competitiveness-funded project "Immersive Virtual Environment for the Evaluation and Training of Children with Autism Spectrum Disorder: T Room" (IDI-20170912) and by the Generalitat Valenciana-funded project REBRAND (PROMETEU/2019/105).

## ACKNOWLEDGMENTS

We thank Zayda Ferrer Lluch for the development of virtual reality environments.

- Bernardini, S., Porayska-Pomsta, K., and Smith, T. J. (2014). ECHOES: an intelligent serious game for fostering social communication in children with autism. *Inf. Sci.* 264, 41–60. doi: 10.1016/j.ins.2013.10.027
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/s15327965pli1302\_01
- Bogdashina, O. (2016). *Sensory Perceptual Issues in Autism and Asperger Syndrome: Different Sensory Experiences-Different Perceptual Worlds*. London: Jessica Kingsley Publishers.
- Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 752–762. doi: 10.1038/nrn3122
- Boucsein, W. (2012). *Electrodermal Activity*. Berlin: Springer Science & Business Media.
- Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1017–1034.
- Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470
- Bujnakova, I., Ondrejka, I., Mestanik, M., Visnovcova, Z., Mestanikova, A., Hrtanek, I., et al. (2016). Autism spectrum disorder is associated with autonomic underarousal. *Physiol. Res.* 65(Suppl. 5), S673–S682.
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27.
- Chang, M. C., Parham, L. D., Blanche, E. I., Schell, A., Chou, C. P., Dawson, M., et al. (2012). Autonomic and behavioral responses of children with autism to auditory stimuli. *Am. J. Occup. Ther.* 66, 567–576. doi: 10.5014/ajot.2012.004242
- Chaytor, N., Schmitter-Edgecombe, M., and Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Arch. Clin. Neuropsychol.* 21, 217–227. doi: 10.1016/j.acn.2005.12.002
- Chen, C. P., Keown, C. L., Jahedi, A., Nair, A., Pflieger, M. E., Bailey, B. A., et al. (2015). Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *Neuroimage Clin.* 8, 238–245. doi: 10.1016/j.nicl.2015.04.002

- Chita-Tegmark, M. (2016). Attention allocation in ASD: a review and meta-analysis of eye-tracking studies. *Rev. J. Autism Dev. Disord.* 3, 209–223. doi: 10.1007/s40489-016-0077-x
- Clark, R. E. (1983). Reconsidering research on learning from media. *Rev. Educ. Res.* 53, 445–459. doi: 10.1097/ACM.0000000000000436
- Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. *Handb. Psychophysiol.* 2, 200–223.
- Delobel-Ayoub, M., Ehlinger, V., Klapouszczak, D., Maffre, T., Raynaud, J. P., Delpierre, C., et al. (2015). Socioeconomic disparities and prevalence of autism spectrum disorders and intellectual disability. *PLoS One* 10:e0141964. doi: 10.1371/journal.pone.0141964
- Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78
- Dudova, I., Vodicka, J., Havlovicova, M., Sedlacek, Z., Urbanek, T., and Hrdlicka, M. (2011). Odor detection threshold, but not odor identification, is impaired in children with autism. *Eur. Child Adolesc. Psychiatry* 20, 333–340. doi: 10.1007/s00787-011-0177-1
- Dunn, W. (2014). *Child Sensory Profile–2 User's Manual*. Bloomington, MN: Pearson.
- Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research*. New York, NY: Dryden Press.
- Fagius, J., and Wallin, B. G. (1980). Sympathetic reflex latencies and conduction velocities in normal man. *J. Neurol. Sci.* 47, 433–448. doi: 10.1016/0022-510x(80)90098-2
- Fenning, R. M., Baker, J. K., Baucom, B. R., Erath, S. A., Howland, M. A., and Moffitt, J. (2017). Electrodermal variability and symptom severity in children with autism spectrum disorder. *J. Autism Dev. Disord.* 47, 1062–1072. doi: 10.1007/s10803-016-3021-0
- Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., et al. (2019). A meta-analysis of procedures to change implicit measures. *J. Pers. Soc. Psychol.* 117, 522–559. doi: 10.1037/pspa0000160
- Francis, K. (2005). Autism interventions: a critical update. *Dev. Med. Child Neurol.* 47, 493–499. doi: 10.1111/j.1469-8749.2005.tb01178.x
- Franzen, M. D., and Wilhelm, K. L. (1996). "Conceptual foundations of ecological validity in neuropsychological assessment," in *Ecological Validity of Neuropsychological Testing*, eds R. J. Sbordone, and C. J. Long (Boca Raton, FL: St Lucie Press), 91–112.
- Friston, K. J., Stephan, K. E., Montague, R., and Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 1, 148–158. doi: 10.1016/S2215-0366(14)70275-5
- Gillberg, C., and Rasmussen, P. (1994). Brief report: four case histories and a literature review of Williams syndrome and autistic behavior. *J. Autism Dev. Disord.* 24, 381–393. doi: 10.1007/bf02172235
- Goldstein, S., Naglieri, J. A., and Ozonoff, S. (eds). (2009). *Assessment of Autism Spectrum Disorder*. New York, NY: The Guilford Press.
- Golestan, S., Soleiman, P., and Moradi, H. (2018). A comprehensive review of technologies used for screening, assessment, and rehabilitation of autism spectrum disorder. *arXiv [Preprint]*.
- Gravetter, F. J., and Forzano, L. A. B. (2018). *Research Methods for the Behavioral Sciences*. Boston, MA: Cengage Learning.
- Großekathöfer, U., Manyakov, N. V., Mihajlović, V., Pandina, G., Skalkin, A., Ness, S., et al. (2017). Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis. *Front. Neuroinform.* 11:9. doi: 10.3389/fninf.2017.00009
- Hirstein, W., Iversen, P., and Ramachandran, V. S. (2001). Autonomic responses of autistic children to people and objects. *Proceedings* 268, 1883–1888. doi: 10.1098/rspb.2001.1724
- Hubert, B. E., Wicker, B., Monfardini, E., and Deruelle, C. (2009). Electrodermal reactivity to emotion processing in adults with autistic spectrum disorders. *Autism* 13, 9–19. doi: 10.1177/1362361308091649
- Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., et al. (2019). Applications of supervised machine learning in autism spectrum disorder research: a review. *Rev. J. Autism Dev. Disord.* 6, 128–146. doi: 10.1007/s40489-019-00158-x
- Joseph, R. M., Ehrman, K., McNally, R., and Keehn, B. (2008). Affective response to eye contact and face recognition ability in children with ASD. *J. Int. Neuropsychol. Soc.* 14, 947–955. doi: 10.1017/S1355617708081344
- Kandalaf, M. R., Didehban, N., Krawczyk, D. C., Allen, T. T., and Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *J. Autism Dev. Disord.* 43, 34–44. doi: 10.1007/s10803-012-1544-6
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010
- Kylliäinen, A., and Hietanen, J. K. (2006). Skin conductance responses to another person's gaze in children with autism. *J. Autism Dev. Disord.* 36, 517–525. doi: 10.1007/s10803-006-0091-4
- Kylliäinen, A., Wallace, S., Coutanche, M. N., Leppanen, J. M., Cusack, J., Bailey, A. J., et al. (2012). Affective-motivational brain responses to direct gaze in children with autism spectrum disorder. *J. Child Psychol. Psychiatry* 53, 790–797. doi: 10.1111/j.1469-7610.2011.02522.x
- Ledoux, K., Coderre, E., Bosley, L., Buz, E., Gangopadhyay, I., and Gordon, B. (2016). The concurrent use of three implicit measures (eye movements, pupillometry, and event-related potentials) to assess receptive vocabulary knowledge in normal adults. *Behav. Res. Methods* 48, 285–305. doi: 10.3758/s13428-015-0571-6
- Leekam, S. R., Nieto, C., Libby, S. J., Wing, L., and Gould, J. (2007). Describing the sensory abnormalities of children and adults with autism. *J. Autism Dev. Disord.* 37, 894–910. doi: 10.1007/s10803-006-0218-7
- Levy, A., and Perry, A. (2011). Outcomes in adolescents and adults with autism: a review of the literature. *Res. Autism Spectr. Disord.* 5, 1271–1282. doi: 10.1016/j.rasd.2011.01.023
- Li, B., Sharma, A., Meng, J., Purushwalkam, S., and Gowen, E. (2017). Applying machine learning to identify autistic adults using imitation: an exploratory study. *PLoS One* 12:e0182652. doi: 10.1371/journal.pone.0182652
- Lieberman, M. D. (2010). Social cognitive neuroscience. *Handb. Soc. Psychol.* 5, 143–193.
- Lisio, S., Graf, L., and Masuch, M. (2018). The relaxing effect of virtual nature: immersive technology provides relief in acute stress situations. *Annu. Rev. Cyberther. Telemed.* 16, 87–93.
- Liu, W., Li, M., and Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework. *Autism Res.* 9, 888–898. doi: 10.1002/aur.1615
- Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., and Pickles, A. (2006). Autism from 2 to 9 years of age. *Arch. Gen. Psychiatry* 63, 694–701.
- Lord, C., Rutter, M., DiLavore, P. C., and Risi, S. A. (2001). *Autism Diagnostic Observation Schedule*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P. C., and Risi, S. A. (1999). *Diagnostic Observation Schedule-WPS (ADOS-WPS)*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/bf02172145
- Loth, E., Spooren, W., Ham, L. M., Isaac, M. B., Auriche-Benichou, C., Banaschewski, T., et al. (2016). Identification and validation of biomarkers for autism spectrum disorders. *Nat. Rev. Drug Discov.* 15, 70–73.
- Louwerse, A., van der Geest, J. N., Tulen, J. H. M., van der Ende, J., Van Gool, A. R., Verhulst, F. C., et al. (2013). Effects of eye gaze directions of facial images on looking behaviour and autonomic responses in adolescents with autism spectrum disorders. *Res. Autism Spectr. Disord.* 7, 1043–1053. doi: 10.1016/j.rasd.2013.04.013
- Lydon, S., Healy, O., Reed, P., Mulhern, T., Hughes, B. M., and Goodwin, M. S. (2016). A systematic review of physiological reactivity to stimuli in autism. *Dev. Neurorehabil.* 19, 335–355. doi: 10.3109/17518423.2014.971975
- McCarthy, C., Pradhan, N., Redpath, C., and Adler, A. (2016). "Validation of the Empatica E4 wristband," in *Proceedings of the 2016 IEEE EMBS International Student Conference (ISC)* (Ottawa: IEEE).
- McCormick, C., Hessel, D., Macari, S. L., Ozonoff, S., Green, C., and Rogers, S. J. (2014). Electrodermal and behavioral responses of children with autism spectrum disorders to sensory and repetitive stimuli. *Autism Res.* 7, 468–480. doi: 10.1002/aur.1382



- Miller, L. J., Anzalone, M. E., Lane, S. J., Cermak, S. A., and Osten, E. T. (2007). Concept evolution in sensory integration: a proposed nosology for diagnosis. *Am. J. Occup. Ther.* 61, 135–140. doi: 10.5014/ajot.61.2.135
- Miller, L. J., Reisman, J. E., McIntosh, D. N., and Simon, J. (2001). “An ecological model of sensory modulation: performance of children with Fragile X syndrome, autistic disorder, attention-deficit/ hyperactivity disorder, and sensory modulation dysfunction,” in *Understanding the Nature of Sensory Integration with Diverse Populations*, eds S. S. Roley, E. I. Blanche, and R. C. Schaff (San Antonio, TX: Therapy Skill Builders).
- Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends Cogn. Sci.* 16, 72–80. doi: 10.1016/j.tics.2011.11.018
- Möricke, E., Buitelaar, J. K., and Rommelse, N. N. (2016). Do we need multiple informants when assessing autistic traits? The degree of report bias on offspring, self, and spouse ratings. *J. Autism Dev. Disord.* 46, 164–175. doi: 10.1007/s10803-015-2562-y
- Murphy, D., and Spooen, W. (2012). EU-AIMS: a boost to autism research. *Nat. Rev. Drug Discov.* 11, 815–816. doi: 10.1038/nrd3881
- Nakai, Y., Takiguchi, T., Matsui, G., Yamaoka, N., and Takada, S. (2017). Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists. *Percept. Mot. Skills* 124, 961–973. doi: 10.1177/0031512517716855
- Nikula, R. (1991). Psychological correlates of nonspecific skin conductance responses. *Psychophysiology* 28, 86–90. doi: 10.1111/j.1469-8986.1991.tb03392.x
- Nosek, B. A., Hawkins, C. B., and Frazier, R. S. (2011). Implicit social cognition: from measures to mechanisms. *Trends Cogn. Sci.* 15, 152–159. doi: 10.1016/j.tics.2011.01.005
- Palkovitz, R. J., and Wiesenfeld, A. R. (1980). Differential autonomic responses of autistic and normal children. *J. Autism Dev. Disord.* 10, 347–360. doi: 10.1007/bf02408294
- Parsons, S. (2016). Authenticity in virtual reality for assessment and intervention in autism: a conceptual review. *Educ. Res. Rev.* 19, 138–157. doi: 10.1016/j.edurev.2016.08.001
- Parsons, T. D. (ed.). (2016). “Telemedicine, mobile, and internet-based neurocognitive assessment,” in *Clinical Neuropsychology and Technology* (Cham: Springer), 99–111. doi: 10.1007/978-3-319-31075-6\_6
- Paulhus, D. L. (1991). “Measurement and control of response bias,” in *Measures of Social Psychological Attitudes, Measures of Personality and Social Psychological Attitudes*, Vol. 1, eds J. P. Robinson, P. R. Shaver, and L. S. Wrightsman (Cambridge, MA: Academic Press), 17–59. doi: 10.1016/b978-0-12-590241-0.50006-x
- Picard, R. W., Fedor, S., and Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emot. Rev.* 8, 62–75. doi: 10.1177/1754073914565517
- Reaven, J. A., Hepburn, S. L., and Ross, R. G. (2008). Use of the ADOS and ADI-R in children with psychosis: importance of clinical judgment. *Clin. Child Psychol. Psychiatry* 13, 81–94. doi: 10.1177/1359104507086343
- Redish, D., and Gordon, J. (2016). *Computational Psychiatry: New Perspectives on Mental Illness (Strüngmann Forum Reports)*. Cambridge, MA: The MIT Press.
- Riby, D. M., Whittle, L., and Doherty-Sneddon, G. (2012). Physiological reactivity to faces via live and video-mediated communication in typical and atypical development. *J. Clin. Exp. Neuropsychol.* 34, 385–395. doi: 10.1080/13803395.2011.645019
- Rogers, S. J., and Ozonoff, S. (2005). Annotation: what do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *J. Child Psychol. Psychiatry* 46, 1255–1268. doi: 10.1111/j.1469-7610.2005.01431.x
- Schmidt, L., Kirchner, J., Strunz, S., Brożus, J., Ritter, K., Roepke, S., et al. (2015). Psychosocial functioning and life satisfaction in adults with autism spectrum disorder without intellectual impairment. *J. Clin. Psychol.* 71, 1259–1268. doi: 10.1002/jclp.22225
- Schoen, S. A., Miller, L. J., Brett-Green, B. A., and Nielsen, D. M. (2009). Physiological and behavioral differences in sensory processing: a comparison of children with autism spectrum disorder and sensory modulation disorder. *Front. Integr. Neurosci.* 3:29. doi: 10.3389/fnro.07.029.2009
- Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. (2000). New support vector algorithms. *Neural. Comput.* 12, 1207–1245. doi: 10.1162/089976600300015565
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138
- Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence* 6, 603–616. doi: 10.1162/pres.1997.6.6.603
- Stevens, S., and Gruzelier, J. (1984). Electrodermal activity to auditory stimuli in autistic, retarded, and normal children. *J. Autism Dev. Disord.* 14, 21245–21260.
- Tomchek, S. D., and Dunn, W. (2007). Sensory processing in children with and without autism: a comparative study using the short sensory profile. *Am. J. Occup. Ther.* 61, 190–200. doi: 10.5014/ajot.61.2.190
- Tomchek, S. D., Huebner, R. A., and Dunn, W. (2014). Patterns of sensory processing in children with an autism spectrum disorder. *Res. Autism Spectr. Disord.* 8, 1214–1224. doi: 10.1016/j.rasd.2014.06.006
- Valenza, G., and Scilingo, E. P. (2013). *Autonomic Nervous System Dynamics for Mood and Emotional-State Recognition: Significant Advances in Data Acquisition, Signal Processing and Classification*. Berlin: Springer Science & Business Media.
- van Engeland, H., Roelofs, J. W., Verbaten, M. N., and Slangen, J. L. (1991). Abnormal electrodermal reactivity to novel visual stimuli in autistic children. *Psychiatry Res.* 38, 27–38. doi: 10.1016/0165-1781(91)90050-y
- Van Hecke, A. V., Stevens, S., Carson, A. M., Karst, J. S., Dolan, B., Schohl, K., et al. (2015). Measuring the plasticity of social approach: a randomized controlled trial of the effects of the PEERS intervention on EEG asymmetry in adolescents with autism spectrum disorders. *J. Autism Dev. Disord.* 45, 316–335. doi: 10.1007/s10803-013-1883-y
- Venables, P. H., and Christie, M. J. (1980). “Electrodermal activity,” in *Techniques in Psychophysiology*, Vol. 54, eds I. Martin, and P. H. Venables (Chichester: Wiley), 3–67.
- Volkmar, F. R., State, M., and Klin, A. (2009). Autism and autism spectrum disorders: diagnostic issues for the coming decade. *J. Child Psychol. Psychiatry* 50, 108–115. doi: 10.1111/j.1469-7610.2008.02010.x
- Wang, X. J., and Krystal, J. H. (2014). Computational psychiatry. *Neuron* 84, 638–654. doi: 10.1016/j.neuron.2014.10.018
- Wang, Y., Hensley, M. K., Tasman, A., Sears, L., Casanova, M. F., and Sokhadze, E. M. (2016). Heart rate variability and skin conductance during repetitive TMS course in children with autism. *Appl. Psychophysiol. Biofeedback* 41, 47–60. doi: 10.1007/s10484-015-9311-z
- White, M. P., Yeo, N. L., Vassiljev, P., Lundstedt, R., Wallergård, M., Albin, M., et al. (2018). A prescription for “nature”—the potential of using virtual nature in therapeutics. *Neuropsychiatr. Dis. Treat.* 14, 3001–3013. doi: 10.2147/ndt.s179038
- White, S. W., Mazefsky, C. A., Dichter, G. S., Chiu, P. H., Richey, J. A., and Ollendick, T. H. (2014). Social-cognitive, physiological, and neural mechanisms underlying emotion regulation impairments: understanding anxiety in autism spectrum disorder. *Int. J. Dev. Neurosci.* 39, 22–36. doi: 10.1016/j.ijdevneu.2014.05.012
- Wing, L., Gould, J., and Gillberg, C. (2011). Autism spectrum disorders in the DSM-IV: better or worse than the DSM-IV? *Res. Dev. Disabil.* 32, 768–773. doi: 10.1016/j.ridd.2010.11.003
- World Health Organization [WHO] (2019). *Autism Spectrum Disorders*. Available at: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders> (accessed November 7, 2019).
- Yan, K., and Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens. Actuators B Chem.* 212, 353–363. doi: 10.1016/j.snb.2015.02.025
- Zahn, T. P., Rumsey, J. M., and Van Kammen, D. P. (1987). Autonomic nervous system activity in autistic, schizophrenic, and normal men: effects of stimulus significance. *J. Abnorm. Psychol.* 96, 135–144. doi: 10.1037/0021-843x.96.2.135

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

Copyright © 2020 Alcañiz Raya, Chicchi Giglioli, Marín-Morales, Higuera-Trujillo, Olmos, Minissi, Teruel Garcia, Sirera and Abad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review

**Cosimo Tuena<sup>1,2\*</sup>, Elisa Pedrolì<sup>1,3</sup>, Pietro Davide Trimarchi<sup>4</sup>, Alessia Gallucci<sup>4</sup>, Mattia Chiappini<sup>1</sup>, Karine Goulene<sup>5</sup>, Andrea Gaggioli<sup>1,2</sup>, Giuseppe Riva<sup>1,2</sup>, Fabrizia Lattanzio<sup>6</sup>, Fabrizio Giunco<sup>4</sup> and Marco Stramba-Badiale<sup>5</sup>**

<sup>1</sup> Applied Technology for Neuro-Psychology, IRCCS Istituto Auxologico Italiano, Milan, Italy, <sup>2</sup> Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy, <sup>3</sup> Faculty of Psychology, University of eCampus, Novedrate, Italy, <sup>4</sup> IRCCS Fondazione Don Carlo Gnocchi, Milan, Italy, <sup>5</sup> Department of Geriatrics and Cardiovascular Medicine, IRCCS Istituto Auxologico Italiano, Milan, Italy, <sup>6</sup> Scientific Direction, IRCCS INRCA, Ancona, Italy

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

Hasan Ayaz,  
Drexel University, United States  
Diogo Morais,  
Lusophone University of Humanities  
and Technologies, Portugal

### \*Correspondence:

Cosimo Tuena  
cosimo.tuena@unicatt.it

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 28 October 2019

**Accepted:** 02 March 2020

**Published:** 08 April 2020

### Citation:

Tuena C, Pedrolì E, Trimarchi PD, Gallucci A, Chiappini M, Goulene K, Gaggioli A, Riva G, Lattanzio F, Giunco F and Stramba-Badiale M (2020) Usability Issues of Clinical and Research Applications of Virtual Reality in Older People: A Systematic Review. *Front. Hum. Neurosci.* 14:93. doi: 10.3389/fnhum.2020.00093

Aging is a condition that may be characterized by a decline in physical, sensory, and mental capacities, while increased morbidity and multimorbidity may be associated with disability. A wide range of clinical conditions (e.g., frailty, mild cognitive impairment, metabolic syndrome) and age-related diseases (e.g., Alzheimer's and Parkinson's disease, cancer, sarcopenia, cardiovascular and respiratory diseases) affect older people. Virtual reality (VR) is a novel and promising tool for assessment and rehabilitation in older people. Usability is a crucial factor that must be considered when designing virtual systems for medicine. We conducted a systematic review with Preferred Reporting Items for Systematic reviews and Meta-analysis (PRISMA) guidelines concerning the usability of VR clinical systems in aging and provided suggestions to structure usability piloting. Findings show that different populations of older people have been recruited to mainly assess usability of non-immersive VR, with particular attention paid to motor/physical rehabilitation. Mixed approach (qualitative and quantitative tools together) is the preferred methodology; technology acceptance models are the most applied theoretical frameworks, however senior adapted models are the best within this context. Despite minor interaction issues and bugs, virtual systems are rated as usable and feasible. We encourage usability and user experience pilot studies to ameliorate interaction and improve acceptance and use of VR clinical applications in older people with the aid of suggestions (VR-USOP) provided by our analysis.

**Keywords:** aging, assessment, rehabilitation, usability, user-experience, virtual reality

## INTRODUCTION

Life expectancy is rapidly increasing and is expected to rise in the years to come, thereby creating an aging population. However, a significant proportion of older people may develop frailty, multi-morbidity, and disability causing a significant impact both on their quality of life and also on health care and social costs (Lutz et al., 2008; World Health Organization, 2015). Aging is associated with physiological changes (e.g., apoptosis, senescence, inflammation) that may lead to

systemic alterations (Flatt, 2012). This potential decline may involve sensory, mental, and physical functioning thus leading to increased morbidity, multi-morbidity, disability, and mortality (World Health Organization, 2015). On the other hand, motor skills, visual, hearing, proprioception, and cognitive abilities (e.g., memory) may be reduced even in healthy older people (Kuehn et al., 2017). In addition, aging hampers psychosocial well-being by adding new developmental tasks or situations (e.g., isolation; Steptoe et al., 2015). In particular, the prevalence of Alzheimer's disease, cancer, chronic obstructive pulmonary disease, maculopathy, osteoarthritis, osteopenia, Parkinson's disease, periodontitis, rheumatoid arthritis, sarcopenia, cardiovascular diseases, and type 2 diabetes increases with age (Tolosa et al., 2006; Dubois et al., 2010; Marengoni et al., 2011; Edwards et al., 2015; Steenman and Lande, 2017; Yakaryilmaz and Öztürk, 2017; Franceschi et al., 2018). Additionally, several clinical conditions may jeopardize the well-being of older people, such as mild cognitive impairment, frailty, or metabolic syndrome (Fried et al., 2001; Petersen, 2004; Portet et al., 2006; Huang, 2009; Xue, 2011; Fedarko, 2012). The main priority of successful management of aging is enabling older people to be healthy, active, and autonomous for as long as possible (World Health Organization, 2002). Accordingly, functional decline is one of the key issues to be managed (World Health Organization, 2015). Among other practices, the use of assistive health technology (AHT; i.e., technologies devoted to maintain or improve functionality, autonomy and well-being) or medical devices (MD; i.e., technologies used for prevention, diagnosis and treatment) may also produce a beneficial effect in older people (Garçon et al., 2016); however, a critical aspect is to ensure accessibility and use of these technologies in the older population (World Health Organization, 2015; Beard et al., 2016).

Virtual reality (VR) is one of the emerging AHT and MD in the field of aging, frailty, and disability (Lange et al., 2010; Bohil et al., 2011). VR is defined as a system based on an interactive computer-simulated 3D environment (Gorini and Riva, 2008), which incorporates mainly auditory and visual feedback, and sometimes also haptic. VR can be divided in non-immersive, semi-immersive, and fully immersive systems (Mujber et al., 2004). The non-immersive system is a desktop-based VR with low interaction (e.g., keyboard, joypad) and immersion (e.g., PC, tablet). The semi-immersive system consists of a large monitor/projector with moderate immersion and interaction (e.g., Kinect, data gloves). The immersive system is characterized by the use of tools such as a head-mounted display (HMD) or the cave automatic virtual environment (CAVE) that enables a high degree of interaction (e.g., trackers) and immersion in the virtual environment (VE). Additionally, VR can be conceptualized as a continuum between reality and virtuality, where some aspects of VE are mixed with the real environment (augmented reality) or vice-versa (augmented virtuality) (Milgram et al., 1995). The sensorimotor channels connected to the VR define the degree of immersion; the psychological consequence of immersion on perception is the sense of presence that felt through being in the VE or, alternatively, the "*perceptual illusion of non-mediation*" with the VE (Riva, 2008; Bohil et al., 2011). Moreover, mobile

applications (e.g., tablet) with tracking systems of the user and/or visors (e.g., Google Cardboard) can be considered mobile VR that allow for different degrees of immersion and interaction with the VE (Pallavicini et al., 2015; Fang et al., 2017).

VR has several requirements for motor and cognitive neurorehabilitation interventions: repetitive practice, feedback about performance, multimodal stimulation, and controlled, secure, and ecologically valid environments (Bohil et al., 2011). It is possible to control and manipulate tailored exercises within meaningful and motivating environments using virtual environments, i.e., transformation of flow (Riva et al., 2006). For these reasons, VR has been utilized for rehabilitation in different fields and, particularly, after stroke. Accordingly, guidelines have recently included the use of VR for both motor and cognitive rehabilitation in patients who suffered a stroke (ISO, 2016b; Winstein et al., 2016). However, access to this kind of technology may be limited by the lack of accessibility in the older population, as compared to other AHT and MD (World Health Organization, 2015). For instance, VR in the context of stroke rehabilitation is facing challenges concerning end-users' interaction, such as feasibility of VR training, lack of functional relevance, patient frustration to feedback, and lack of integration of environmental factors that link to motor performance (Teo et al., 2016).

On the macroscopic level, access to AHT and MD is limited by socio-demographic and economic terms, while on the microscopic level, access is the use itself of a device. Indeed, according to the MOLD-US framework (Wildenbos et al., 2018), the use of technology among older people is hampered by different barriers: (1) cognitive (e.g., reduced working memory, spatial cognition, attention, language, and reasoning) and motivational (e.g., self-efficacy, self-confidence, benefits identification, computer literacy, integration in daily life) that affect the use with errors, efficiency, learnability, memorability and satisfaction; (2) physical (e.g., motor speed, flexibility, hand-eye coordination, strength) and perception (e.g., vision, auditory, haptic) that influence errors and efficiency. According to Nielsen (Nielsen, 2012), usability is defined by learnability (is it easy to accomplish the task?), efficiency (once learned, is the user fast in performing the task?), memorability (is the user able to reestablish proficiency with the design after a period of stop?), errors (how many errors does the user make?) and satisfaction (how pleasant is the design?). Along with usability (i.e., easiness and pleasure), the technology should provide the attributes needed by the user (i.e., utility). Usability can be assessed by a means of a wide range of methods, such as the system usability scale (SUS), heuristic evaluation, cognitive and pluralistic walkthrough, formal usability, pluralistic, consistency, and standard inspections (Brooke, 1986; Nielsen, 1994).

Nevertheless, usability tends to focus more on the task rather than on the experience (Vermeeren et al., 2010). Indeed, researchers investigating user experience (UX) point out a role of factors that go beyond the technology and its usability/usefulness. UX facets embrace emotion and affective reactions toward the technology and experiential, hedonic, holistic, and aesthetic factors. The interaction with a technology is "*a subjective, situated, complex, and dynamic encounter*" (Hassenzahl and Tractinsky, 2006). If it is true that satisfaction

plays a critical role in usability, UX takes into account emotions, motivation, and expectation of human-computer interaction (Vermeeren et al., 2010). For instance, the user experience questionnaire (UEQ) aims at evaluating six factors: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty (Laugwitz et al., 2008), or the usability metric for user experience (UMUX) taps UX facets of usability (Finstad, 2010). Additionally, 96 UX methods (<http://www.allaboutux.org/all-methods>) have been identified in the UX research field (Vermeeren et al., 2010). UX methods range from qualitative to quantitative techniques, target technology, period of assessment (e.g., developmental, conceptual), time, information source (e.g., experts, specific users, individual, group), and location (e.g., lab, online, field). Methods range from semantic differential, checklists, heuristics, think-aloud, psychophysiological measures, self-report, questionnaires, *in situ* observation, and video analysis (Vermeeren et al., 2010). A critical aspect of UX is the prototype development (Novak, 2008), which follows the concept (idea) and pre-production (demo) phases and precedes production & localization (development), Alpha, Beta, and Gold/post-production phases.

A wide range of theories have been proposed to understand and explain user acceptance and use of technology (for a literature review see Taherdoost, 2018). The most inclusive model is the unified theory of acceptance and use of technology model (UTAUTM) (Venkatesh et al., 2003), which includes the technology acceptance model (TAM), theory of reasoned action, theory of planned behavior (TPB), combined TAM and TPB, model of PC utilization, the diffusion of innovation model, motivational model, and social cognitive theory. In this model, the significant factors are: effort, expectancy, performance expectancy, social influence, and facilitating conditions. Interestingly, starting from the TPB (Fishbein and Ajzen, 1975), TAM (Davis et al., 1989), and UTAUTM (Chen and Shou, 2014), developed the senior technology acceptance model (STAM). Controlling age, gender, educational level, and economic status, their model included gerontechnology self-efficacy and anxiety, facilitating health conditions, cognitive abilities, social relationships, attitude to life and satisfaction, and physical functioning as factors that influenced perceived usefulness, usage behavior, and perceived ease of use, which in turn affects general attitude toward the use. A similar model (senior citizens' acceptance of information systems; SCAIS) was developed by Phang et al. (2006). This model takes into account preference for human contact, self-actualization, resource saving, anxiety, computer support, physiological decline which influences perceived usefulness, ease of use, internet safe perception and in turn, intention. Another theoretical framework used to approach technology use and acceptance is the user-centered design (UCD). UCD enables technology systems to be made more usable and interactive to end-users, but it can also be applied to assess needs, wants, and limitations of general products (Sebe, 2010; ISO, 2016a; Brox et al., 2017). UCD can be investigated using a variety of qualitative and quantitative methods such as field studies, user requirements analyses, iterative design, usability evaluation, task analyses, focus groups, user interviews, participatory design, and

prototypes (Vredenburg et al., 2002). UX can be explored with the playability model (i.e., immersion, socialization, emotion, satisfaction, effectiveness) that is crucial when building games for clinical purposes (Sánchez et al., 2012; Valladares-Rodriguez et al., 2019); emotive design for VR should be followed for designing human-computer interaction systems (see Vredenburg et al., 2002).

Lastly, human-computer interfaces are also conceptualized in terms of architecture and layers needed to provide a service (Tsai et al., 2012; Nikitina et al., 2018). For instance, the user remote console (URC) is a framework used for telemedicine systems to define abstract user interface layers, hubs, and devices. If a researcher wishes to consider a VR AHT or MD for healthcare purposes, in addition to the usability and UX aspects, they may want to assess the sense of presence in the VE. According to the Inner Presence theory (Lee, 2004; Riva and Waterworth, 2014), presence is not necessarily related to media characteristics (e.g., graphic realism) but rather to an everyday life flow that controls actions through a constant intentions-perceptions comparison. In this sense, a VR user may experience the system as usable, as they are able to enact actions thanks to an easy-to-learn interface that tracks user's movements, an understandable game/training structure, and engaging storytelling (Triberti and Riva, 2016). These elements are particularly relevant for videogames and serious games used also for therapeutic purposes (Sáenz-de-Urturi et al., 2015). This conceptualization of presence has relevant consequences when taking clinical practice and change into consideration. VR clinical applications should exploit the transformation of flow (transformative and optimal experience allowed by the sense of presence) to discover and use new and unexpected resources to foster clinical change (Riva et al., 2006, 2016) and consider sensorimotor and cognitive impairments in the old population to customize VR for cognitive (Tuena et al., 2019) or physical (Pedroli et al., 2018) rehabilitation.

This paper aims at systematically reviewing the studies that evaluated feasibility, usability, and UX of assessment and treatment VR systems in healthy aging and age-related clinical conditions. In order to provide an overview of the current research status we analyzed characteristics of participants involved, technological apparatus and use, usability/UX assessments, theoretical framework, and primary outcomes. VR use is classified as the task being accomplished and the training sessions and the aims, which include assessment and rehabilitation. Additionally, we outlined suggestions to assess usability of VR applications for older people in clinical and research contexts.

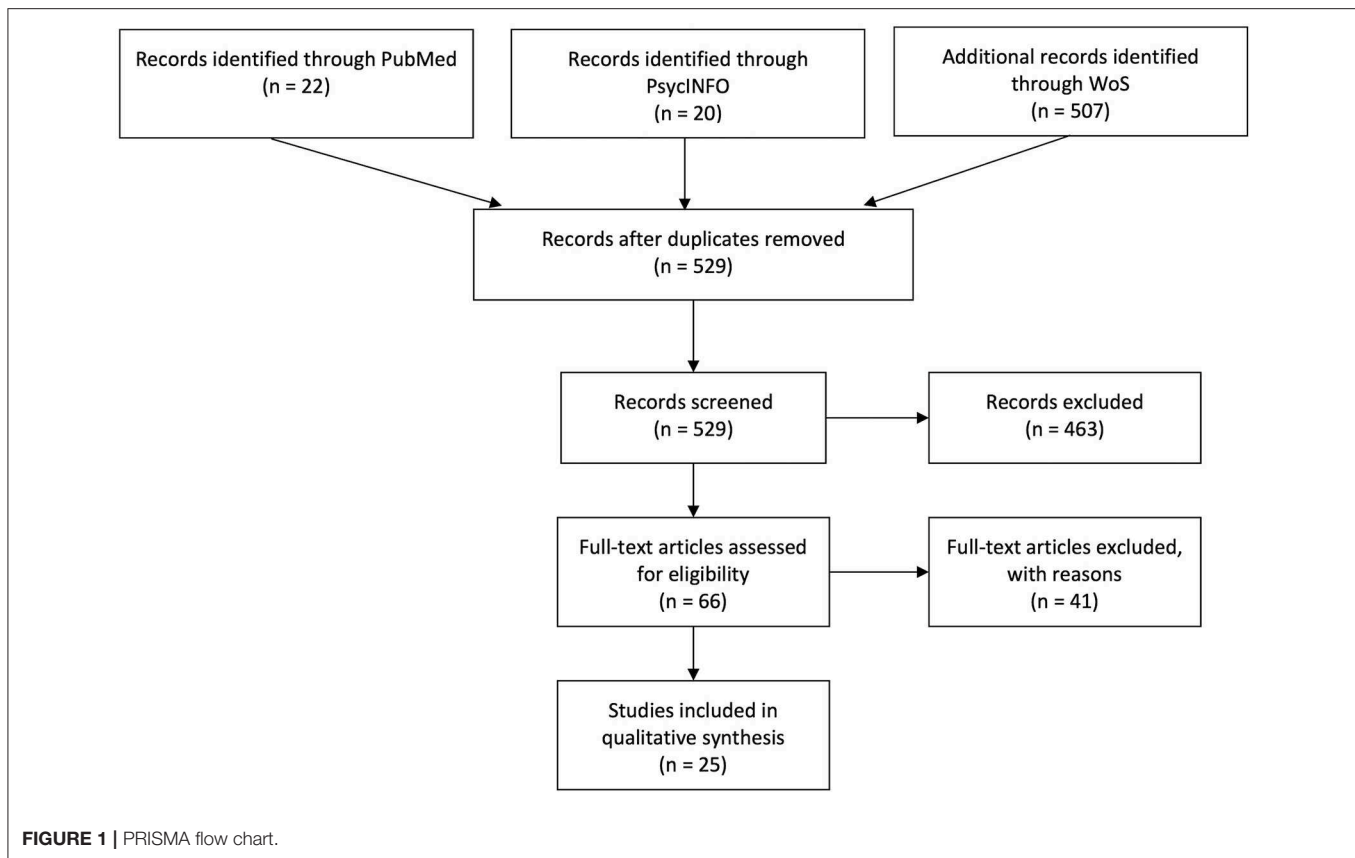
## METHODS

Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) guidelines were followed (Moher et al., 2009).

### Search Strategy

Three high-profile databases (PubMed, PsycINFO, and Web of Science) were used to perform the computer-based research on 3 September 2019. The string used to carry out the search





(Title/Abstract for PubMed, Topic for Web of Science, Abstract for PsycINFO) was as follows: (“aging” OR “frailty” OR “elder\*” OR “multimorbidity”) AND (“usability” OR “user experience” OR “UX” OR “user centered design” OR “human centered design” OR “human computer interaction”) AND (“virtual”). The search resulted in 507 articles for Web of Science, 22 for PubMed, and 20 for PsycINFO (total of 529). We made a first selection by reading titles and abstracts after removing duplicates. A total of 66 manuscripts were chosen for full-text screening. This procedure resulted in 25 experimental studies. See the flow diagram (**Figure 1**) for the paper selection procedure.

## Selection Criteria

Studies concerning the usability, UX, and feasibility of VR (see introduction for definition) systems for assessment/monitoring and rehabilitation/empowerment in healthy and pathological aging were included. In particular, we focused on the age-related clinical conditions in older people. We excluded articles that did not involve usability of VR clinical systems in non-age-related conditions that do not fall into the context of frailty, multimorbidity, or chronicity in aging and with technologies that do not meet VR definition. Additionally, studies for which the full text was not available or for which the abstract lacked basic information for review were removed. Non-English papers, reviews, meeting abstracts, conference proceedings, notes, case reports, letters to the editor, research protocols, patents, editorials, and other editorial materials were also excluded.

## Quality Assessment and Data Abstraction

PRISMA guidelines were strictly followed; search results found by the first author (CT) were shared with the review author (MC) for individual selection of papers in order to reduce the risk of bias, and disagreements were resolved through consensus. The risk of bias for each single study was assessed following the Cochrane guidelines (Higgins et al., 2011) by CT and MC. The research question was formulated according to suggested PICO (Population: older people with age  $\geq 65$ , Intervention: VR for assessment or rehabilitation in age-related conditions and diseases, Comparison: N/A as usability at this time adopt quasi-experimental or pilot study designs (see also risk of bias **Supplementary Figure 1**), Outcome: measures of usability and acceptance) research question guidelines (Abigail et al., 2014). The Comparison is mainly applied to randomized clinical trials and within our search only one study (Schwenk et al., 2014) satisfied this criterion. Consequently, data extracted from each included study were as follows: reference, year, sample (s), aims, technology, VR training, technology design framework, usability/UX/feasibility assessment tools, primary outcomes, and type (assessment/rehabilitative) of VR system.

## RESULTS

Our search identified several usability, user experience (UX), and feasibility studies in healthy aging and age-related clinical

**TABLE 1 |** Summary of the studies included.

References	Sample(s)	Aims	VR technology	VR training	Design framework	Usability assessment tools	Primary outcomes	Clinical field
Brox et al. (2017)	10 OA (age range= 66–90, MMSE > 25) with strength/balance impairments and recent illness/surgery	Recording UX and usability of exergame for physical training in OA	Semi-immersive VR with Kinect	Every second week for 3 years to play exergames and participate in the UCD protocol	Senior UCD	UCD-based questionnaire, semi-structured and structured interviews, observation, group discussions	Results show that VR features (e.g., realism, interaction), usability assessment, and physical impairments are critical factors to be taken into account in the older people	R
Valladares-Rodriguez et al. (2019)	64 older people (16 MCI, mean age = 76.87, SD = 9.33; 20 AD, mean age = 79.15, SD = 4.91; 28 HC, mean age = 75.57, SD = 7.14)	Evaluate UX and PX of game-based battery Panoramix	Non-immersive: Samsung Galaxy Note Pro (SM-P900)	Patients played each game twice during two different sessions (45 min)	TAM, playability model and EMOLVE guidelines	Videogame, technology and TAM questionnaires, PSSUQ and PSSUQ-playability-based to administrators	The Panoramix battery is usable and playable by patients, regardless of their socio-cultural level and their technological dexterity	A
Tsai et al. (2012)	52 OA (age range = 64–91)	Exploring the usability of Sharetouch system to encourage social integration for senior users	Semi-immersive VR with infrared LED	One 10 min session	TAM and architecture design	TAM questionnaire	Sharetouch can enrich the users' social network experience through its hardware and software architecture	R
Nikitina et al. (2018)	60 OA (age range = 59–83) with non-to-mild frailty	Exploring the usability of home-based online group training for home physical training (high vs. low social cohesion and interaction vs. individual group)	Non-immersive VR: PC or tablet app (Gymcentral)	8 weeks at least two sessions (30–40 min) per week	SCAIS	SUS, acceptance questionnaire, VR data (e.g., ratio of copresence sessions, time), MOS, PACES	Group exercise app has a high usability and future use. Copresence was found to be related to social cohesion factor	R
Sáenz-de-Urturi et al. (2015)	14 OA (mean age = 81.28, SD = 8.94 MMSE = 20–26) with mixed age-related disabilities	Assessing usability of Kinect-based training for physical exercises	Semi-immersive VR with Kinect	Three 9 min sessions	Playability model and architecture design	Heuristic evaluation, videotaping, written observation, think aloud, CEGEQ, modified SUS, physical exercise questionnaire	Results from CEGEQ and SUS suggest a high game playability and usability. End-users and experts are critical during the design phase	R
Pedroli et al. (2018)	5 OA (mean age = 70, SD = 11.70; MMSE > 20)	Evaluating usability, characteristics and experience of the Positive Bike for cognitive and physical therapy in frailty	Immersive VR; CAVE with Cosmed Eurobike 320, Vicon motion tracking system and controller	15 min ride in virtual park with a dual interference task (cognitive vs. physical)	ToF	SUS, flow state scale, semi-structured interview	The Positive Bike was evaluated as usable and provided a positive flow experience	R
Cook and Winkler (2016)	11 OA (mean age = 71.2) who completed the training and 8 OA (mean age = 71.2) non-completers	Exploring usability and engagement of VE for health care	Non-immersive SL environments	Four educational sessions on SL	TAM	TAM-based questionnaire	VE are evaluated as adequate and applicable for health care uses after proper training	R
Castilla et al. (2013)	8 OA (age range = 60–72) with no cognitive deterioration and proper vision and audition level	Development and assessment of Butler, a VR telemedicine system for older people	Non-immersive VR	Conceptual design	Not reported	Group enquiry method, cognitive walkthrough method, and heuristic evaluation method	Older people mental model require accurate user interface design in order to facilitate usability	R

(Continued)

TABLE 1 | Continued

References	Sample(s)	Aims	VR technology	VR training	Design framework	Usability assessment tools	Primary outcomes	Clinical field
Corno et al. (2014)	10 OA (age $\geq 60$ . MMSE range: 27–30)	Evaluate the usability of V-MT for executive functions assessment in older people	Immersive VR: HMD with wand	One session with eight tasks of the V-MT	Not reported	Familiarity with technology questionnaire, SSQ, think aloud, SUS, semi-structured usability interview	Usability was found to be crucial for detecting issues of immersive VR (instructions, movements, and realism)	A
Morán et al. (2015)	32 OA ( $M = 64.96$ ; $SD = 6.31$ ) with no apparent cognitive and functional problems were divided according to their experience of technology	The aim of the study is to discuss usability aspects of Gesture Therapy for stroke rehabilitation according to technology experience	Non-immersive VR with hand sensor	Three games (15 min) in one session	TAM2	TAM-based questionnaire, indirect observation (verbal and non-verbal language)	The study shows that expert and non-expert older people differ in terms of anxiety and enjoyment. Two strategies approach were found for the users ( <i>score and compete vs. explore and learn</i> ). Based on these factors, authors provided feedback guidelines for VR trainings	R
Vanbellingen et al. (2017)	13 OA (mean age = 68.2, $SD = 17.5$ )	Evaluating the usability, compliance and efficacy of VBT using the LMC to train fine manual dexterity rehabilitation of stroke patients	Non-immersive VR with LMC	Nine training sessions of 30 min, spread out over 3 weeks	Not reported	SUS, VR data (e.g., time), PRPS, interview.	VBT using LMC is a usable rehabilitation tool to train dexterity in stroke patients	R
Trombetta et al. (2017)	10 OA (age range = 61–75)	The aim of the study is to offer a tool (i.e., Motion Rehab AVE 3D) to improve upper limb motor and balance rehabilitation for stroke patients	Immersive VR with HMD and Kinect and semi-immersive with Smart TV 3D	Motion Rehab AVE 3D contemplates six physical activities	Not reported	Device preference questionnaire and physical training interview	Regarding this pilot study, all participants classified the experience as interesting and excellent for older people. For stroke patients authors suggest semi-immersive apparatus	R
Im et al. (2015)	18 OA (mean age = 64.7, $SD = 7.27$ , mean MMSE = 29.06, $SD = 1$ )	The aim of the study is to assess a novel 3D ARS balance program	Semi-immersive with Kinect	Ten sessions (30 min, three games) over the course of 4 weeks	Not reported	PRPS, side effects interview (e.g., dizziness, headache, falling and joint pain)	Participants were engaged in the training across the sessions without any adverse effects. 3D ARS is a safe, well-tolerated, motivating and efficacious method	R
Wüest et al. (2014)	16 OA (age > 64, MMSE $\geq 22$ )	Assessing the usability of a stroke rehabilitation program (REWIRE project) for motor training	Non-immersive VR with force platform	36, 30-min sessions over 12 weeks (five exergames)	Abridged TAM	TAM questionnaire, think aloud, number of drop-outs and completed sessions	The findings revealed high level of acceptance, positive attitude, future use toward the program	R
Rebsamen et al. (2019)	12 OA (mean age = 72.3, $SD = 4.44$ , MoCA range = 26–30)	Investigating the feasibility and efficacy of a physical exergame on cardiovascular fitness	Semi-immersive VR: Senso system	4 weeks training with three sessions per week (eight exergames; 30 min circa)	TAM	Think aloud, SUS, TAM questionnaire, enjoyment scale, computer use, VR data	Senso has excellent usability, is fun and well-accepted	R

(Continued)

TABLE 1 | Continued

References	Sample(s)	Aims	VR technology	VR training	Design framework	Usability assessment tools	Primary outcomes	Clinical field
Plechátá et al. (2019)	36 OA (mean age = 69.47, <i>SD</i> = 7.39) vs. 25 YA (mean age = 25.4, <i>SD</i> = 5.13)	Assessing age-related differences on immersive vs. non-immersive version of the vSST for episodic memory evaluation	Non-immersive vs. immersive VR: desktop PC and HTC Vive	One session (4–10 min)	Not reported	Ad-hoc usability questionnaire	OA memory was worst in the immersive compared to desktop-based VR. YA prefer HMD and generally reported more usability of VR systems. OA did not show a specific preference	A
Money et al. (2019)	15 participants (age range = 50–70)	Exploring and evaluating usability of Falls Sensei 3D for fall prevention	Non-immersive VR	One exergame session (~17 min)	UTAUTM and architecture design	Think aloud, post-experience interview, SUS	Fall Sensei was rated as engaging and feasible serious game for fall prevention	R
Kiselev et al. (2015)	4 participants with fall risk (1 = control group; 3 = intervention group, age > 55)	The aim of the study is to investigate the usability and user acceptance of VR home-based training (i.e., Interactive Trainer) for fall prevention	Semi-immersive VR with Kinect and sensors	6 weeks training (balance exercises)	UCD	Semi-structured interviews, focus group and VR data	Participants stated that the Interactive Trainer was easy to use and exercises challenging but some technical and interaction problems were reported	R
Shubert et al. (2015)	21 OA (mean age = 69.2, <i>SD</i> = 5.8) with mixed chronic diseases (no neurodegenerative)	Exploring usability of ST as a possible platform to provide a fall prevention program	Non-immersive VR; VERA software, Kinect and laptop	90 min session of system navigation and physical exercise	Not reported	Debrief survey, think aloud, SUS, interview	OA well-accepted this system and show the potential of ST to provide OEP	R
Schwenk et al. (2014)	33 OA with risk fall. Intervention (mean age = 84.3, <i>SD</i> = 7.3) Control (mean age = 84.9, <i>SD</i> = 6.6; MMSE > 23).	Evaluating the effectiveness and UX of a balance-training program.	Semi-immersive VR with sensors	Training session of 45 min twice a week for 4 weeks	Not reported	GEQ	Training was rated as fun, well-designed and adequate	R
van Beek et al. (2019)	10 PD (mean age = 65.4, <i>SD</i> = 7.01, Hoehn and Yahr range = 2–4, MoCA range = 22–29)	Evaluating the usability of a dexterity exergame in PD	Non-immersive VR with LMC	Eight 30 min sessions (5 games) for 4 weeks	Not reported	VR data (i.e., time/planned time × 100), PRPS, interview, SUS	Patients showed high adherence, motivation, enjoyment and good usability	R
Desteghe et al. (2017)	15 AF patients (mean age = 69.2, <i>SD</i> = 3.7).	The aim of this pilot study was to assess the feasibility and usability of the Health Buddies app in AF patient	Non-immersive (PC, tablet or mobile)	Training lasted every day for 3 months	Not reported	Focus group, UEQ, MMAS-8, MEMS, Helping Hand, VR data	The app was positively rated by its users; nevertheless adherence to medication was low and need user-friendly interface	A
Epelde et al. (2014)	13 medical professionals and 19 orthopedic patients (mean age = 69.31, <i>SD</i> = 7.38)	Assessing the acceptance of a universal remote rehabilitation leaded by avatar	Semi-immersive VR with inertial sensors	One session	URC	Ad-hoc usability questionnaire, focus group	Medical professionals were positive regarding the virtual therapists and patients showed good acceptance of the system	R
Fordell et al. (2011)	31 stroke patients (mean age = 74.1, <i>SD</i> = 11) with no severe comorbidity and with ( <i>N</i> = 9) or without ( <i>N</i> = 22) neglect	Assesses effectiveness and usability of VR-DiSTRO compared to gold-standard neglect assessment	Immersive VR 3D glasses and interaction pen	One session (VR 15 min “paper and pencil” 50 min)	Not reported	Ad-hoc usability	Patients felt focused, amazed and comfortable with the VR assessment. Any severe side effects were reported.	A

(Continued)



TABLE 1 | Continued

References	Sample(s)	Aims	VR technology	VR training	Design framework	Usability assessment tools	Primary outcomes	Clinical field
Kizony et al. (2006)	12 OA (mean age = 70.6, SD = 4.4) 4 OA patients (3 stroke and 1 spinal stenosis)	Assess usability of TheraGame for neurorehabilitation	Non-immersive VR	One session (30 min) and 2 weeks and half for one stroke patient	Not reported	SUS, SFQ, Borg scale	Both groups showed good level of usability and enjoyment during the session. Also the caregiver who followed the patient during the 2 weeks confirmed the usability.	R

A, assessment; AD, Alzheimer's disease; AF, atrial fibrillation; ARS, augmented reality system; CAVE, cave automatic virtual environment; CEGEQ, core elements of the gaming experience questionnaire; EMOVLE, emotive virtual learning environment; GEQ, game experience questionnaire; HMD, head-mounted display; HC, healthy controls; LMC, leap motion controller; MEMS, medication event monitoring system; MCI, mild cognitive impairment; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; MMAS-8, Morisky medication adherence scale; OA, older adults; OEP, Orago exercise program; PD, Parkinson's disease; PACES, physical activity enjoyment scale; PAPS, Pittsburgh rehabilitation participation scale; PX, player experience; PSSUQ, post-study system usability questionnaire; R, rehabilitation; SL, second life; SCALS, senior citizens' acceptance of information systems; SFQ, short feedback questionnaire; SSQ, simulator sickness questionnaire; ST, stand tall; SUS, system usability scale; TAM, technology acceptance model; ToF, transformation of flow; UTAUTM, unified theory of acceptance and use of technology model; UX, user experience; UEQ, user experience questionnaire; UCD, user-centered design; URC, user remote console; VB1, videogame-based training; VE, virtual environment; V-MIT, virtual multitasking test; VR, virtual reality; vSST, virtual supermarket shopping task; YA, young adults.

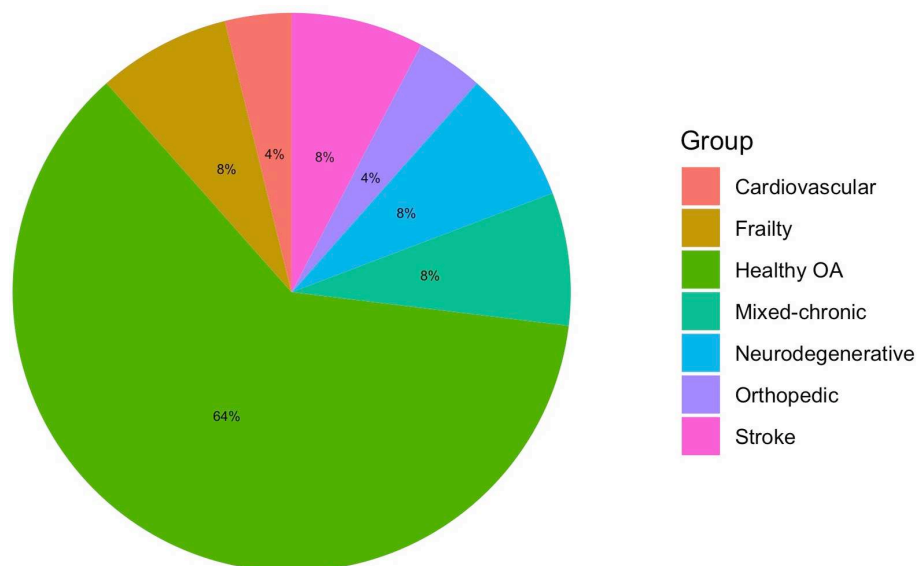
conditions. A critical aspect of virtual reality (VR) and new technologies is their interaction with humans and in particular, those whose physical, psychological, or social barriers hamper the use of technological devices. The aim of this systematic review was to analyze the current research in the field of usability of clinical VR systems in older people and to provide an overview on this topic. Findings are shown in **Table 1** according to reference, year, sample(s), aims of the study, VR technology, VR training, theoretical framework, usability assessment, primary outcomes, and clinical aims. **Figures 2–8** summarize the results as well.

## Which Are the Samples Involved in VR Usability Studies?

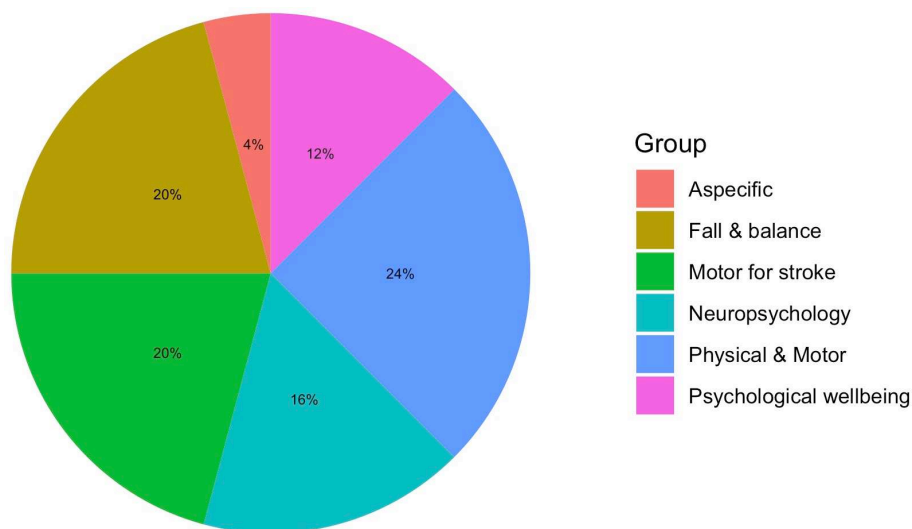
The majority of the studies (Kizony et al., 2006; Tsai et al., 2012; Castilla et al., 2013; Corno et al., 2014; Wüest et al., 2014; Im et al., 2015; Morán et al., 2015; Cook and Winkler, 2016; Trombetta et al., 2017; Vanbellingen et al., 2017; Plechatá et al., 2019; Rebsamen et al., 2019) recruited healthy older adults (OA) to assess the usability of clinical VR systems. Two studies collected data from the fifth decade to old age (Kiselev et al., 2015; Money et al., 2019). Nevertheless, in these studies, systems were created for clinical conditions such as stroke (Wüest et al., 2014; Morán et al., 2015; Trombetta et al., 2017; Vanbellingen et al., 2017) or movement disorders (e.g., balance, physical frailty; Pedroli et al., 2018; Money et al., 2019). Indeed, only two studies recruited stroke patients for stroke VR systems (Kizony et al., 2006; Fordell et al., 2011). Sáenz-de-Urturi et al. (2015) and Pedroli et al. (2018) recruited OA and, among these individuals, some had mild or moderate cognitive impairment (O'Bryant et al., 2017). Patients with mixed age-related conditions (e.g., Parkinson's disease, macular degeneration, muscular dystrophy, arthritis, diabetes, hypertension) were recruited in Sáenz-de-Urturi et al. (2015) and Shubert et al. (2015). Frail patients were collected in Nikitina et al. (2018), mixed frail and physical-motor patients in Brox et al. (2017), participants at risk of falling in Schwenk et al. (2014) and Kiselev et al. (2015), mild cognitive impairment (MCI) and Alzheimer's disease (AD) individuals in Valladares-Rodriguez et al. (2019), Parkinson's disease (PD) individuals in van Beek et al. (2019), OA with atrial fibrillation (AF) in Desteghe et al. (2017), and with orthopedics impairments in Epelde et al. (2014). Experts and medical professionals were included in some pilot studies for their opinion on the design or on the VR system (Castilla et al., 2013; Epelde et al., 2014; Morán et al., 2015; Sáenz-de-Urturi et al., 2015; Brox et al., 2017; Desteghe et al., 2017; Valladares-Rodriguez et al., 2019).

## Which are the Aims and the Clinical Fields of the Studies?

All of the studies—except one that was principally devoted to the clinical efficacy of the training (Schwenk et al., 2014)—were mainly designed for the usability, UX and feasibility of VR systems in aging (Tsai et al., 2012; Castilla et al., 2013; Corno et al., 2014; Epelde et al., 2014; Wüest et al., 2014; Im et al., 2015; Kiselev et al., 2015; Morán et al., 2015; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Cook and Winkler, 2016; Brox et al., 2017; Desteghe et al., 2017; Trombetta et al., 2017;



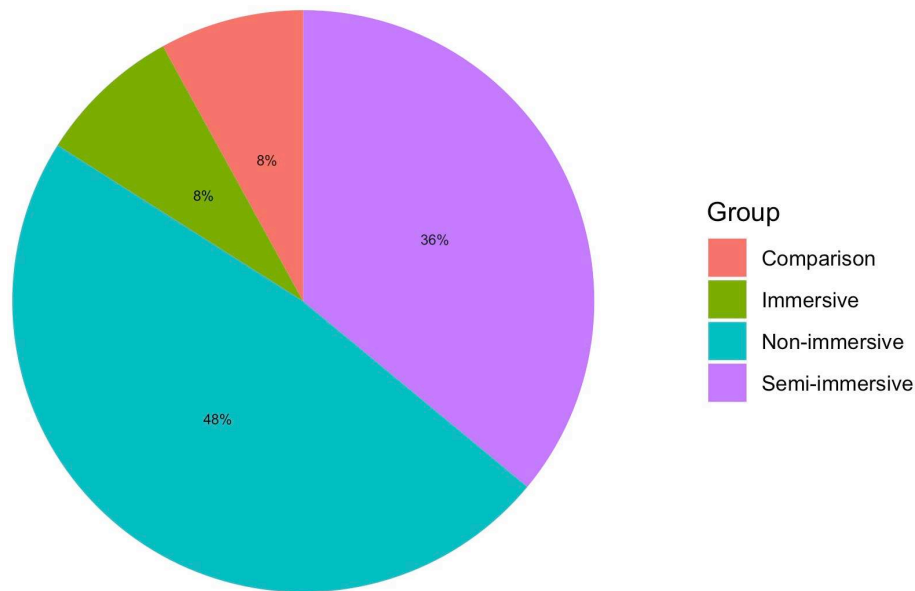
**FIGURE 2 |** Clinical conditions; OA, older adults.



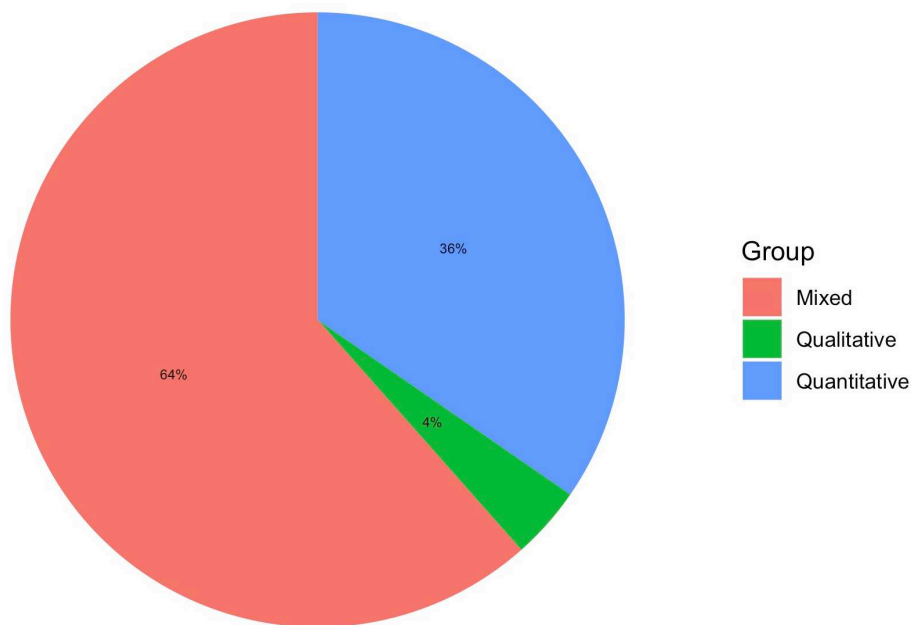
**FIGURE 3 |** Clinical applications.

Vanbellingen et al., 2017; Nikitina et al., 2018; Pedroli et al., 2018; Money et al., 2019; Plechatá et al., 2019; Rebsamen et al., 2019; Valladares-Rodriguez et al., 2019; van Beek et al., 2019). Most of the studies concerned the assessment of therapeutic (i.e., rehabilitative or psychological empowerment) VR systems Tsai et al., 2012; Castilla et al., 2013; Corno et al., 2014; Epelde et al., 2014; Wüest et al., 2014; Im et al., 2015; Kiselev et al., 2015; Morán et al., 2015; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Cook and Winkler, 2016; Brox et al., 2017; Desteghe et al., 2017; Trombetta et al., 2017; Vanbellingen et al., 2017; Nikitina et al., 2018; Pedroli et al., 2018; Money et al., 2019;

Plechatá et al., 2019; Rebsamen et al., 2019; Valladares-Rodriguez et al., 2019; van Beek et al., 2019, whereas only a few were on assessment or monitoring tools (Fordell et al., 2011; Corno et al., 2014; Desteghe et al., 2017; Plechatá et al., 2019; Valladares-Rodriguez et al., 2019). The intervention/assessment of the studies included were physical-motor (e.g., limb physiotherapy, physical activity, hand motricity; Kizony et al., 2006; Epelde et al., 2014; Schwenk et al., 2014; Wüest et al., 2014; Im et al., 2015; Kiselev et al., 2015; Morán et al., 2015; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Brox et al., 2017; Trombetta et al., 2017; Vanbellingen et al., 2017; Nikitina et al., 2018;



**FIGURE 4 |** Degree of virtual immersion; VR, virtual reality.

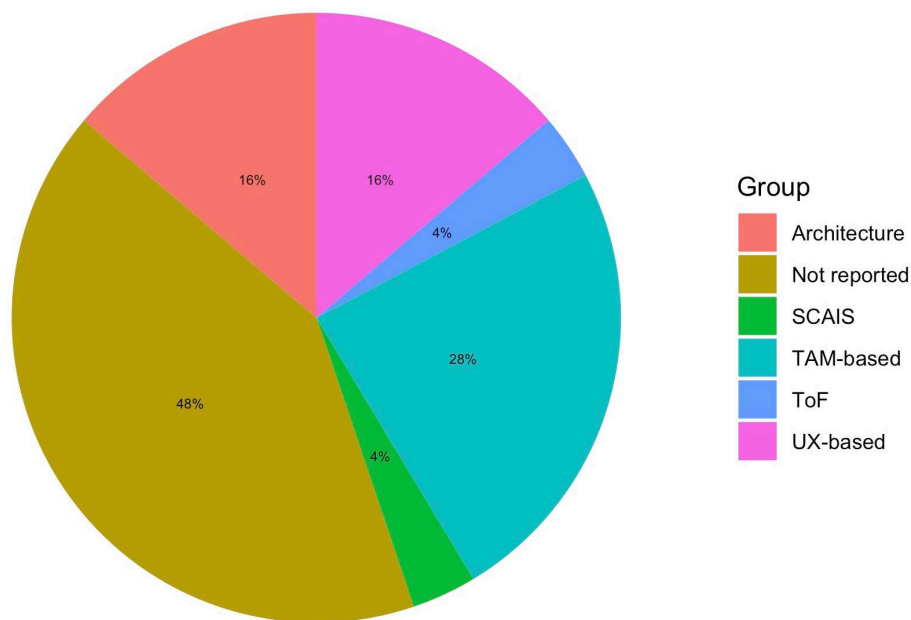


**FIGURE 5 |** Methodological approach.

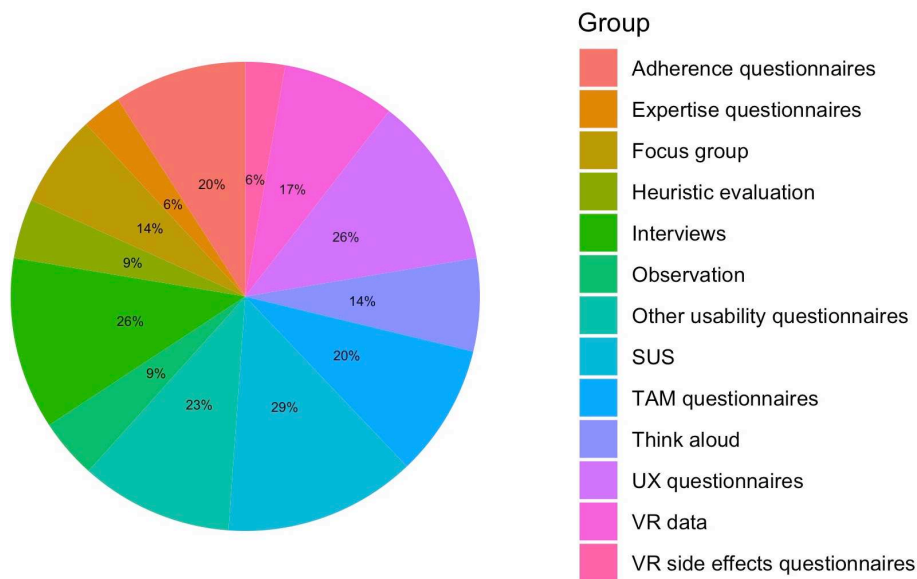
Pedroli et al., 2018; Money et al., 2019; van Beek et al., 2019), neuro/psychological (Fordell et al., 2011; Tsai et al., 2012; Castilla et al., 2013; Corno et al., 2014; Desteghe et al., 2017; Plechatá et al., 2019; Valladares-Rodriguez et al., 2019), cardiovascular fitness (Rebsamen et al., 2019), or non-specific healthcare applications (Cook and Winkler, 2016).

### Which Are the VR Technologies Used and the Training?

Non-immersive VR (i.e., desktop-based VR, tablet, and mobile app) were used in most of the studies (Kizony et al., 2006; Castilla et al., 2013; Wüest et al., 2014; Morán et al., 2015; Shubert et al., 2015; Cook and Winkler, 2016; Desteghe et al.,



**FIGURE 6 |** Technology usability and acceptance models; SCAIS, senior citizens acceptance of information systems; TAM, technology acceptance model; ToF, transformation of flow; UX, user experience.

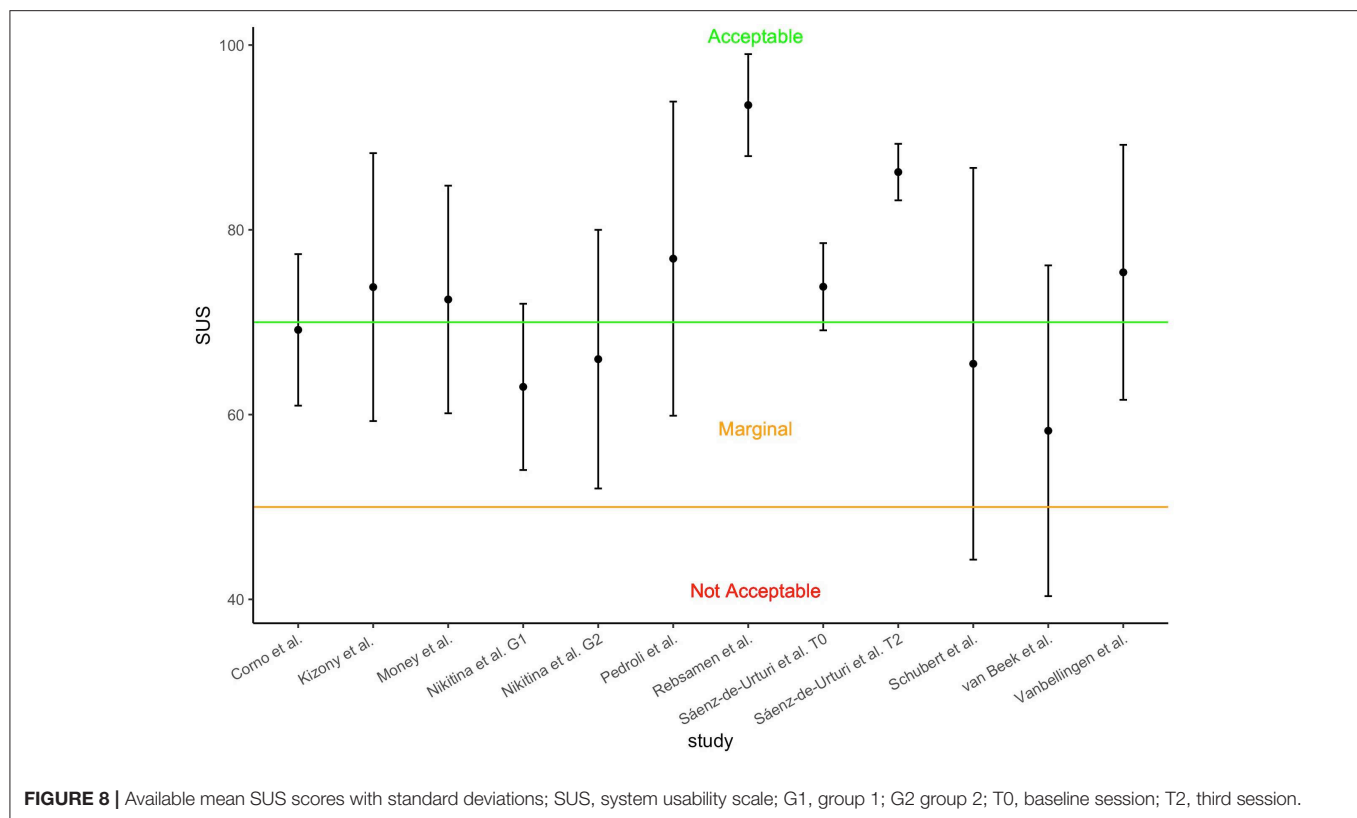


**FIGURE 7 |** Assessment tools by the total of unique instruments; SUS, system usability scale; TAM, technology acceptance model; UX, user experience; VR, virtual reality.

2017; Vanbellingen et al., 2017; Nikitina et al., 2018; Money et al., 2019; Valladares-Rodriguez et al., 2019; van Beek et al., 2019). Application exclusively for tablets were used in Valladares-Rodriguez et al. (2019) and multi-device (i.e., PC, tablet or mobile) apps in Castilla et al. (2013), Desteghe et al. (2017),

and Nikitina et al. (2018). Semi-immersive VR (i.e., large TV or projector screens with sensors for interaction) systems were tested in some studies (Tsai et al., 2012; Epelde et al., 2014; Schwenk et al., 2014; Im et al., 2015; Kiselev et al., 2015; Sáenz-de-Urturi et al., 2015; Brox et al., 2017; Rebsamen et al., 2019),





whereas full immersive VR (i.e., visors or CAVE with interaction devices) was tested only in Fordell et al. (2011), Corno et al. (2014), and Pedroli et al. (2018). Interestingly, Trombetta et al. (2017) compared a semi-immersive vs. an immersive version of the training to evaluate their usability, whereas Plechatá et al. (2019) tested a VR memory test with non-immersive vs. immersive VR.

Sessions lasted from nine to 90 min (mean time = 30 min ca.), ranging from one to 36 sessions spread over the course of 1 single day to 4 years; indeed, usability along with effectiveness of training was tested for three (Vanbellinghen et al., 2017), four (Im et al., 2015; Rebsamen et al., 2019; van Beek et al., 2019), six (Kiselev et al., 2015), eight (Nikitina et al., 2018), and 12 weeks (Wüest et al., 2014; Desteghe et al., 2017), and three (Brox et al., 2017) and 4 years (Schwenk et al., 2014). Exergames (i.e., serious games used for balance and fall risk training) were used in most of the motor training (Kizony et al., 2006; Schwenk et al., 2014; Wüest et al., 2014; Kiselev et al., 2015; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Brox et al., 2017; Trombetta et al., 2017; Money et al., 2019; Rebsamen et al., 2019; van Beek et al., 2019), physiotherapy exercise in others (Epelde et al., 2014; Im et al., 2015), cognitive-physical dual-task in Pedroli et al. (2018), neuropsychological testing in three studies (Fordell et al., 2011; Corno et al., 2014; Plechatá et al., 2019; Valladares-Rodriguez et al., 2019), gesture therapy in Morán et al. (2015) and Vanbellinghen et al. (2017), and psychosocial support/educational in four studies (Tsai et al., 2012; Castilla et al., 2013; Cook and Winkler, 2016; Desteghe et al., 2017).

## Which Are the Theories and Tools Used to Assess VR?

A key component regarding the use and acceptance of technology is understanding elements that facilitate or reduce its use in terms of human factors, not only in terms of technical ones (Wildenbos et al., 2018). An architecture structure model, the user remote console (URC), was used as a theoretical background to design the physical training in Epelde et al. (2014), whereas the majority of the studies used psychological models to develop the VR systems. Technology acceptance model (TAM) and modified versions were used in most of the studies (Tsai et al., 2012; Wüest et al., 2014; Morán et al., 2015; Cook and Winkler, 2016; Money et al., 2019; Rebsamen et al., 2019; Valladares-Rodriguez et al., 2019), transformation of flow (ToF) in Pedroli et al. (2018), UX playability in two studies (Sáenz-de-Urturi et al., 2015; Valladares-Rodriguez et al., 2019), and user-centered design model (UCD) in Kiselev et al. (2015) and Brox et al. (2017). Importantly, some studies adopted technological theoretical frameworks adapted for older people, such as the senior UCD or the senior citizens' acceptance of information systems (SCAIS) (Brox et al., 2017; Nikitina et al., 2018). However, several studies did not report a theoretical model to design their systems (Kizony et al., 2006; Castilla et al., 2013; Corno et al., 2014; Schwenk et al., 2014; Im et al., 2015; Shubert et al., 2015; Desteghe et al., 2017; Trombetta et al., 2017; Vanbellinghen et al., 2017; Plechatá et al., 2019; van Beek et al., 2019). Concerning the assessment of usability of VR systems, a wide range of quantitative and qualitative methods have been

used (see **Table 1** for specific information and **Figures 5–7** for models and methods overviews). Concerning quantitative data, system usability scale (SUS) (Corno et al., 2014; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Vanbellingen et al., 2017; Nikitina et al., 2018; Pedroli et al., 2018; Money et al., 2019; Rebsamen et al., 2019; van Beek et al., 2019), TAM-based questionnaires (Tsai et al., 2012; Wüest et al., 2014; Morán et al., 2015; Cook and Winkler, 2016; Rebsamen et al., 2019; Valladares-Rodriguez et al., 2019), UX questionnaires (Schwenk et al., 2014; Sáenz-de-Urturi et al., 2015; Brox et al., 2017; Desteghe et al., 2017; Rebsamen et al., 2019), UCD-based questionnaire (Brox et al., 2017), flow of experience scale (Pedroli et al., 2018), other usability questionnaires (Fordell et al., 2011; Epelde et al., 2014; Trombetta et al., 2017; Plechatá et al., 2019; Valladares-Rodriguez et al., 2019), adherence or motivation to training questionnaires (Im et al., 2015; Desteghe et al., 2017; Vanbellingen et al., 2017; Nikitina et al., 2018; Rebsamen et al., 2019; van Beek et al., 2019) or with VR data (Wüest et al., 2014; Kiselev et al., 2015; Desteghe et al., 2017; Vanbellingen et al., 2017; Nikitina et al., 2018; Rebsamen et al., 2019; van Beek et al., 2019), cybersickness assessment (Corno et al., 2014; Im et al., 2015; Plechatá et al., 2019), technology expertise (Corno et al., 2014; Rebsamen et al., 2019; Valladares-Rodriguez et al., 2019), and video analysis (Morán et al., 2015) were used. Regarding qualitative data, think aloud technique (Corno et al., 2014; Wüest et al., 2014; Shubert et al., 2015; Money et al., 2019; Rebsamen et al., 2019), heuristic evaluation or cognitive walkthrough (Castilla et al., 2013; Sáenz-de-Urturi et al., 2015), focus group (Castilla et al., 2013; Epelde et al., 2014; Kiselev et al., 2015; Brox et al., 2017; Desteghe et al., 2017), and semi-structured or structured usability post-experience interviews (Corno et al., 2014; Kiselev et al., 2015; Shubert et al., 2015; Brox et al., 2017; Vanbellingen et al., 2017; Pedroli et al., 2018; Money et al., 2019; van Beek et al., 2019) were used. The sense of presence was assessed only in three studies (Kizony et al., 2006; Nikitina et al., 2018; Pedroli et al., 2018).

Concerning the tools used, a variety of quantitative and qualitative methods are reported. However, it is important to remember that each of these instruments assess different aspects of usability and acceptance; some are more concerned about the task to perform (e.g., SUS) while others tap the emotional/motivational elements of the interaction (e.g., UX questionnaires) or the factors that hamper/facilitate the use of a technology (e.g., TAM-based tools). Qualitative tools are able to grasp different perspectives (individual or group) of the experience or the design by asking experts in the sector or the end-user itself. A multidimensional approach emerged in our search and should be preferred when selecting assessment tools.

## Are VR Clinical Systems for the Older People Usable?

In this section we outlined the findings of the included studies, reporting their strengths and weaknesses. **Figure 8** shows mean and standard deviation for the available SUS scores, which display moderate to acceptable usability despite some cases of wide variation.

Cook and Winkler (2016) showed that OA find virtual environments (VE) from Second Life (SL) as feasible and applicable for healthcare purposes, especially for improving social interactions. Despite a high number of drop-outs, participants liked the realism and virtual experience (e.g., sports, changing avatar, teleporting, shopping) but bugs frustrated them and they found it hard to control the avatar and to learn SL. According to users, SL might be improved by clear training (i.e., individualized, small group), step-by-step teaching, by enlarging the screen, and facilitating the interaction. The exergame Falls Sensei was rated as engaging and usable for educating OA about risk fall (Money et al., 2019). Falls Sensei was rated as having a good usability (score SUS > 70, Bangor et al., 2009), especially by older users. Unified theory of acceptance and use of technology (UTAUT) thematic analysis on interviews (i.e., performance expectancy, effort, social influence) showed that users rated the training as a useful, positive experience, relevant for specific populations. Similarly, the Positive Bike (Pedroli et al., 2018) was rated as having good usability (mean SUS = 76.88, SD = 17). Problems were found concerning the size of items on the screen and low realism or interaction users felt in the VE, but still had a positive experience and found the system useful. Stand Tall (ST) (Shubert et al., 2015) was rated by participants as having a nearly good usability (mean SUS = 65.5, SD = 21.2) and agreed in using ST to improve balance autonomously and accepted the Kinect sensor and the avatar. Senso system (Rebsamen et al., 2019) had high adherence, usability (mean SUS = 93.5, SD = 5.52), enjoyment, usefulness, and acceptability, also confirmed by think aloud technique. Similarly, van Beek et al. (2019) found optimal adherence and motivation toward their VR training. Despite some interaction issues with LMC and difficulty of the exercises, the system had marginal usability (mean SUS = 58.25, SD = 17.9) and was also rated positively at the interviews. Lineage was evaluated with high satisfaction by its users (Sáenz-de-Urturi et al., 2015). Gaming experience was positive, exercise adequate, and participants stated that they would use the game again. SUS improved across the three sessions (first mean SUS = 73.84, SD = 4.72; third mean SUS = 86.25, SD = 3.06). Acceptable usability was reported by OA and stroke patients for the TheraGame (mean SUS = 73.8, SD = 14.5) that also found the VR training adequate and enjoyable (Kizony et al., 2006). Good usability (first session mean SUS = 75.4, SD = 13.8) was found by Vanbellingen et al. (2017) in their upper limb video game with a leap motion controller; however, usability did not change across the nine sessions. The training had a compliance of 87.4% and the adherence was rated as very good and remained stable across time. Users expressed that a 30 min session is the best time to not overload arm fatigue. Optimal (100%) adherence and good acceptance (e.g., ease, usefulness, intention to use) were found by Wüest et al. (2014). Nikitina et al. (2018) found that usability of the virtual gym App did not differ between groups with social interaction (mean SUS = 63, SD = 9) or interaction with coach only (mean SUS = 66, SD = 14). Moreover, the participants positively accepted the app, with high co-presence for the interaction group (interactions occurred especially with private messages), but adherence was similar for individual vs. group exercises with social support

predicating adherence when social connections are low. Despite Corno et al. (2014) finding that virtual-multitasking test (V-MT) induces cybersickness symptoms, it was rated as usable (mean SUS = 69.17,  $SD = 8.2$ ), the head-mounted display (HMD) was comfortable, interaction with the wand was difficult, instructions hard to remember, and realism sufficient. Similar results on HMD were found by Plechatá et al. (2019). HMD lead to the worst memory performances compared to non-immersive VR in OA, with users preferring neither desktop-based VR nor immersive VR, whereas young users liked immersive versions of the virtual supermarket shopping task (vSST). However, authors suggest non-immersive scenarios for OA. Fordell et al. assessed VR-DiSTRO, an immersive VR version of “paper and pencil” neglect neuropsychological battery, and showed that stroke patients tolerated and were engaged during the assessment, which was much faster than the classic evaluation (Fordell et al., 2011).

In order to design Game Up exergames and a senior-UCD model (Brox et al., 2017), it is crucial to involve older people and experts to create safe, fun, and usable games. Three-point Likert scale short questionnaires are suggested for end evaluations, whereas in the requirement, design, and implementation phases, interviews, observations, and group discussions are preferred for senior UCD. Similarly, in order to develop the Butler app (Castilla et al., 2013) it is important to gather information from end-users and experts from the first stage of the development and to create prototypes of the app. Graphics and navigations systems must be adequate and understandable for older people in order to reduce mental load. In the same way, the Health Buddies app (Desteghe et al., 2017) was initially designed with the end-users (AF patients and grandchildren). Participants, especially patients, were motivated to use the app but its usage decreased across 90 days. Despite adherence improving only in one patient, the UX with the app was easy to use and educational, and 60% of patients would use the app again. Experts and end-users of a joint rehabilitation virtual therapy were also involved in the evaluation phase in Epelde et al. (2014). Medical professionals and patients positively accepted the virtual therapist and training but patients stated that the avatar was too serious and lacked empathy. A team of experts developed an augmented reality exergame (Im et al., 2015), which did not have any side effects (e.g., cybersickness) and led to high adherence to the training.

The Interactive Trainer (Kiselev et al., 2015), despite some technical problems being reported, was evaluated according to interviews as easy to use, challenging, and motivating. Schwenk et al. (2014) assessed the gaming UX of a exergame with sensors, which was found to be effective, fun, easy to learn thanks to feedback, adequate, and well-designed. Interestingly, Valladares-Rodriguez et al. (2019) aimed at assessing UX and player eXperience (PX) of Panoramix neuropsychological touchscreen battery in OA, mild cognitive impairment (MCI), and Alzheimer's disease individuals. They found that Panoramix perception and acceptance were positive after the pilot study in the groups but was judged as more playable by OA, MCI, and AD in this order; nevertheless, PX improved after the second interaction in all groups. Additionally, administrators also evaluated the battery as playable, usable, useful, and with a good interface. Morán et al. (2015) used a TAM-based questionnaire

and video analysis to assess usability. Users rated the VR gesture therapy (GT) as useful, easy, and with high UX and found that even technological expertise did not affect task performance. By analyzing verbal and non-verbal reactions, raters judged the system as more usable and fun for non-expert participants. Conversely, anxiety was low for expert users. Authors defined two approach strategies according to expertise, *explore-and-learn* and *score-and-complete*, respectively, for inexperienced and experienced participants that guided behaviors (e.g., anxiety, interaction strategies with the games) and reactions through the experience.

A comparison of semi vs. full immersive versions of Motion Rehab AVE 3D was done by Trombetta et al. (2017). Training was feasible for users and participants evaluated as important for usability feedback, third-person perspective, comfort (semi-immersion version), and immersion (full immersion). Authors suggested that, for post-stroke rehabilitation, semi-immersive systems are more comfortable than full-immersive VR. Tsai et al. (2012) showed that Sharetouch is a well-designed, easy, and usable system, independent of gender or age, and facilitates social interactions in OA. Importantly, significant effects of the rehabilitative training on different motor/physical measures were found in all the studies that tested efficacy and usability (Schwenk et al., 2014; Wüest et al., 2014; Im et al., 2015; Vanbellingen et al., 2017; Rebsamen et al., 2019; van Beek et al., 2019). However, risk of bias (see **Supplementary Figure 1**) is high for most of the categories (randomization, allocation, blinding, missing data, and reporting bias), as the majority of the research is quasi or non-experimental. Of note, the risk of incomplete data outcome was low.

In general, despite some technical weaknesses (e.g., realism, bugs), interaction constraints and physical/psychological barriers to technology use, the included VR studies showed that with adequate usability design methods, it is possible to develop effective and usable systems for clinical purposes in aging.

## DISCUSSION

In the present paper we reviewed the current research on usability, user experience (UX), and feasibility of virtual reality (VR) clinical systems in older people.

Our work can be summarized in the following points: (1) most of the usability pilots involved healthy or heterogeneous diseased older people; (2) usability mainly concerned VR physiotherapy training; (3) most of the studies involved non-immersive scenarios; (4) quantitative (e.g., SUS) and qualitative (e.g., interviews) methods are the most used and suggested approach in usability piloting and technology acceptance model (TAM) is the main theoretical framework; (5) despite some interaction issues, VR systems are rated as having good usability by end-users.

Usability is a critical and complex task when specific end-users with particular needs are involved. Conditions that hamper the interaction with the device (Wildenbos et al., 2018), and also cultural and technology background, should be taken into account (Corno et al., 2014; Nikitina et al., 2018). For instance,

Tuena et al. found that executive functions are overloaded by input device use in older people and this leads to worse memory performances (Tuena et al., 2019). Design guidelines should be used to avoid basic sensorimotor and interaction issues (see Phiriyapokanon, 2011; Loureiro and Rodrigues, 2014).

If, on the one hand, the studies included collected data from the target population (e.g., Parkinson's disease patients tested usability for Parkinson's disease rehabilitation), several others assessed usability with healthy older people or mixed-pathologies patients (e.g., Wüest et al., 2014; Sáenz-de-Urturi et al., 2015; Shubert et al., 2015; Trombetta et al., 2017; Vanbellinghen et al., 2017); in this sense, diagnostic criteria were not clear or end-users characteristic do not match potential technology barriers of end-users. Future research should use strict inclusion/exclusion criteria according to diagnostic criteria of the diseases or syndromes. Moreover, in the context of healthcare, the end-users are also the medical professionals that use the technology with the patients. Usability should be assessed via questionnaires or interviews in the design and test phases (e.g., Castilla et al., 2013; Valladares-Rodriguez et al., 2019). Finally, despite some studies reporting the number of participants as a limitation (Corno et al., 2014; Desteghe et al., 2017; Vanbellinghen et al., 2017; van Beek et al., 2019), a number of 5–10 individuals is sensible enough to identify a minimum of 80% circa of usability issues (Wüest et al., 2014; Brox et al., 2017).

The uses of VR systems in our review were mainly focused on motor rehabilitation. In healthcare, VR is mainly applied for the assessment and rehabilitation of sensorimotor, physical, and psychological deficits via non-immersive to immersive technologies (Lange et al., 2010; Bohil et al., 2011; García-Betances et al., 2015; Muratore et al., 2019; Tuena et al., 2019). We also encourage the use of pilot studies in other domains where VR is used for clinical purposes. For instance, it is important to evaluate usability of assessment tools (e.g., Pedrolí et al., 2015; Desteghe et al., 2017). Mean usability session testing lasted 30 min; nevertheless, depending on the aims of the studies (e.g., memorability), longitudinal usability studies can be done as usability might improve after some sessions (Valladares-Rodriguez et al., 2019). Lastly, future research should focus more on immersive technology as technical development will lead to new forms of immersive VR and costs will be reduced. It is important to also assess these systems because they might lead to reduced cybersickness compared to desktop-based VR (Lange et al., 2010; Bohil et al., 2011; Plechatá et al., 2019).

Several studies (see **Table 1**) did not report a model on which usability and acceptance of a technology can be assumed. TAM-based and UX-based are useful for investigating and understanding psychological factors, whereas architecture design and user remote control (URC) are more useful for technical development. Indeed, usability, and in particular UX, are devoted not only to the ease of use and the technical bugs but also to the psychological domains (e.g., emotions, motivations; Vermeeren et al., 2010). However, as researchers in the context of aging face specific needs and barriers, adapted models with relevant variables should be used as the senior user-centered design (UCD) by Brox et al. (2017) or the senior citizens' acceptance of information systems (SCAIS) by Phang et al. (2006). Surprisingly,

none of the authors used the senior technology acceptance model (STAM) by Chen and Shou (2014), which could be more suitable than TAM models not adapted to older people. Interestingly clinical researchers interested in technology usability, sense of presence, and clinical change may want to use the transformation of flow (ToF) theory, as presence and flow experiences might facilitate clinical change by means of VR (Riva et al., 2006).

Usability assessment (see **Table 1** and **Figure 7**) tools should include a mix of quantitative methods (e.g., SUS, TAM-based questionnaires, UX-based questionnaires) and qualitative techniques (e.g., experience interviews, think aloud, heuristic evaluation). The systematic review on telemedicine systems by Klaassen et al. (2016) recommend SUS, TAM2, and PSSUQ and state that questionnaires along with interviews, which are both low-cost and flexible methods, can be used from early to final phases of usability. Indeed, questionnaires give useful quantitative data that, however, still need qualitative information to tap individual sources of variation. Therefore, a mixed approach composed of quantitative and qualitative tools is the preferred way to carry out complete, interpretable, and useful usability studies in older people. Additionally, we encourage a critical adoption of assessment tools according to the aims of the study, thus considering the aspects (e.g., individual, group, task, emotions/motivation, acceptance, adherence) to be engaged during the VR interaction.

Additionally, innovative quantitative techniques could be useful to track unexpected information about psychophysiological (e.g., eye-tracking, heart-rate, galvanic skin response, non-verbal communication) responses of the users to assess their affective and cognitive reactions to the VR system (Morán et al., 2015; Sáenz-de-Urturi et al., 2015). VR can also be used for evaluating usability and adherence (good >80%) by using time spent, number of log-ins, or interaction modality, giving additional quantitative data (Cipresso, 2015; Rebsamen et al., 2019; van Beek et al., 2019). Importantly, when testing immersive VR, cybersickness should always be assessed because it may negatively influence clinical practice and its reduction is a key objective of pilot studies (Kober et al., 2013; Corno et al., 2014; Tuena et al., 2017; Plechatá et al., 2019) and virtual embodiment with questionnaires if avatars are used (Kilteni et al., 2012; Gonzalez-Franco and Tabitha, 2018). Finally, in the early design phases, information from end-users (e.g., patients, medical professionals) could be gathered from group interviews or focus groups, where ideas from experts' opinions and needs can be used to guide VR development (Castilla et al., 2013; Brox et al., 2017). For instance, Brox et al. (2017) developed a senior-UCD with a mixed use of quantitative and qualitative methods to design a semi-immersive exergame for older people, through iteration from the early phases to the prototype. Researchers should be aware that step-by-step UCD (e.g., prototype development) and pretesting are critical for clinical VR settings (Novak, 2008; Im et al., 2015). However, we know that time is a limitation to some research projects and, in some occasions, there is no time for longitudinal and proper VR design. When this is not possible, we strongly encourage the use of qualitative and quantitative evaluation of the VR experience. In the same manner, it would be better to assess usability and



acceptance separately from efficacy of a VR system, as quality of patients' healthcare services is intertwined with usability, acceptance, and adherence (Middleton et al., 2013).

Despite some technical and interaction issues (e.g., bugs, interaction difficulties, realism, sensors application), the included studies showed that usability of a wide range of VR clinical systems is good, well-accepted, adequate, effective, and useful. Skepticism of older people and digital divide are walls that could be successfully broken after the use of VR devices (Desteghe et al., 2017) and comfort of immersive VR can be improved by replacing visors with CAVE, although non-HMD systems are considered better for older people (Corno et al., 2014; Pedrolí et al., 2018; Plechatá et al., 2019). Nevertheless, a recent study shows that OA positively accept and tolerate HMD VR (Huygelier et al., 2019). Indeed, Fordell et al. showed that stroke patients enjoyed the immersive VR assessment (Fordell et al., 2011). However, a future objective in the field is to make sensors application and use easier for this population, as home-based training, where no professional is present to provide assistance, is rising in popularity in VR clinical practice (Schwenk et al., 2014). Moreover, online assistance could be useful to help patients with set-up and exercises (Im et al., 2015; Nikitina et al., 2018). Morán et al. (2015) provided some guidelines concerning the feedback the VR training should give to older users:

- “Provide timely feedback on successful actions in a simple and salient manner”;
- “Provide feedback on erroneous actions in a simple and salient manner”;
- “Provide simple and salient instructions on how to recover or solve an error”;
- “Provide feedback that fosters or inhibits specific behaviors from the user in a salient and concise manner.”

Additionally, Teo et al. (2016) provide specific suggestions in their review for VR training in individuals with stroke-related impairments, such as flexible activity according to patients' objectives, possibility to adapt online the task by the therapist according to patient's needs, multiplayer services, and automated recording of patient tracking. Moreover, Teo et al. (2016) show that VR can be enriched with neurophysiological tools (e.g., EEG, fNIRS) that the researcher or the clinician can use to adapt the task according to individual effort or needs.

Finally, it is worth mentioning some solutions provided by the Cochrane guidelines (Higgins et al., 2011) that avoid risk of bias in usability experiments. Despite blinding procedures in cognitive/motor rehabilitation trials (VR vs. treatment as usual) being a hard task to fulfill, still randomization, attrition bias, and reporting bias can be improved, respectively, with random number generators, shuffling cards, or throwing dice, with adequate missing data manipulation (e.g., balanced observation, imputation) and via adequate hypotheses and primary/secondary outcomes specification in the introduction and then in the discussion and adequate analyses in the result section.

The present review outlined current VR usability piloting issues and strengths in healthy aging and age-related clinical

**TABLE 2 | VR-USOP.**

1	Identify Barriers and Facilitators	<ul style="list-style-type: none"> <li>• Use UTAUTM, STAM, MOLD-US or SCAIS models (Venkatesh et al., 2003; Phang et al., 2006; Chen and Shou, 2014; Wildenbos et al., 2018)</li> <li>• Clinical diagnosis and expert clinicians</li> </ul>
2	Develop adequate VR and task	<ul style="list-style-type: none"> <li>• Architecture design</li> <li>• Use older people technology design guidelines (Phiriyapokanon, 2011; Loureiro and Rodrigues, 2014)</li> <li>• Apply senior-UCD (Brox et al., 2017)</li> <li>• If training, use feedback guidelines (Morán et al., 2015)</li> <li>• Iterative prototyping</li> </ul>
3	Define usability assessment	<ul style="list-style-type: none"> <li>• Quantitative methods (e.g., SUS, TAM-based, UX-based questionnaires, PSSUQ)</li> <li>• Qualitative methods (e.g., post-experience interviews, think-aloud)</li> <li>• Additional methods (VR data, observation, psychophysiological measurements)</li> <li>• Assess usability and feasibility from medical professionals</li> <li>• If training, evaluate adherence</li> <li>• If immersive VR, evaluate cybersickness</li> <li>• If immersive VR with avatar, assess virtual embodiment</li> <li>• Consider PX for serious games and sense of presence for VR</li> <li>• Usability session should last 30 min approx.; more sessions can be included if there are experimental reasons</li> </ul>
4	Test clinical use	<ul style="list-style-type: none"> <li>• If usability results are unsatisfying, adjust VR system before clinical testing</li> <li>• If usability results are acceptable, start clinical efficacy testing</li> </ul>

PSSUQ, post-study system usability questionnaire; PX, player eXperience; SCAIS, senior citizens' acceptance of information systems; STAM, senior technology acceptance model; SUS, system usability scale; UCD, user-centered design; UTAUTM, unified theory of acceptance and use of technology model; UX, user eXperience; VR, virtual reality.

conditions. In the following paragraph, we will provide suggestions for researchers who wish to run usability testing in the context of clinical application of VR systems for older patients.

## VR-USABILITY SUGGESTIONS FOR THE OLDER PEOPLE (VR-USOP)

In the present paragraph we presented some suggestions we derived from findings of the systematic review. VR-USOP will be mainly focused on human-interaction factors rather than on technical aspects of developing VR clinical systems. **Table 2** summarizes some suggestions in four steps to follow if researchers and clinicians wish to design and test their VR clinical apparatus to older end-users.

The assessment of potential barriers and facilitators of the end-users, which can also include the medical professionals and technology acceptance models, is the first step. In our opinion, this is crucial as it allows the identification

and the development of adequate characteristics of VR interaction and task (step 2). The latter aspects will be provided also by adopting architecture design, senior-UCD, and guidelines and prototyping, thus allowing the definition of usability assessment. In addition, we encourage ameliorating the methodology (risk of bias, see **Supplementary Figure 1**; i.e., randomization, allocation, blinding, manipulation of missing data, and reporting bias) to overcome the limitations of the available studies analyzed in the present review. VR usability and acceptance assessment should be defined and developed in accordance to the aims of the study (step 3). We suggest a mixed-approach with quantitative and qualitative methods (mainly focused on psychological experience of usability) and additional aspects to consider (see **Table 2**). Lastly, we suggest ensuring usability before clinical testing (step 4).

## CONCLUSIONS

This systematic review aimed at describing an overview of state of the art VR clinical systems for older people in relation to usability and providing researchers with suggestions based on the results of the review. Despite some limitations concerning the criteria used to recruit the samples, the low number of immersive technologies so far tested, and the high risk of bias of the studies, VR systems show good usability and acceptance among older people. A wide variety of quantitative and qualitative methods can be used to evaluate usability. We suggest adopting mixed-methodology with appropriate tools in order to grasp different aspects of the usability, acceptability, and user experience and to plan sessions according to objectives of usability. Piloting is a critical aspect of clinical studies with VR technology and we encourage future research to test usability of their applications following VR-USOP.

## REFERENCES

- Abigail, M. M., Campbell, S., Chew-Graham, C., McNally, R., and Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv. Res.* 14:579. doi: 10.1186/s12913-014-0579-0
- Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Stud.* 4, 114–123. doi: 10.5555/2835587.2835589
- Beard, J. R., Officer, A. M., and Cassels, A. K. (2016). The world report on ageing and health. *J. Gerontol. B. Psychol. Sci. Soc. Sci.* 56, S163–S166. doi: 10.1093/geront/gnw037
- Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 752–762. doi: 10.1038/nrn3122
- Brooke, J. (1986). “SUS: A ‘quick and dirty’ usability scale,” in *Usability Evaluation in Industry*. eds P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland (London: Taylor & Francis), 189–194.
- Brox, E., Konstantinidis, S. T., and Evertsen, G. (2017). User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. *JMIR Serious Game* 5:e2. doi: 10.2196/games.6254
- Castilla, D., Garcia-Palacios, A., Breto, J., Mari, R., Etchemendy, E., Farfallini, L., et al. (2013). Process of design and usability evaluation of a telepsychology web and virtual reality system for the elderly: Butler. *Int. J. Hum. Comput. Stud.* 71, 350–362. doi: 10.1016/j.ijhcs.2012.10.017

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## AUTHOR CONTRIBUTIONS

CT wrote the first draft of the manuscript. MS-B supervised and wrote the following drafts of the manuscript. MS-B, FG, CT, EP, AGal, and PT defined the methodology and objectives of the manuscript. MC assessed risk of bias and made the second search strategy. AGag gave framework for the VR-USOP. KG provided clinical expertise and support. GR and FL revised the manuscript. All authors contributed to the revision and final approval of the manuscript.

## FUNDING

This work was funded by the Italian Ministry of Health IRCCS Network on Aging Research roadmap on aging and age-related diseases RRC-2018-2365820.

## ACKNOWLEDGMENTS

CT wishes to thank all the authors for providing useful ideas and supervision during the conceptualization, writing, and revision of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2020.00093/full#supplementary-material>

**Supplementary Figure 1** | Risk of bias assessment.

- Chen, K., and Shou, A. H. C. (2014). Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). *Ergonomics* 57, 635–652. doi: 10.1080/00140139.2014.895855
- Cipresso, P. (2015). Modeling behavior dynamics using computational psychometrics within virtual worlds. *Front. Psychol.* 6:1725. doi: 10.3389/fpsyg.2015.01725
- Cook, N., and Winkler, S. L. (2016). Acceptance, usability and health applications of virtual worlds by older adults: a feasibility study. *JMIR Res. Protoc.* 5:e81. doi: 10.2196/resprot.5423
- Corno, G., Bouchard, S., and Forget, H. (2014). Usability assessment of the virtual multitasking test (v-mt) for elderly people. *Annual Rev. Cyberther. Telemed.* 199, 168–172.
- Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Manag. Sci.* 35, 982–1003. doi: 10.1287/mnsc.35.8.982
- Desteghe, L., Kluts, K., Vijgen, J., Koopman, P., Dilling-Boer, D., Schurmans, J., et al. (2017). The health buddies app as a novel tool to improve adherence and knowledge in atrial Fibrillation patients: a pilot study. *JMIR Mhealth Uhealth* 5:e98. doi: 10.2196/mhealth.7420
- Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., et al. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 9, 1118–1127. doi: 10.1016/S1474-4422(10)70223-4
- Edwards, M. H., Dennison, E. M., Sayer, A. A., Fielding, R., and Cooper, C. (2015). Osteoporosis and sarcopenia in older age. *Bone* 80, 126–130. doi: 10.1016/j.bone.2015.04.016

- Epelde, G., Carrasco, E., Rajasekharan, S., Jimenez, M., Vivanco, K., Gomez-Fraga, I., et al. (2014). Universal remote delivery of rehabilitation : validation with seniors ' joint rehabilitation therapy. *Cybern. Syst.* 45, 109–122. doi: 10.1080/01969722.2014.874807
- Fang, W., Zheng, L., Deng, H., and Zhang, H. (2017). Real-time motion tracking for mobile augmented/virtual reality using adaptive visual-inertial fusion. *Sensors* 17:1037. doi: 10.3390/s17051037
- Fedarko, N. S. (2012). The biology of aging and frailty. *Clin. Geriatr. Med.* 27, 27–37. doi: 10.1016/j.cger.2010.08.006
- Finstad, K. (2010). The usability metric for user experience. *Interact. Comput.* 22, 323–327. doi: 10.1016/j.intcom.2010.04.004
- Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.
- Flatt, T. (2012). A new definition of aging? *Front. Genet.* 3:148. doi: 10.3389/fgene.2012.00148
- Fordell, H., Bodin, K., Bucht, G., and Malm, J. (2011). A virtual reality test battery for assessment and screening of spatial neglect. *Acta Neurol. Scand.* 123, 167–174. doi: 10.1111/j.1600-0404.2010.01390.x
- Franceschi, C., Garagnani, P., Morsiani, C., Conte, M., Santoro, A., Grignolio, A., et al. (2018). The continuum of aging and age-related diseases :common mechanisms but different rates. *Front. Med.* 5:61. doi: 10.3389/fmed.2018.00061
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., et al. (2001). Frailty in older adults : evidence for a phenotype. *J. Gerontol.* 56, 146–157. doi: 10.1093/gerona/56.3.M146
- García-Betances, R. I., Arredondo Waldmeyer, M. T., Fico, G., and Cabrera-Umpiérrez, M. F. (2015). A succinct overview of virtual reality technology use in Alzheimer's disease. *Front. Aging Neurosci.* 7:80. doi: 10.3389/fnagi.2015.00235
- Garçon, L., Khasnabis, C., Walker, L., Nakatani, Y., Lapitan, J., Borg, J., et al. (2016). Medical and assistive health technology : meeting the needs of aging populations. *J. Gerontol. A Biol. Soc. Sci. Med.* 56, S293–S302. doi: 10.1093/geront/gnw005
- Gonzalez-Franco, M., and Tabitha, C. P. (2018). Avatar embodiment. Towards a standardized questionnaire. *Front. Rob. AI* 5:74. doi: 10.3389/frobt.2018.00074
- Gorini, A., and Riva, G. (2008). Virtual reality in anxiety disorders: the past and the future. *Expert Rev. Neurother.* 8, 215–233. doi: 10.1586/14737175.8.2.215
- Hassenzahl, M., and Tractinsky, N. (2006). User experience - a research agenda. *Behav. Inform. Technol.* 25, 91–97. doi: 10.1080/01449290500330331
- Higgins, J. P., Altman, D. G., Gotzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928. doi: 10.1136/bmj.d5928
- Huang, P. L. (2009). A comprehensive definition for metabolic syndrome. *Dis. Model. Mech.* 2, 231–237. doi: 10.1242/dmm.001180
- Huygheer, H., Schraepen, B., van Ee, R., Abeele, V. V., and Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality un older adults. *Sci. Rep.* 9:4519. doi: 10.1038/s41598-019-41200-6
- Im, D. J., Jeunghun, K., Joon, K. Y., Sangwoo, C., Kyung, C. Y., Teo, L., et al. (2015). Utility of a three-dimensional interactive augmented reality program for balance and mobility rehabilitation in the elderly :a feasibility study. *Ann. Rehabil. Med.* 39, 462–472. doi: 10.5535/arm.2015.39.3.462
- ISO (2016a). *Ergonomics of Human-System Interaction—Part 210: Human-Centred Design For Interactive Systems*.
- ISO (2016b). *SPREAD: Linee Guida Italiane di Prevenzione e Trattamento*. Retrieved from: [http://www.iso-stroke.it/?page\\_id=200](http://www.iso-stroke.it/?page_id=200) (accessed October 15, 2019).
- Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence Teleo. Virt. Environ.* 21, 373–387. doi: 10.1162/PRES\_a\_00124
- Kiselev, J., Haesner, M., Gövercin, M., and Steinhagen-Thiessen, E. (2015). Implementation of a home-based interactive training. *J. Gerontol. Nurs.* 41, 14–19. doi: 10.3928/00989134-20141201-01
- Kizony, R., Weiss, P. L., Shahar, M., and Rand, D. (2006). Theragame: a home based virtual reality rehabilitation system. *Int. J. Dis. Hum. Develop.* 5, 265–270. doi: 10.1515/IJDHD.2006.5.3.265
- Klaassen, B., van Beijnum, B. J. F., and Hermens, H. (2016). Usability in telemedicine systems – a literature survey. *Int. J. Med. Inform.* 93, 57–69. doi: 10.1016/j.iimedinf.2016.06.004
- Kober, S. E., Wood, G., Hofer, D., Kreuzig, W., Kiefer, M., and Neuper, C. (2013). Virtual reality in neurologic rehabilitation of spatial disorientation. *J. Neuroeng. Rehabil.* 10:17. doi: 10.1186/1743-0003-10-17
- Kuehn, E., Perez-lopez, M. B., Diersch, N., Döhler, J., Wolbers, T., and Riemer, M. (2017). Embodiment in the aging mind. *Neurosci. Biobehav. Rev.* 86, 207–225. doi: 10.1016/j.neubiorev.2017.11.016
- Lange, B. S., Requejo, P., Flynn, S. M., and Rizzo, A. A. (2010). The potential of virtual reality and gaming to assist successful aging with disability. *Phys. Med. Rehabil. Clin.* 21, 339–356. doi: 10.1016/j.pmr.2009.12.007
- Laugwitz, B., Schrepp, M., and Held, T. (2008). “Construction and evaluation of a user experience questionnaire,” in *Symposium of the Austrian HCI and Usability Engineering Group* (Berlin: Springer), 63–76. doi: 10.1007/978-3-540-89350-9\_6
- Lee, K. M. (2004). Why presence occurs: evolutionary psychology, media equation, and presence. *Presence* 13, 494–505. doi: 10.1162/1054746041944830
- Loureiro, B., and Rodrigues, R. (2014). “Design guidelines and design recommendations of multi-touch interfaces for elders,” in *Proceedings of the ACHI* (Barcelona), 41–47.
- Lutz, W., Sanderson, W., Scherbov, S., and Asia, S. (2008). The coming acceleration of global population ageing. *Nature* 451, 716–719. doi: 10.1038/nature06516
- Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., et al. (2011). Aging with multimorbidity : a systematic review of the literature. *Ageing Res. Rev.* 10, 430–439. doi: 10.1016/j.arr.2011.03.003
- Middleton, B., Bloomrosen, M., Dente, M. A., Hashmat, B., Koppel, R., Overhage, J. M., et al. (2013). Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *J. Am. Med. Inform. Assoc.* 20, e2–e8. doi: 10.1136/amiajnl-2012-001458
- Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). “Augmented reality: a class of displays on the reality-virtuality continuum,” in *SPIE 2351 Telemanipulator and Telepresence Technologies* (Boston, MA), 2351. doi: 10.1117/12.197321
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Grp, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement (reprinted from annals of internal medicine). *Phys. Ther.* 89, 873–880. doi: 10.1093/ptj/89.9.873
- Money, A. G., Atwal, A., Boyce, E., Gaber, S., Windeatt, S., and Alexandrou, K. (2019). Falls Sensei :a serious 3D exploration game to enable the detection of extrinsic home fall hazards for older adults. *BMC Med. Inform. Decis. Mak.* 19:85. doi: 10.1186/s12911-019-0808-x
- Morán, A. L., Ramírez-Fernández, C., Meza-Kubo, V., Orihuela-Espina, F., García-Canseco, E., Grimaldo, A. I., et al. (2015). On the effect of previous technological experience on the usability of a virtual rehabilitation tool for the physical activation and cognitive stimulation of elders. *J. Med. Syst.* 39:104. doi: 10.1007/s10916-015-0297-0
- Mujber, T. S., Szecsi, T., and Hashmi, M. S. J. (2004). Virtual reality applications in manufacturing process simulation. *J. Mat. Process. Technol.* 156, 1834–1838. doi: 10.1016/j.jmatprotec.2004.04.401
- Muratore, M., Tuena, C., Pedroli, E., Cipresso, P., and Riva, G. (2019). Virtual reality as a possible tool for the assessment of self-awareness. *Front. Behav. Neurosci.* 13:62. doi: 10.3389/fnbeh.2019.00062
- Nielsen, J. (1994). “Usability inspection methods,” in *Usability Inspection Methods*, eds J. Nielsen and R. L. Mack (New York, NY: John Wiley & Sons), 413–414. doi: 10.1145/259963.260531
- Nielsen, J. (2012). *Usability 101: Introduction to Usability*. Fremont, CA: Nielsen Norman Group.
- Nikitina, S., Didino, D., Baez, M., and Casati, F. (2018). Feasibility of virtual tablet-based group exercise among older adults in siberia : findings from two pilot. *JMIR Mhealth Uhealth* 6:e40. doi: 10.2196/mhealth.7531
- Novak, J. (2008). *Game Development Essentials: An Introduction. 3rd Edn*. Clifton Park, NY: Delmar Cengage Learning.
- O'Bryant, S. E., Humphreys, J. D., Smith, G. E., Ivnik, R. J., Graff-radford, N. R., Petersen, R. C., et al. (2017). Detecting Dementia with the mini-mental state examination in highly educated individuals. *Arch. Neurol.* 65, 963–967. doi: 10.1001/archneur.65.7.963
- Pallavicini, F., Pedroli, E., Serino, S., Dell'Isola, A., Cipresso, P., Cisari, C., et al. (2015). Assessing unilateral spatial neglect using advanced technologies: the potentiality of mobile virtual reality. *Technol. Health Care* 23, 795–807. doi: 10.3233/THC-151039
- Pedroli, E., Greci, L., Serino, S., Cipresso, P., Arlati, S., Mondellini, M., et al. (2018). Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment :positive bike. *Sensors* 18:2343. doi: 10.3390/s18072343



- Pedroli, E., Serino, S., Cipresso, P., Pallavicini, F., Chicchi Giglioli, I., Guastafierro, E., et al. (2015). "Neglect App. Usability of a new application for assessment and rehabilitation of neglect," in *REHAB '15 Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques* (Lisbon), 139–143. doi: 10.1145/2838944.2838978
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x
- Phang, C. W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B. C., and Teo, H. H. (2006). Senior Citizens' acceptance of information systems: a study in the context of e-government services. *IEEE Trans. Eng. Manage.* 53, 555–569. doi: 10.1109/TEM.2006.883710
- Phiriapokanon, T. (2011). *Is a Big Button Interface Enough for Elderly Users: Towards user Interface Guidelines for Elderly Users*. Kittery, ME: LAP LAMBERT Academic Publishing.
- Plechátá, A., Sahula, V., Fayette, D., and Fajnerová, I. (2019). Age-related differences with immersive and non-immersive virtual reality in memory assessment. *Front. Psychol.* 10:1330. doi: 10.3389/fpsyg.2019.01330
- Portet, F., Ousset, P. J., Visser, P. J., Frisoni, G. B., Nobili, F., Scheltens, P., et al. (2006). Report of the MCI working group of the European consortium on Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* 77, 714–719. doi: 10.1136/jnnp.2005.085332
- Rebsamen, S., Knols, R. H., Pfister, P. B., Jung, M. E., and Elliot, C. A. (2019). Exergame-Driven high-intensity interval training in untrained community dwelling older adults : a formative one group quasi- experimental feasibility trial. *Front. Physiol.* 10:1019. doi: 10.3389/fphys.2019.01019
- Riva, G. (2008). From virtual to real body: virtual reality as embodied technology. *J. Cyber Ther. Rehabil.* 1, 7–22.
- Riva, G., Baños, R. M., Botella, C., Mantovani, F., and Gaggioli, A. (2016). Transforming experience : the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Front. Psychiatry* 7:164. doi: 10.3389/fpsyg.2016.00164
- Riva, G., Castelnuovo, G., and Mantovani, F. (2006). Transformation of flow in rehabilitation: the role of advanced communication technologies. *Behav. Res. Methods* 38, 237–244. doi: 10.3758/BF03192775
- Riva, G., and Waterworth, J. A. (2014). "Being present in a virtual world," in *The Oxford Handbook of Virtuality*, ed M. Grimshaw (Oxford: Oxford University Press), 205–221. doi: 10.1093/oxfordhb/9780199826162.013.015
- Sáenz-de-Urturi, Z., Zapirain, B. G., and Zorrilla, A. M. (2015). Elderly user experience to improve a Kinect-based game playability. *Behav. Inform. Technol.* 34, 1040–1051. doi: 10.1080/0144929X.2015.1077889
- Sánchez, J. L., Vela, F. L., Simarro, F. M., and Zea, N. P. (2012). Playability: analysing user experience in video games. *Behav. Inform. Technol.* 31, 1033–1054. doi: 10.1080/0144929X.2012.710648
- Schwenk, M., Grewal, G. S., Honarvar, B., Schwenk, S., Mohler, J., Khalsa, D. S., et al. (2014). Interactive balance training integrating sensor-based visual feedback of movement performance : a pilot study in older adults. *J. Neuroeng. Rehabil.* 11:164. doi: 10.1186/1743-0003-11-164
- Sebe, N. (2010). "Human-centered computing," in *Handbook of Ambient Intelligence and Smart Environments*, (New York, NY: Springer). 349–370. doi: 10.1007/978-0-387-93808-0\_13
- Shubert, T. E., Basnett, J., Chokshi, A., Barrett, M., Hill, C., and Carolina, N. (2015). Are virtual rehabilitation technologies feasible models to scale an evidence-based fall prevention program? A pilot study using the kinect camera. *JMIR Rehabil. Assist. Technol.* 2:e10. doi: 10.2196/rehab.4776
- Steenman, M., and Lande, G. (2017). Cardiac aging and heart disease in humans. *Biophys. Rev.* 9, 131–137. doi: 10.1007/s12551-017-0255-9
- Steptoe, A., Deaton, A., and Stone, A. A. (2015). Psychological wellbeing, health and ageing. *Lancet* 385, 640–648. doi: 10.1016/S0140-6736(13)61489-0
- Taherdoost, H. (2018). "A review of technology acceptance and adoption models and theories," in *A Review of Technology Acceptance and Adoption Models and Theories*, eds L. Moldovan and A. Gligor (Tirgu Mures: Elsevier B.V.), 960–967. doi: 10.1016/j.promfg.2018.03.137
- Teo, W.-P., Muthalib, M., Yamin, S., Hendy, A. M., Bramstedt, K., Kotsopoulos, E., et al. (2016). Does a combination of virtual reality, neuromodulation and neuroimaging provide a comprehensive platform for neurorehabilitation? – a narrative review of the literature. *Front. Hum. Neurosci.* 10:284. doi: 10.3389/fnhum.2016.00284
- Tolosa, E., Wenning, G., and Poewe, W. (2006). The diagnosis of Parkinson's disease. *Lancet Neurol.* 5, 75–86. doi: 10.1016/S1474-4422(05)70285-4
- Triberti, S., and Riva, G. (2016). Being present in action: a theoretical model about the "interlocking" between intentions and environmental affordances. *Front. Psychol.* 6:2052. doi: 10.3389/fpsyg.2015.02052
- Trombetta, M., Paula, P., Henrique, B., Brum, M. R., Colussi, E. L., Carolina, A., et al. (2017). Motion Rehab AVE 3D: a VR-based exergame for post-stroke rehabilitation. *Comput. Methods Programs Biomed.* 151, 15–21. doi: 10.1016/j.cmpb.2017.08.008
- Tsai, T., Chang, H., Chang, Y., and Huang, G. (2012). A system to enrich social network experiences for the elderly. *J. Syst. Software* 85, 1363–1369. doi: 10.1016/j.jss.2012.01.023
- Tuena, C., Serino, S., Dutriaux, L., Riva, G., and Piolino, P. (2019). Virtual enactment effect on memory in young and aged populations : a systematic review. *J. Clin. Med.* 8:620. doi: 10.3390/jcm8050620
- Tuena, C., Serino, S., Gaston-Bellegarde, A., Orriols, E., Makowski, D., Riva, G., et al. (2017). How virtual embodiment affects episodic memory functioning : a proof-of-concept study. *Ann. Rev. Cyberther. Telemed.* 98–103.
- Valladares-Rodríguez, S., Fernández-Iglesias, M. J., Anido-Rifón, L., Facal, D., Rivas-Costa, C., and Pérez-Rodríguez, R. (2019). Touchscreen games to detect cognitive impairment in senior adults. A user- interaction pilot study. *Int. J. Med. Inform.* 127, 52–62. doi: 10.1016/j.ijmedinf.2019.04.012
- van Beek, J. J. W., van Wegen, E. E. H., Bohlhalter, S., and Vanbellinghen, T. (2019). Exergaming-based dexterity training in persons with parkinson disease: a pilot feasibility study. *J. Neurol. Phys. Ther.* 43, 168–174. doi: 10.1097/NPT.0000000000000278
- Vanbellinghen, T., Filius, S. J., Nyffeler, T., and van Wegen, E. E. (2017). Usability of videogame-based dexterity training in the early rehabilitation phase of stroke patients : a pilot study. *Front. Neurol.* 8:654. doi: 10.3389/fneur.2017.00654
- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Q* 425–478. doi: 10.2307/30036540
- Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). "User experience evaluation methods : current state and development needs," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction*, 521–530. doi: 10.1145/1868914.1868973
- Vredenburg, K., Mao, J., Smith, P., and Carey, T. (2002). "A survey of user-centered design practice," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Edinburgh), 3460–3478. doi: 10.1145/503376.503460
- Wildenbos, G. A., Peute, L., and Jaspers, M. (2018). Aging barriers in influencing mobile health usability for older adults : a literature based framework (MOLD-US). *Int. J. Med. Inform.* 114, 66–75. doi: 10.1016/j.ijmedinf.2018.03.012
- Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., et al. (2016). Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 47, e98–e169. doi: 10.1161/STR.0000000000000098
- World Health Organization (2002). *Active Ageing: A Policy Framework*. Work.
- World Health Organization (2015). *World Report on Ageing and Health*.
- Wüest, S., Borghese, N. A., Pirovano, M., Mainetti, R., van de Langenberg, R., and de Bruin, E. D. (2014). Usability and effects of an exergame-based balance training program. *Games Health J.* 3, 106–114. doi: 10.1089/g4h.2013.0093
- Xue, Q. (2011). The frailty syndrome: definition and natural history. *Clin. Geriatr. Med.* 27, 1–15. doi: 10.1016/j.cger.2010.08.009
- Yakaryilmaz, F. D., and Öztürk, Z. A. (2017). Treatment of type 2 diabetes mellitus in the elderly. *World J. Diabetes* 8:278. doi: 10.4239/wjd.v8.i6.278

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

Copyright © 2020 Tuena, Pedroli, Trimarchi, Gallucci, Chiappini, Goulene, Gaggioli, Riva, Lattanzio, Giunco and Stramba-Badiale. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Agency and Performance of Reach-to-Grasp With Modified Control of a Virtual Hand: Implications for Rehabilitation

Raviraj Nataraj<sup>1,2\*</sup>, Sean Sanford<sup>1,2</sup>, Aniket Shah<sup>1,2</sup> and Mingxiao Liu<sup>1,2</sup>

<sup>1</sup> Movement Control Rehabilitation (MOCORE) Laboratory, Stevens Institute of Technology, Hoboken, NJ, United States,

<sup>2</sup> Department of Biomedical Engineering, Stevens Institute of Technology, Hoboken, NJ, United States

## OPEN ACCESS

### Edited by:

Pietro Cipresso,  
Italian Auxological Institute (IRCCS),  
Italy

### Reviewed by:

Joon-Ho Shin,  
National Rehabilitation Center,  
South Korea  
Maria Vittoria Bulgheroni,  
Ab.Acus, Italy

### \*Correspondence:

Raviraj Nataraj  
rnataraj@stevens.edu

### Specialty section:

This article was submitted to  
Motor Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 25 November 2019

**Accepted:** 19 March 2020

**Published:** 23 April 2020

### Citation:

Nataraj R, Sanford S, Shah A and  
Liu M (2020) Agency and  
Performance of Reach-to-Grasp With  
Modified Control of a Virtual Hand:  
Implications for Rehabilitation.  
Front. Hum. Neurosci. 14:126.  
doi: 10.3389/fnhum.2020.00126

This study investigated how modified control of a virtual hand executing reach-to-grasp affects functional performance and agency (perception of control). The objective of this work was to demonstrate positive relationships between reaching performance and grasping agency and motivate greater consideration of agency in movement rehabilitation. We hypothesized that agency and performance have positive correlation across varying control modes of the virtual hand. In this study, each participant controlled motion of a virtual hand through motion of his or her own hand. Control of the virtual hand was modified according to a specific control mode. Each mode involved the virtual hand moving at a modified speed, having noise, or including a level of automation. These specific modes represent potential control features to adapt for a rehabilitation device such as a prosthetic arm and hand. In this study, significant changes in agency and performance were observed across the control modes. Overall, a significant positive relationship ( $p < 0.001$ ) was observed between the primary performance metric of reach (tracking a minimum path length trajectory) and an implicit measurement of agency (intentional binding). Intentional binding was assessed through participant perceptions of time-intervals between grasp contact and a sound event. Other notable findings include improved movement efficiency (increased smoothness, reduced acceleration) during expression of higher agency and shift toward greater implicit versus explicit agency with higher control speed. Positively relating performance and agency incentivizes control adaptation of powered movement devices, such as prostheses or exoskeletons, to maximize both user engagement and functional performance. Agency-based approaches may foster user-device integration at a cognitive level and facilitate greater clinical retention of the device. Future work should identify robust and automated methods to adapt device control for increased agency. Objectives include how virtual reality (VR) may identify optimal control of real-world devices and assessing real-time agency from neurophysiological signals.

**Keywords:** cognitive agency, reach to grasp, movement rehabilitation, virtual reality, visual feedback

## INTRODUCTION

Sense of agency during movement intuitively leads to better physical function, but it is not a primary rehabilitation target compared to increased strength or practiced skill (Shepherd, 2001; Yang et al., 2006; Timmermans et al., 2009). Powered devices such as exoskeletons (Rosen et al., 2001; Heo et al., 2012) and prosthetics (Childress, 1973; Li et al., 2010), can inject the mechanical energy to physically assist the user. However, functional performance depends on how well the person can control the device toward intended actions. The ability to control these assistive devices primarily depends on a robust command interface from which the user can reliably trigger device actions. The command interface can infer user intention from mechanical triggers such as switches (Bhadra et al., 2002; Peckham and Knutson, 2005). More “natural” interfaces involve command detection from computational processing of recorded physiological signals such as muscle electromyography (EMG) (Boostani and Moradi, 2003) or brain electroencephalography (EEG) (Wolpaw and McFarland, 2004). Despite the interface, functional control is generated from the user’s ability to cognitively integrate their intention with observed device actions toward desired performance outcomes. This study investigated how modifying control of a virtual hand executing reach-to-grasp contributed to performance of functional reach and sense of grasp agency. It was hypothesized that control modes inducing higher agency would also demonstrate greater performance. To verify this relationship as broadly applicable, we investigate control modes that are diverse (changes in speed, presence of noise, addition of automation). Such positive associations should motivate greater consideration of agency in movement rehabilitation.

Sense of agency is defined as the perception of control over actions and related sensory consequences (Moore and Obhi, 2012). Since sensorimotor control of functional movements involves sequences of motor actions continually modulated by sensory feedback (Todorov, 2004), measuring agency by action-consequence events may be especially pertinent and effective in methods to rehabilitate movement. Significant previous work has demonstrated conditions under which sense of agency is generated and modulated (Moore, 2016; Haggard, 2017, 2019). These conditions include voluntary versus involuntary movements (Haggard et al., 2002), matching actual and expected consequences (Frith et al., 2000; Blakemore et al., 2002), and the effects of external cues (Moore et al., 2009). Thus, experimental conditions may be constructed to provide cues that boost agency, but it is unclear if greater agency is related to better movement performance and which conditions may precipitate both. If clear links between agency and movement performance were established, methods to adapt device control for better cognitive engagement and ability with a device may be better pursued. Greater perception of control would naturally engage the user, and user ability is inherently reflected through greater performance. Engagement and ability are vital factors for clinical retention of device-based rehabilitation. Such approaches are especially beneficial for developing sensorimotor prostheses (Marasco et al., 2018) and

powered exoskeletons (Farris et al., 2013) that restore function after neurological trauma. Individuals with brain injury, spinal cord injury (SCI), or amputation may undergo intensive therapy to improve both physical and cognitive skills in re-learning functional movements with devices.

A major advancement in rehabilitation device technology would be the creation of methods that not only optimize user-device mechanics but also cognitive engagement of the user. Systematically identifying user agency and adapting device control accordingly may produce better performing, cognition-driven rehabilitation devices. Ultimately, clinical retention of rehabilitation devices is predicated on user perception of utility (Phillips and Zhao, 1993; Hughes et al., 2014). Methods that leverage perception metrics, such as agency, can also facilitate more usage of rehabilitation devices. Devices for rehabilitation are those that improve movement function for persons with neuromuscular dysfunction. We classify devices either providing powered movement assistance or training for independent function through robotic and computer interfaces as rehabilitation devices. In both cases, greater cognitive engagement and involvement due to user agency in controlling the device should facilitate better, and more natural, performance.

Intentional binding is an established implicit measure for agency. It indicates how coupled one perceives an intended action to an expected sensory consequence (Haggard et al., 2002; Moore and Obhi, 2012). Intentional binding refers to the perceived compression in time between a movement and its consequences during voluntary control (Haggard et al., 2002). The classical construct for intentional binding involved action of a key press to trigger the delayed onset of a sound tone. Participants would judge the time duration between key press and tone. A perceptual shift toward compression of time was shown when the key press was voluntary versus an involuntary twitch induced by transcranial magnetic stimulation. This binding effect is considered implicit since it is specific to voluntary action while passively induced actions can produce a reversal of this effect (Moore et al., 2012). Intentional binding has been used to show the influence of sensorimotor processes on agency through internal prediction and external action outcomes (Haggard et al., 2002; Moore and Haggard, 2008; Moore and Obhi, 2012; Frith and Haggard, 2018). Physical rehabilitation methods could be well served to monitor agency during the recovery and reformulation of sensorimotor pathways after neurotrauma. Intentional binding metrics for agency have already been used for human computer interaction to show the sensitivity of implicit agency to particular input modalities (Coyle et al., 2012; Limerick et al., 2014). Furthermore, it has been shown that brain machine interfaces (BMIs) can generate experiences of explicit agency in users similar to bodily movements (Evans et al., 2015). Explicit agency requires subjects to provide higher-order, conscious assessments of perception of control for given conditions (Moore et al., 2012). Given the sensitivity of both implicit and explicit agency to external cues, a variety of sensory feedback paradigms may be employed to train user-device integration centered on agency. As such, the effects of varying device control on both implicit and explicit agency should be examined.

Virtual reality (VR) is an attractive platform to develop customized methods for user-device integration and agency-based rehabilitation. For the user, VR is proven to enhance cognitive engagement in performing repetitive physical therapy movements (Sveistrup, 2004; Saleh et al., 2017). VR is readily programmable (Todorov et al., 2012) to customize visual projections of user actions and their consequences in functional task performance. Visual feedback from VR can modulate for both sense of agency (Moore and Fletcher, 2012) and control of functional movements like reaching (Desmurget and Grafton, 2000; Saunders and Knill, 2003; Nataraj et al., 2014b) and grasping (Winges et al., 2003; Nataraj et al., 2014a). Reach-to-grasp is a fundamental human action and is commonly targeted for rehabilitation following neuromuscular dysfunction (Lin et al., 2007; Loureiro and Harwin, 2007) and can be assisted with powered devices triggered by user command actions (Popovic, 2003; Kotecha et al., 2014). With neurotrauma such as SCI, visual capabilities are still largely intact and can be leveraged further in VR to partially compensate loss of other senses (Ghez et al., 1995) such as touch and proprioception. For rehabilitation devices, such as prostheses and exoskeletons, VR platforms can be flexibly constructed to train complex interfaces involving direct physiological access (Kuiken et al., 2009; Marasco et al., 2018) or powered actuation of limbs (Hartigan et al., 2015). VR could be employed to match user intentions to optimal parameters for controlling a device using visual projections of device actions following user commands. Control parameters include feedback gains to maximize performance and minimize effort (Nataraj and van den Bogert, 2017) and to achieve desired movement features such as smoothness (Hogan and Sternad, 2009). Ultimately, VR platforms may be utilized to efficiently identify control parameters of rehabilitation devices that optimize not only functional mechanics but also user agency prior to eventual translation to real-world systems (Caldwell et al., 1995, 1998; Bar-Cohen, 2003; Perry et al., 2007).

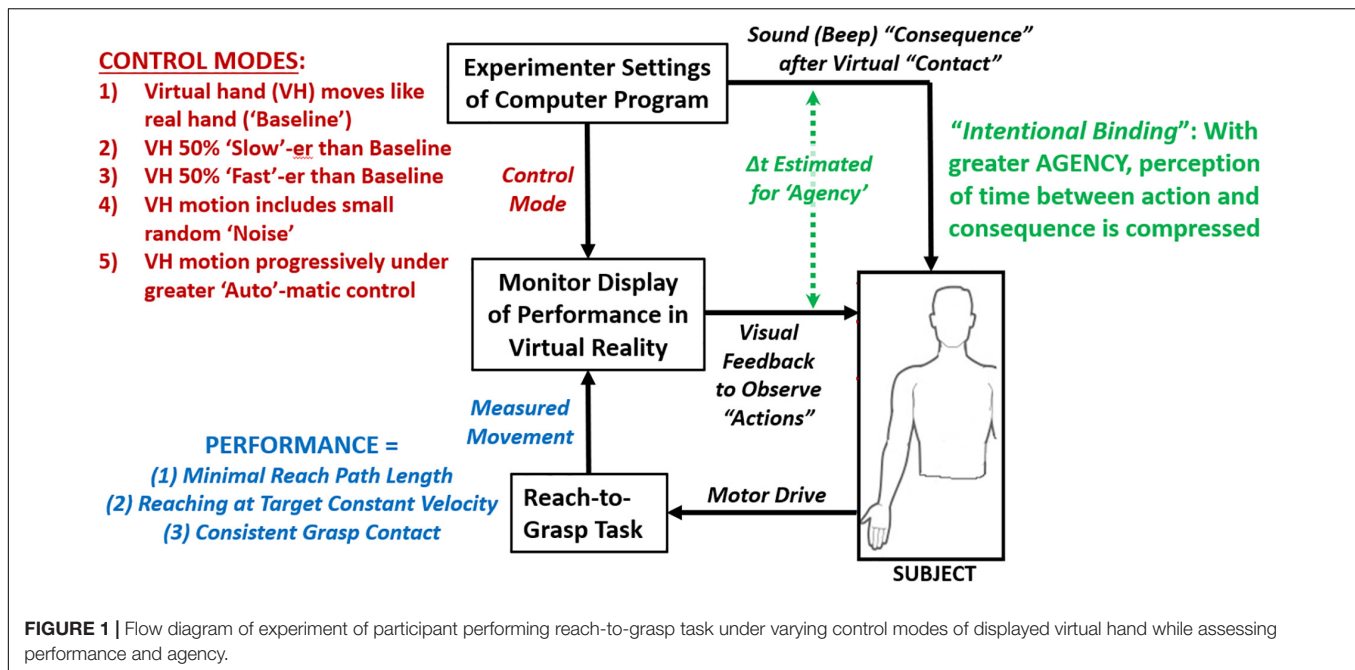
In this study, a VR environment was utilized to couple reach and grasp “actions” to programmed sensory “consequences” (visual and sound events). Participants triggered movement control of the virtual hand through movement of their own hand. The visually observed movement of the virtual hand depended on the specific control mode. The control mode defined at what fixed speed the virtual hand would move proportional to the real hand and if virtual movement included noise or assisted automation. We investigated how changes in user control of a virtual hand prosthesis (Johannes et al., 2011) during reach-to-grasp may generate effects across both sense of agency and functional task performance. Visual cues informed the participant about initiating and pacing the reach, where to grasp, and when grasp action was successfully completed. The primary performance metric was reducing position error of the participant’s hand to a minimal path-length trajectory at a fixed velocity. As with previous intentional binding studies (Moore and Obhi, 2012), a sound cue (beep) was used as the consequence to an intended action (grasp). Participants provided verbal estimates of lapsed time intervals between action and consequence to infer agency implicitly via intentional binding across the various control modes. The control modes of the

virtual hand were consistent with parameters commonly adapted for a movement rehabilitation device, and included: setpoints for speed (Blaya and Herr, 2004; Wege et al., 2005), noise mitigation (Taylor et al., 2002; Agostini and Knaflitz, 2012), and a level of automated assistance (Ronsse et al., 2010, 2011). Speed, noise, and automation are fundamental control parameters that a device engineer can *ad hoc* tune based on stated user preferences or anecdotal observation of performance (Terenzi, 1998). Alternatively, these parameters can also be determined through optimization of mechanical performance (e.g., effort, tracking) for a model system (Davoodi et al., 2007; Nataraj and van den Bogert, 2017). Neither approach systematically adapts the device according to user agency. The major implication of this study is how a subjective metric of perception in control of a virtual device (hand) can be related to objective performance (reaching) with that device. In this study, the control modes were enacted as deviations from an optimal (“Baseline”) mode, at which the virtual hand moved to match the actual hand movements and agency is expected to be highest.

Unlike previous studies that identified agency for movement initiation (Haggard et al., 2002), this study investigated how agency of grasp execution was modulated by the control mode of the preceding reach. In this way, it was inferred how control *during* reaching may facilitate or inhibit agency of the terminating action of grasp and performance of the reach itself. Previous studies have shown the direct link of agency between continuous movements and terminal events (Wen et al., 2015; Oishi et al., 2018). In this study, we prioritized and considered implicit agency by time-interval estimation as a less biased (more sub-conscious) perceptive measure. With time-interval estimation, a quantifiable measure was provided each trial that was not readily linked to a conscious preference to a control mode. The main hypotheses of this VR reach-to-grasp study were: (1) implicit grasp agency and reaching performance are positively related across a broad class of control modes typically considered for rehabilitation devices, (2) significant differences in both implicit agency and performance are observable between these control modes. While our primary hypotheses considered implicit agency, we additionally examined explicit perception of each control mode with Likert-scale survey responses. The purpose of the survey responses was to observe how implicit and explicit agency may be related through the presented control modes of this study. Another important implication of translating agency to more effective rehabilitation device control is greater performance efficiency. Thus, the secondary hypotheses of this study were: (1) there are significant shifts between implicit and explicit agency across control modes, (2) agency is positively related to performance efficiency, and (3) significant differences in efficiency exist between control modes.

## MATERIALS AND METHODS

In this experimental protocol, participants controlled a virtual hand to perform reach-to-grasp through movements of their own hand (**Figure 1**). The observed movement of the virtual hand were initially based on those of the real hand (“Baseline” case)



but modified depending on the other control modes tested. The modifications from Baseline involved fixed changes in speed, addition of noise, or inclusion of automation. Participants were asked to maximize performance (primarily moving own hand to minimize reaching path length at a target velocity) and provide verbal estimates of perceived time-intervals between grasp action and a sound consequence for implicit assessment of agency.

## Participants

A total of 16 able-bodied volunteers (12 male, 4 female,  $20.9 \pm 3.2$  years) were recruited to participate in this study. A power analysis for one-way ANOVA at 95% suggested that eight-participant samples would show significant differences ( $\alpha = 0.05$ ) in implicit agency and reaching performance. In this power analysis, performance was for minimizing path length (see "Data and statistical analysis") across the tested control modes during a pilot study (Shah et al., 2018). Only right-handed participants were tested for right-hand reach-to-grasp to avoid considering effects of hand dominance. All participants had normal or corrected-to-normal vision and did not previously report nor demonstrate a history of disease, injury or complications involving cognition or upper extremity function. All participants signed an informed consent form approved by the Stevens Institutional Review Board.

## Equipment (Hardware and Software)

A marker-based motion capture system was used to track 3-D hand motions and correspondingly control a virtual model of a prosthetic hand [MPL, *Modular Prosthetic Limb* (Johannes et al., 2011)]. The hand was viewed in a VR environment with advanced contact mechanics [*Multi-Joint Dynamics with Contact*, MuJoCo, Roboti LLC, Seattle, Washington, United States (Todorov et al., 2012)]. The motion capture system included nine infra-red

cameras (*Prime 17W* by Optitrack, NaturalPoint Inc., Corvallis, OR, United States) to track 3-D position and orientation of three retroreflective marker clusters. The first cluster included three markers (9 mm diameter) that were Velcro-affixed in a non-colinear arrangement on a worn glove at the dorsal side of the hand (midpoint of third metacarpal). This "hand" cluster served as a reference coordinate system mapping real-time changes in position and orientation to the virtual hand. Similarly, two additional clusters with smaller markers (4 mm diameter) were placed on the nails of the index finger and thumb. These nail clusters were affixed to 3-D printed platforms that attached to the nails using double-side adhesive tape. Coordinate systems represented by these nail clusters drove position and orientation of the distal segments of the respective digits. Joint angle changes across the digits were based on real-time inverse kinematics solutions sufficiently satisfying the position and orientation constraints of all three clusters. Position constraints for the nail clusters were relative to the hand cluster and scaled for each participant hand size to match the virtual hand size. Only the thumb and index finger were tracked and animated on the virtual hand as the functional task was reach to precision grasp (Nataraj et al., 2014a), requiring focus onto smaller objects. Real-time streaming of marker data to manipulate the VR environment was done using the motion capture software (*Motive* by Optitrack) and API code written in MATLAB (Mathworks Inc., Natick, MA, United States) running on a Dell Workstation. All data was processed at 120 Hz.

## Protocol

### Participant Preparation

Upon arriving to the laboratory, participants were re-informed about protocol and their right-hand size was measured. Hand size was measured as the maximum spread distance from tip



of thumb to tip of index finger. The average hand size was  $15.2 \pm 0.95$  cm. For each participant, hand size was used to spatially calibrate motions of the index finger and thumb clusters relative to the hand cluster of the real hand to those of the virtual hand. Each participant was seated with chair height adjusted so that the reaching arm would be table-supported to initially have: the elbow at a right angle, shoulders comfortably level, and upper-arm at the participant's side (**Figure 2A**). Each participant then wore a glove (**Figure 2B**) with hand marker cluster attached. A marker cluster was then added to each of the index finger and thumb nails. The participant then had placed over their head and eyes an Oculus® *Rift* headset (Facebook Technologies, LLC) displaying a custom virtual environment (*MuJoCo*) as seen in **Figure 2C**. The participant then had placed over their ears a noise canceling headset (Bose® *QuietComfort 35*) to minimize audible distractions and primarily only hear an occasional beep tone (sound consequence) as part of the experimental task.

### Virtual Reality Calibration Procedures

The Oculus display filled the participant's entire field of view with the virtual environment. Participants were able to find an initial starting position for their real hand based on tactile sensation of a Velcro strip on the support table. The view within the virtual environment was initially calibrated such that the hand marker cluster position of the real hand was coincident with the same landmark position of the virtual hand. In front of the participant's virtual view was a sphere (7 cm diameter) that served as the target the participant reached toward and grasped each trial. The virtual sphere was located 20 cm above and 25 cm anterior to the initial hand cluster position. Two tracks for speed pacers were also within view. One pacer moved forward and the other vertically to inform the participant about the target hand velocity in each dimension. The tracks were semi-transparent to subtly cue the participant about speed without distracting visual focus from the virtual hand. The pacer speeds were set to traverse each dimension in 4 s.

### Virtual Reality Task

Each trial, the participant was cued by countdown to begin performing reach-to-grasp (**Figure 3**). The countdown for a trial was represented by color transitions of the target sphere as follows: red at trial time ( $t$ ) = -2 s, to yellow at  $t$  = -1 s, and to green at  $t$  = 0 sec, at which time the speed pacers, moving at constant velocity, began to move and the participant should initiate hand movement. The pacers ceased movement after  $t$  = 4 s or earlier when the participant made premature grasp contact. Participants were told to maximize reach-to-grasp performance across three criteria: (1) minimize reaching path length, (2) match hand reaching velocity to speed pacers and complete reach-to-grasp in precisely 4 s, and (3) grasp the target sphere with thumb and index finger at consistent locations. Participants were told that reaching performance was primarily evaluated in this study but to self-consider all three performance criteria to promote task consistency. Each trial lasted up to 10 s as the participant had 7 s to complete reach-to-grasp with the goal to complete in precisely 4 s. Although natural reach-to-grasp is executed nominally at 1 s (van Vliet and Sheridan, 2007), reaching time

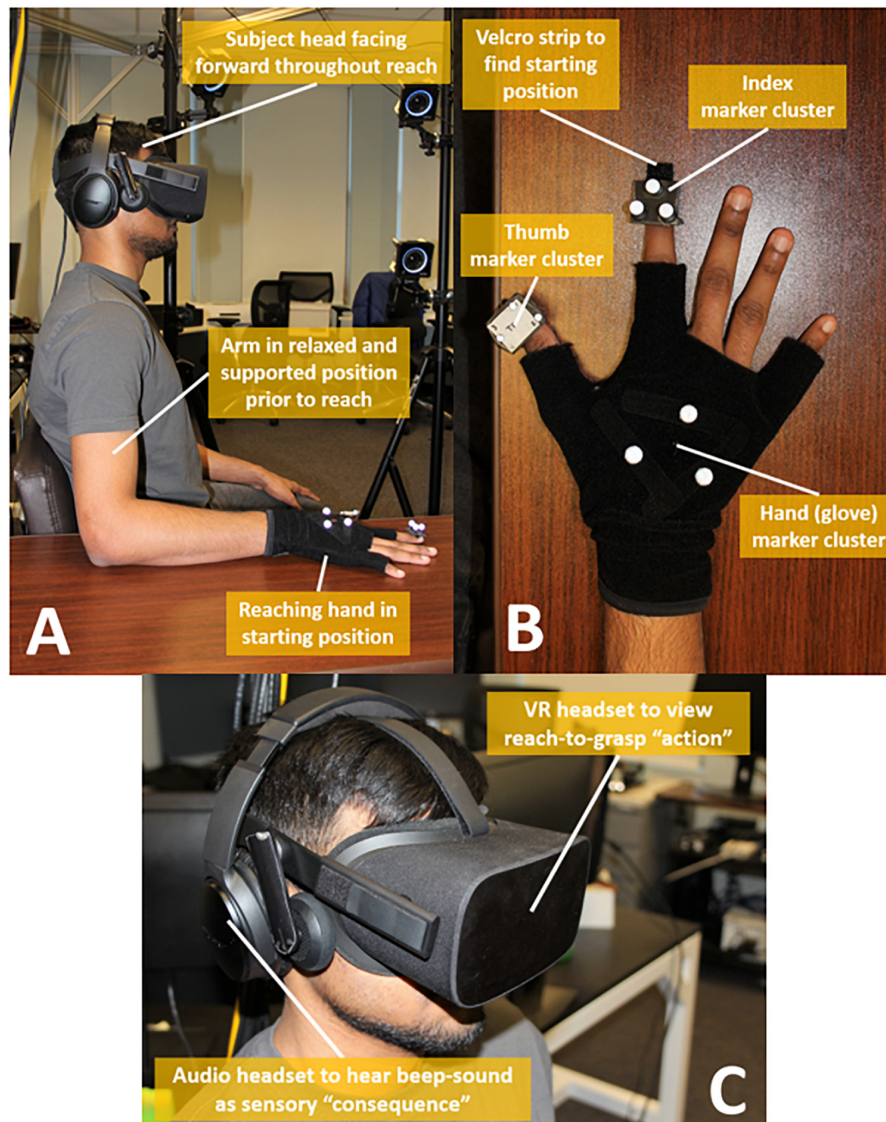
with a neural controlled robotic device can be notably slower ( $\sim 6$  s) (Hochberg et al., 2012). In this study, ecological validity for reach performance and grasp agency was intended more for device control.

When the virtual hand grasped the target sphere with both the index and thumb digits, the sphere instantly changed color from green to black and the virtual environment froze in place. This color-change event cued the participant that grasp action was successfully completed. A short-duration ( $\sim 100$  ms), moderate-pitch beep was sounded to the participant's headset at a variable time-interval following grasp action. The participant was asked to verbally estimate the time-interval to the best of their abilities after each beep. The participant was previously instructed that the interval for each trial was anywhere from 100 to 1000 ms in denominations of 100 ms. The actual intervals were always 100, 300, 500, 700, or 900 ms. For each block of trials to test a specific control mode, the number of trials presented at a given time-interval was based on a Gaussian distribution centered at 500 ms. This approach in presenting time-intervals was modified from previous intentional binding experiments that assessed agency with a uniform distribution of intervals at 300, 500, and 700 ms (Caspar et al., 2015). Pilot data revealed that these modifications facilitated a distribution of estimates necessary to infer differences in agency across several control modes (6 in all, see next section). Greater underestimation of time-intervals indicated greater compression (shortening) of the perceived time-interval and implicitly demonstrated greater agency (Haggard et al., 2002). As with other intentional binding experiments, our implicit measure of agency served as a more sub-conscious perception of control.

### Varying Control Modes

Each participant performed the reach-to-grasp task under six different control modes. As previously described, the control modes examined in this study considered modifications in speed, addition of mild noise, and automation. The participant was aware of each control mode being tested through visual feedback of the virtual hand in reference to their own moving hand. The test cases of control modes were as follows:

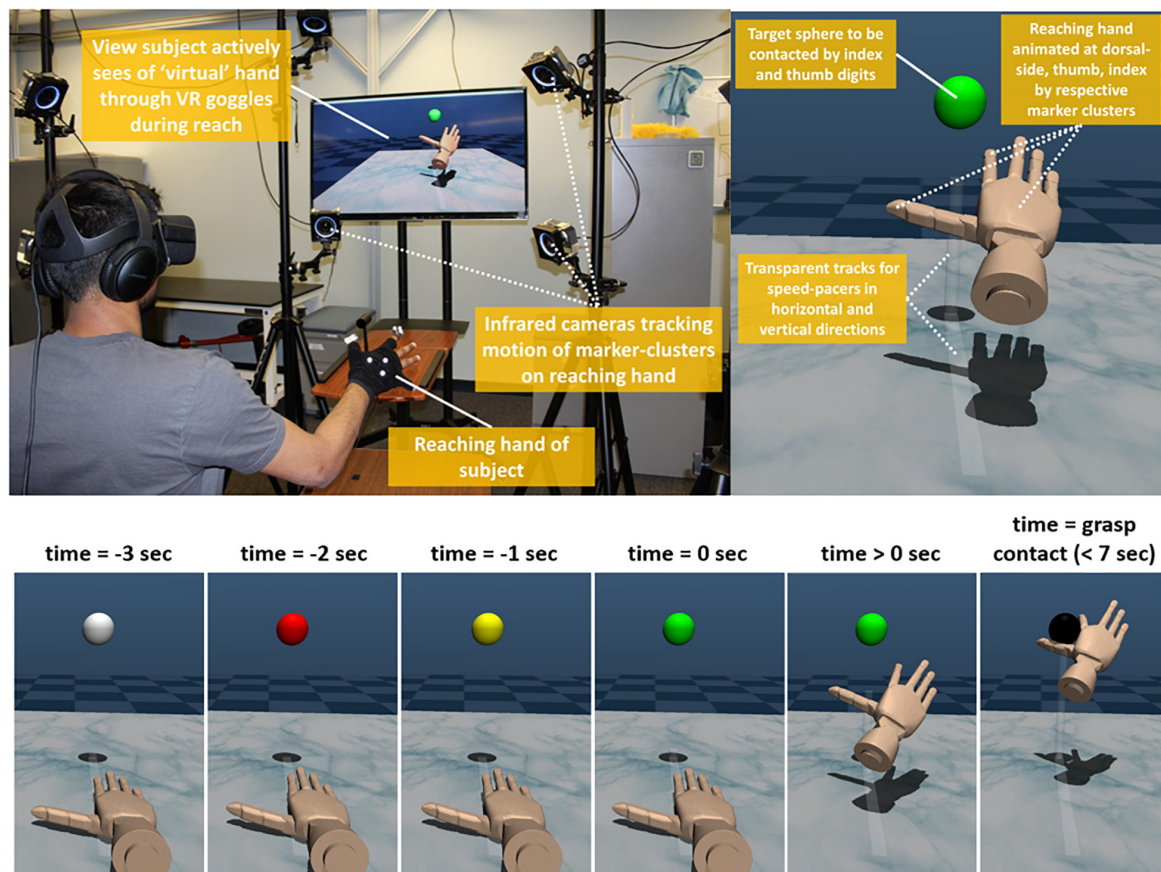
- (1) **Grasp-Only** –The virtual hand was initially placed near the target sphere whereby no reach was required, and only grasp action was needed to complete the trial. This test case served as a control for all subsequent control modes to observe how a preceding movement phase (reach) may affect agency of the terminal action (grasp). As previously described, other studies have investigated agency in relation to the intent to move with a simple key press (Haggard et al., 2002). Our study examined a more complex functional task with two components (reach and grasp) whereupon movement was already initiated prior to grasp. This case with grasp-only was analogous to initiation of key press. All other control modes tested included reach and grasp.
- (2) **Baseline** –The virtual hand moved in equal proportion to the real hand in all three dimensions. This control mode was tested once at the beginning of the session (after grasp-only) and repeated at the end. The first test block was



**FIGURE 2 |** Experimental set-up elements. **(A)** Participant positions body and arm at start of each trial, **(B)** Hand shown with motion capture marker clusters, **(C)** Participant head mounted with VR headset and audio headset.

used for comparison to other cases. The second block was done to compare agency and performance to the first block and verify possible changes due to fatigue or learning across the session.

- (3) **Slow** –The virtual hand moved in all three dimensions at a speed that was 50% slower than the real hand. The virtual hand appeared “sluggish,” and the participant needed to move the real hand 50% faster and further as compensation to control the virtual hand and complete reach-to-grasp as intended.
- (4) **Fast** –The virtual hand moved in all three dimensions at a speed that was 50% faster than the real hand. The virtual hand appeared “hyperactive,” and the participant moved the real hand 50% slower and shorter to compensate and control the virtual hand as intended.
- (5) **Noise** –The virtual hand was infected by mild to moderate noise. A small random value was added in each of the three dimensions for the position of the real hand. The random value was  $\pm X$ , where  $X = 10 \times \text{displacement from previous time-step}$ . Given the sampling frequency of 120 Hz, the noise amplitude was proportional to hand velocity as  $\pm 1$  cm per 12 cm/sec. This noise-level produced light visual tremor to the moving hand that was clearly noticeable but not overtly distracting or challenging to complete the reach-to-grasp task.
- (6) **Auto** –The virtual hand was progressively (linear with time) under automatic control. At the start-time of reach ( $t_{\text{reach}} = 0$ ), the participant controlled the virtual hand just as in “Baseline.” Over the designated 4-s reach duration, the position of the virtual hand ( $pos_{\text{VR-hand}}$ ) was a weighted



**FIGURE 3 |** Virtual reality environment. (Top Left) Participant actively reaches and concurrently views virtual hand, (Top Right) Close-up of virtual hand reaching to target sphere against transparent speed pacer tracks, (Bottom) Sphere color changes with hand transitions across trial time  $t = -3$  to  $+7$  s (10 s total). Countdown occurs from  $t = -3$  to 0 s. After countdown, hand should be in “motion” during time sphere is green.

average of the participant's real hand position ( $pos_{subj}$ ) and a pre-defined optimal position ( $pos_{opt}$ ) corresponding to the minimal path trajectory. The virtual hand position was given as:  $pos_{VR-hand} = \left(1 - \frac{t_{reach}}{4}\right) \times pos_{real} + \left(\frac{t_{reach}}{4}\right) \times pos_{opt}$ . At  $t_{reach} = 4$  s, the virtual hand was guaranteed to be very near the sphere, but the participant must still volitionally perform grasp to complete the trial. This automated case was akin to user initiation of movement to trigger device assistance and auto-complete the movement (Lucas et al., 2004).

### Experimental Testing Blocks

Participants would perform a block of 20 consecutive trials for each of the six control modes. The first three trials of every block were “practice” with the time-interval between grasp contact and the beep fixed at 1 s. The participant was aware these practice trials served to gain mild familiarity with the control mode and to re-calibrate their internal reference of a 1 s time-interval. The remaining 17 test trials were used for agency and performance assessment with time-intervals to be estimated ranging from 100 to 1000 ms as previously described. After each trial, the VR hand was reset to the initial position prior to the 3-s countdown to

initiate movement for the second trial. Each participant was given up to 5 min between blocks to rest and complete a survey to rate their experience for that block.

### Surveys

After each block, the participant was presented with a 1-statement survey to express their subjective perception of the control mode presented. Participants were asked to rate, on a 5-point Likert scale ( $-2$  = strongly disagree,  $+2$  = strongly agree), to what extent they agreed that the visualized hand motions reflected their intentions. The specific statement read “the visualized hand motions reflected your intentions.” The survey responses served as an explicit, or conscious, measure of agency (Moore et al., 2012; Dewey and Knoblich, 2014) for each control mode. The single survey was presented at the end of each block to ensure subjects accommodated to a control mode prior to making a conscious subjective assessment.

### Data and Statistical Analysis

The primary performance metric evaluated across control modes was the inverse of path length error to a minimal path length trajectory occurring at constant velocity over 4 s. The total 3D minimal pathlength was 0.32 m, and for completion in



4 s, the target constant velocity is 0.08 m/s. The total error in three dimensions (3D) was computed for the position of the hand cluster from the target position trace over time. In each dimension, the target trajectory was a linear (constant velocity) position trace that directly (straight line) connects the initial hand position to a position near the sphere from which it can immediately be grasped. The time course of each target trajectory was coincident with the 4 s duration of the constant-speed pacers. Additional performance metrics evaluated in this study involved efficiency of movement. These metrics included greater smoothness (Hogan and Sternad, 2009) and lower 3D acceleration given a constant velocity target. These movement performance metrics were explicitly computed for each trial as follows:

**Pathlength**(over entire reach)  $\rightarrow$

$$P = \sum_{i=1}^N \sqrt{(px_{i+1} - px_i)^2 + (py_{i+1} - py_i)^2 + (pz_{i+1} - pz_i)^2}$$

where

$i$  = time index

$N$  = total number of time-points until grasp contact at sampling frequency (120 Hz)

$px, py, pz = x, y, z$  position of hand marker-cluster

**Inverse Pathlength**  $\rightarrow P^{-1} = \frac{1}{P}$

**Kinematics** (at each time index)  $\rightarrow$

$$vx_{i+1} = \frac{px_{i+1} - px_i}{\Delta t}, ax_{i+1} = \frac{vx_{i+1} - vx_i}{\Delta t}, jx_{i+1} = \frac{ax_{i+1} - ax_i}{\Delta t}$$

where

$vx, ax, jx$  = velocity, acceleration, and jerk of hand marker-cluster in  $x$ -dimension (repeated for  $y$ - and  $z$ -dimension) at given time index.  $\Delta t = 1/120$  s. A moving mean window of 12 time points (0.1 s given sampling frequency of 120 Hz) was employed for smoothing kinematic trajectories.

**Total 3D Acceleration** (at each time index)  $\rightarrow$

$$Acc_i = \sqrt{ax_i^2 + ay_i^2 + az_i^2}$$

**Total Smoothness** (over entire reach)  $\rightarrow S_{tot} = Sx + Sy + Sz$

where

$Sx = \sum_{i=1}^N jx_{i+1}^2$  (smoothness in each dimension, e.g.,  $x$ -dimension)

$Sx' = Sx \frac{D^3}{vx^2}$  (unitless smoothness in each dimension)

$D$  = total duration of reach

$vx$  = mean velocity in  $x$ -dimension during reach

**Inverse Smoothness**  $\rightarrow S_{tot}^{-1} = \frac{1}{S_{tot}}$

**Controller Efficiency**  $\rightarrow CE = \frac{S_{tot}^{-1}}{Acc}$

where

$Acc$  = mean total 3D acceleration during reach

The following statistical analyses were performed:

- For comparisons across tested control modes, a Kolmogorov–Smirnov confirmed normality in analyzed data sets and use of parametric statistical tests. Repeated-measures one-way ANOVA was done independently on data for agency and each reach-to-grasp performance metric across the single factor of control modes. *Post hoc* comparisons between paired test cases were made with Bonferroni correction for multiple comparisons. The  $p$ -value, F-statistic, and eta-squared metric were reported for significance and effect size.
- In assessing dependence of a performance metric on agency, a linear regression analysis was applied to identify evident relationships of performance or explicit agency to implicit agency. The F-statistic and  $p$ -value was computed to refute the null hypothesis that the slope coefficient was equal to zero and suggest significant dependence on implicit agency. A significant non-zero slope indicated a simple relationship between either a performance metric or explicit agency to implicit agency. The actual slope value indicated the magnitude of dependence of each variable on implicit agency.
- An unpaired  $t$ -test (two-tailed) was used to assess possible significant difference in agency between the grasp-only and the five test cases for reach-to-grasp.

A paired  $t$ -test (two-tailed) was used to assess difference in agency and performance between the Baseline test block at the start of the session versus the end of the session.

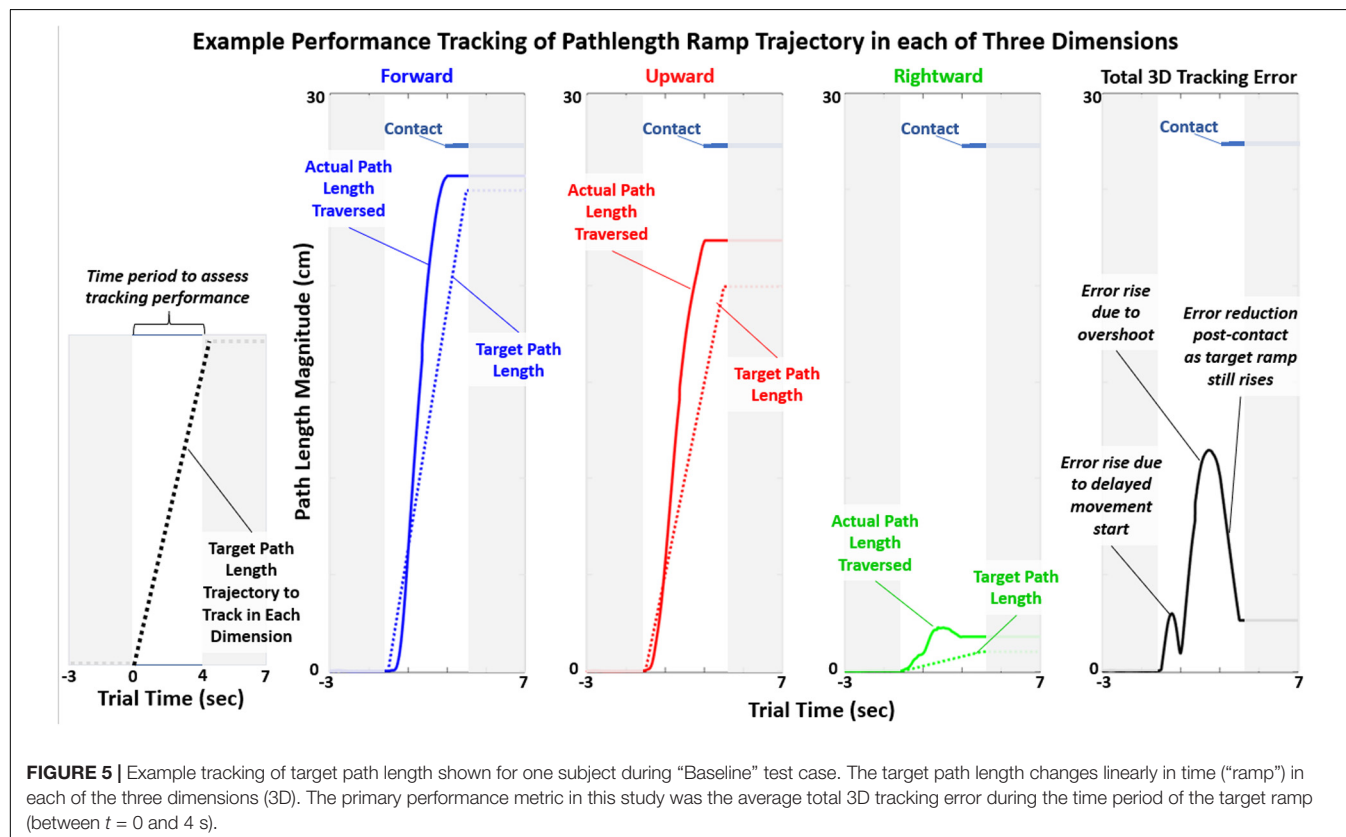
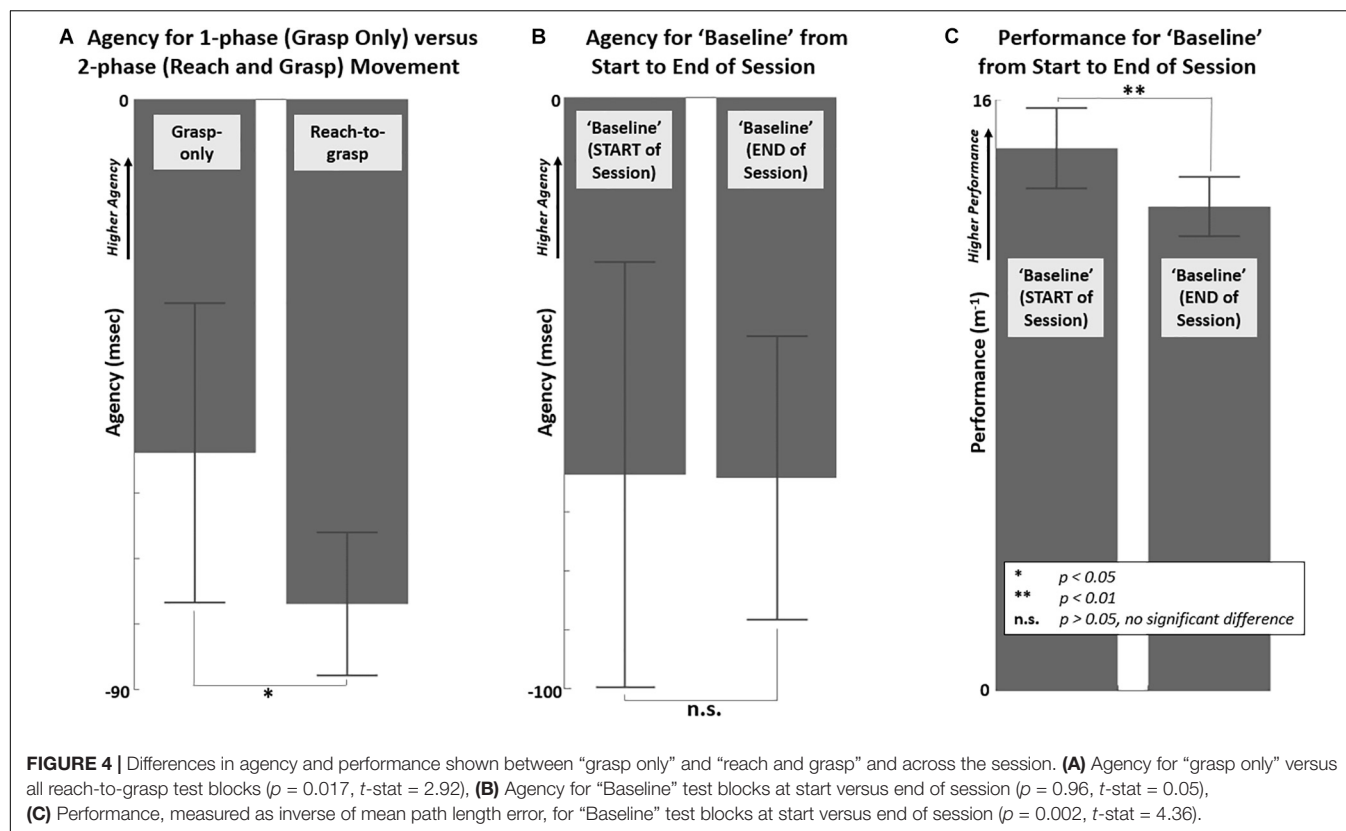
## RESULTS

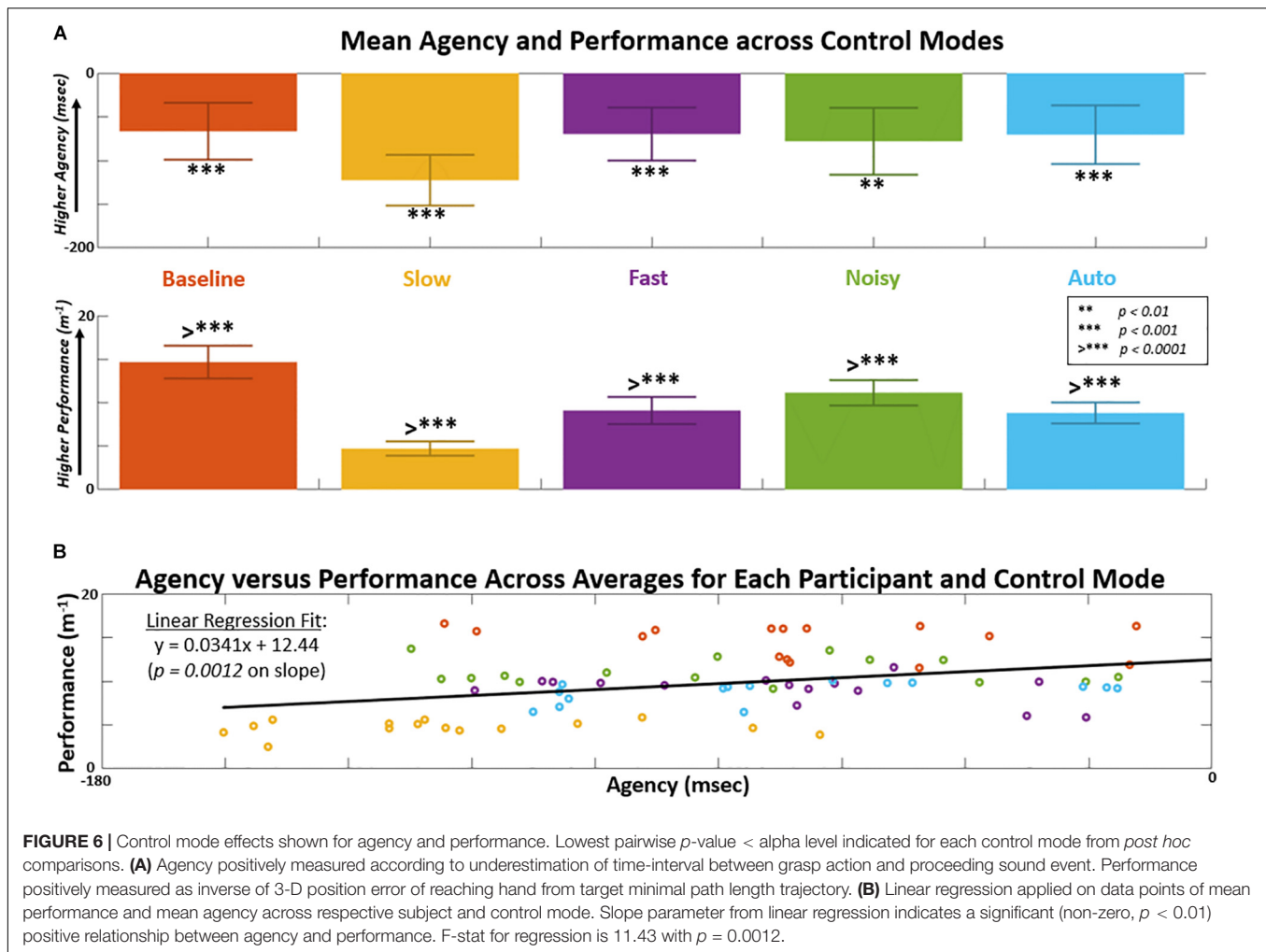
This study demonstrates the effects of varying control modes of a virtual hand on agency and performance of reach-to-grasp. Results are organized as follows: preliminary considerations of agency and performance of the reach-to-grasp task, agency and performance across control modes, changes in movement efficiency (e.g., smoothness) across control modes, and path length kinematics during high agency versus low agency.

### Preliminary Considerations of Reach-to-Grasp Agency and Performance

The reach phase decreased agency of grasp compared to the grasp-only test case as shown in **Figure 4A**. No significant change in agency was observed for between the Baseline test blocks across the session (**Figure 4B**). There was a significant reduction in reaching performance (inverse of mean error to minimal path length trajectory) between the Baseline test block from start ( $14.7 \text{ m}^{-1}$ ) to end ( $13.1 \text{ m}^{-1}$ ) of the session (**Figure 4C**). Due







to the observed reduction in Baseline performance, performance data across the session were adjusted by a linear correction factor. The correction factor was applied uniformly across sequential test blocks proportional to the reduction in Baseline performance from start to end of the session.

## Effect of Control Mode on Agency and Reaching Performance

The mean total 3D tracking error of the target minimal pathlength across time was the primary performance metric in this study. Example performance to track a minimal path length trajectory is shown in **Figure 5**. There was typically a delay in movement initiation despite a preparatory countdown cue. There was also tendency to move the virtual hand faster than the target constant velocity. This resulted in a quick overshoot of the target and completion of contact prior to completion of the target ramp trajectory.

One-way ANOVA indicated significant differences in both agency ( $p < 0.001$ ) and performance ( $p < 0.0001$ ) across the single factor of control modes (**Figure 6A** and **Table 1**). The highest mean value in agency and performance was observed

for the Baseline control mode. The lowest mean value in agency and performance was observed for the Slow control mode. The F-stat for both agency and performance were notably greater than 1 and with notable effect size ( $\eta^2 > 0.30$ ). A linear regression was applied to subject-averaged sample points for agency versus performance across all control modes tested (**Figure 6B**). The slope parameter was significantly greater than zero ( $p < 0.01$ ) indicating a positive relationship between agency and performance.

Implicit measures of agency using intentional binding are shown against survey-based explicit measures of agency in **Figure 7** and **Table 2**. Significant differences ( $p < 0.05$ ) in explicit agency were not observed across control modes (**Figure 7B**). Implicit and explicit agency results across subject-mode pairs were self-normalized [mean = 0, range over  $(-1, 1)$ ] and plotted against each other in **Figure 7C** to suggest an inverse relationship (linear regression slope < 0,  $p < 0.05$ ) in this study. The average difference in normalized explicit agency from implicit agency for each control mode is shown in **Figure 7D**. Across control modes, the normalized differences between explicit and implicit agency produced notable F-stat (9.88) and effect size ( $\eta^2 = 0.36$ ). The largest differences were observed for the

**TABLE 1A |** Mean value comparisons for implicit agency and performance of minimizing reach pathlength across control modes.

Metric	Control mode					ANOVA		
	Baseline	Slow	Fast	Noisy	Auto	F-Stat	p-val	$\eta^2$
Implicit agency (ms)	-66 ± 32	-122 ± 29	-69 ± 30	-77 ± 38	-70 ± 33	7.58	3.94E-05	0.30
Performance (m <sup>-1</sup> )	14.7 ± 1.8	4.7 ± 0.8	9.1 ± 1.6	11.1 ± 1.5	8.8 ± 1.2	96.6	9.81E-28	0.85

**TABLE 1B |** *Post hoc* comparisons (*p*-values) between control modes for implicit agency.

Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	–	<b>1E-04</b>	0.99	0.87	0.99
Slow	–	–	<b>4E-04</b>	<b>4E-03</b>	<b>4E-04</b>
Fast	–	–	–	0.96	0.99
Noisy	–	–	–	–	0.97

**TABLE 1C |** *Post hoc* comparisons (*p*-values) between control modes for performance (minimizing reach pathlength).

Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	–	<b>1E-08</b>	<b>1E-08</b>	<b>4E-08</b>	<b>1E-08</b>
Slow	–	–	<b>1E-08</b>	<b>1E-08</b>	<b>1E-08</b>
Fast	–	–	–	<b>1E-08</b>	0.98
Noisy	–	–	–	–	<b>3E-04</b>

All *post hoc* comparisons made with Bonferroni correction. Significant *post hoc* *p*-values (<0.05) bolded.

Slow and Fast mode with a shift toward explicit and implicit agency, respectively.

## Effect of Control Mode on Movement Efficiency

The mean kinematic trajectory for reach in each direction is shown for Baseline in **Figure 8**. Given the reach-to-grasp task is continuous with clear initiation and termination, the movement smoothness was computed based on minimization of integrated squared-jerk (Flash and Hogan, 1985) for each control mode. To remove dependencies on movement duration or amplitude, the squared-jerk term is made unitless (Hogan and Sternad, 2009) based on movement time and mean velocity in each direction.

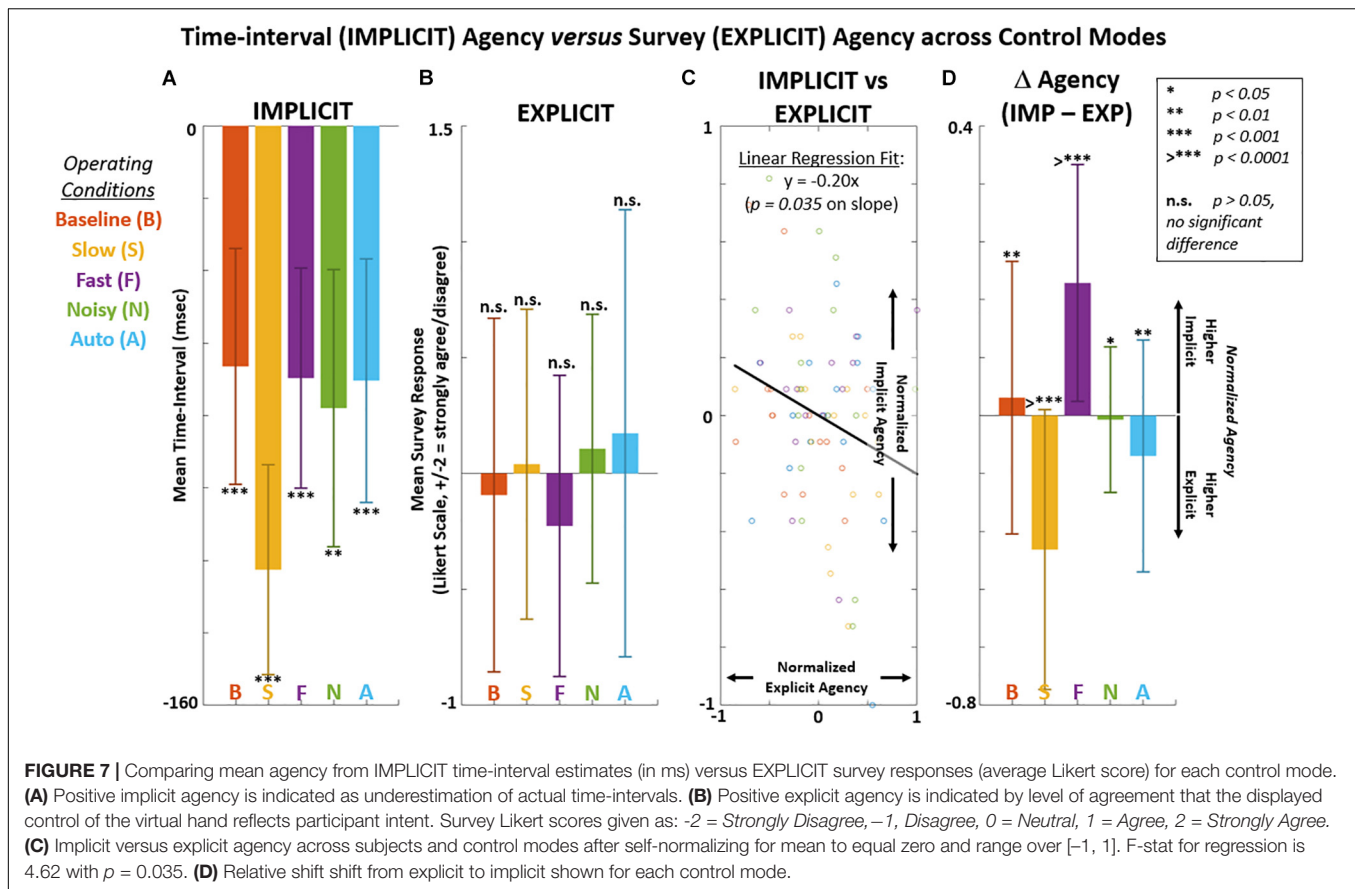
Results for select metrics of movement efficiency across control modes are shown in **Figure 9** and **Table 3**. Smoothness (**Figure 9A**) is shown as the inverse of the integrated unitless squared-jerk metric summed in all three directions. The inverse operation presents higher smoothness by higher positive value. Highest smoothness was observed for the Slow control mode. However, the highest total 3-D acceleration (**Figure 9B**) was also observed for the Slow control mode. Higher acceleration indicates greater corrections were made online in tracking a constant-velocity movement target. When smoothness is normalized by total 3-D acceleration (**Figure 9C**), then the highest smoothness per unit acceleration was achieved during the Baseline and Fast control modes.

Higher smoothness per unit acceleration suggests greater sensitivity of efficiency to a given correction, i.e., “correction sensitivity.” Correction sensitivity is plotted against agency for data points across subjects and control modes in **Figure 9D**. A linear regression on that data indicates a positive relationship (slope > 0, *p* < 0.05) between correction sensitivity and agency.

## Effect of High Versus Low Agency on Path Length Kinematics

The general effects of high versus low implicit agency on path length position and velocity over the reach cycle are shown in **Figure 10**. The mean path kinematic trajectories are shown across the top (high) 50% of trials in agency versus the bottom (low) 50% of trials across all participants and control modes. High agency trials generally demonstrate shorter path length trajectories and slower path length velocities throughout the reach cycle.

**Figure 11** indicates that high agency trials produce significant (*p* < 0.001) reductions in the following movement features of path length: maximum path length, mean path length velocity, and maximum path length velocity. These high agency effects were desirable given the performance task was to minimize path length, ideally by following a minimum path length trajectory of 0.32 m at a constant velocity of 0.08 m/s. **Figure 11** also indicates a significant increase (*p* < 0.05) in movement smoothness in path length with high agency.



**FIGURE 7 |** Comparing mean agency from IMPLICIT time-interval estimates (in ms) versus EXPLICIT survey responses (average Likert score) for each control mode. **(A)** Positive implicit agency is indicated as underestimation of actual time-intervals. **(B)** Positive explicit agency is indicated by level of agreement that the displayed control of the virtual hand reflects participant intent. Survey Likert scores given as: -2 = Strongly Disagree, -1 = Disagree, 0 = Neutral, 1 = Agree, 2 = Strongly Agree. **(C)** Implicit versus explicit agency across subjects and control modes after self-normalizing for mean to equal zero and range over [-1, 1]. F-stat for regression is 4.62 with  $p = 0.035$ . **(D)** Relative shift from explicit to implicit shown for each control mode.

**TABLE 2A |** Mean value comparisons for implicit and explicit agency across control modes.

Metric	Control mode					ANOVA		
	Baseline	Slow	Fast	Noisy	Auto	F-Stat	p-val	$\eta^2$
Explicit Agency (Likert)	$-0.09 \pm 0.76$	$0.04 \pm 0.67$	$-0.23 \pm 0.65$	$-0.11 \pm 0.58$	$0.17 \pm 0.96$	0.7113	0.59	0.04
Normalized $\Delta$ Agency (Implicit - Explicit)	$0.03 \pm 0.24$	$-0.23 \pm 0.24$	$0.23 \pm 0.20$	$-0.007 \pm 0.13$	$-0.07 \pm 0.20$	9.88	2.2E-06	0.36

**TABLE 2B |** Post hoc comparisons ( $p$ -value) between control modes for difference (shift) in normalized agency,  $\Delta$ agency = implicit - explicit.

Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	—	<b>7E-03</b>	7E-02	0.99	0.67
Slow	—	—	<b>5E-07</b>	<b>3E-02</b>	0.21
Fast	—	—	—	<b>2E-02</b>	<b>2E-03</b>
Noisy	—	—	—	—	0.92

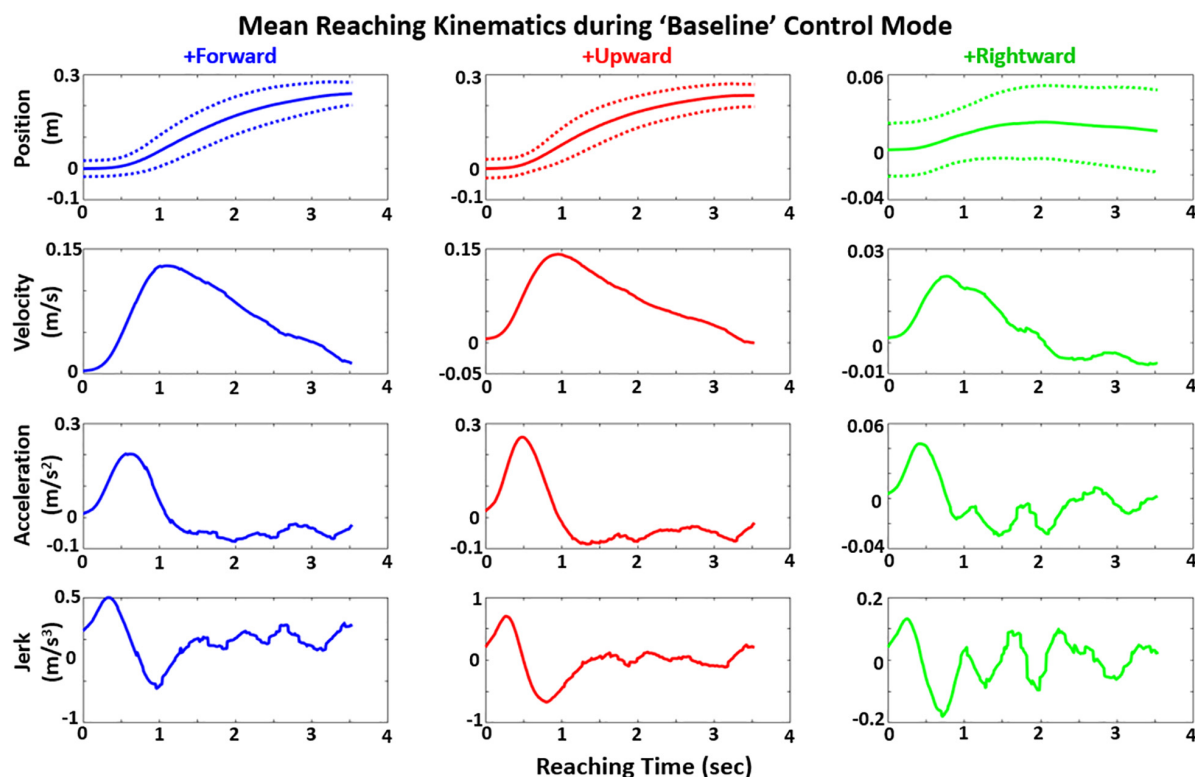
All post hoc comparisons made with Bonferroni correction. Significant post hoc  $p$ -values ( $< 0.05$ ) bolded.

## DISCUSSION

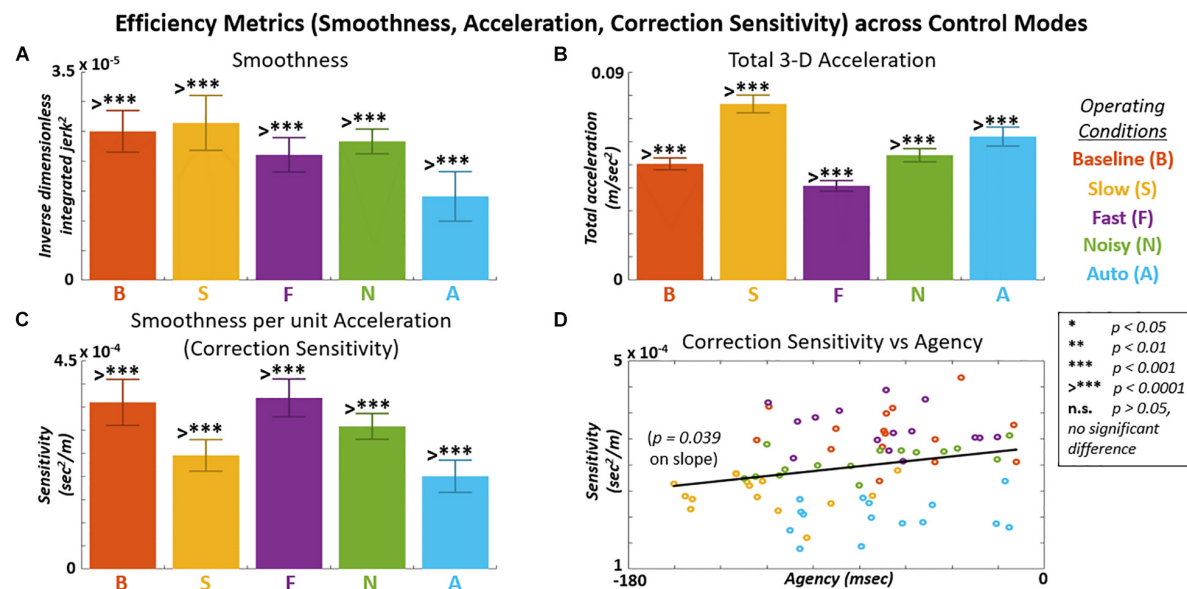
This study demonstrated a positive relationship between agency of grasp and performance of reach-to-grasp across various control modes of the virtual hand. Implicit agency was measured through intentional binding of grasp action, and performance was primarily assessed as inverse of mean reaching error to

a minimized path length trajectory. The results of this study may establish motivation for adapting user-device interfaces to co-maximize agency and performance. Of special interest are devices for movement assistance and rehabilitation, such as prostheses and exoskeletons. Clinical paradigms for motor rehabilitation that offer high value in both user engagement and functional utility have the best chances for retention and success





**FIGURE 8** | Mean position (with standard deviation dotted bands), velocity, acceleration, and jerk profiles for reaching hand shown in each movement direction for “Baseline” control mode across all subjects. For these mean trajectory plots, the trajectory of each subject was projected to fit across the average reach-to-contact time of 3.49 s for all subjects prior to trajectory averaging.



**FIGURE 9** | Various metrics of movement efficiency shown across control modes. **(A)** Smoothness values made dimensionless and inverted so higher values indicate greater smoothness. **(B)** Total acceleration in 3-D indicates magnitude of corrections made in tracking constant velocity target trajectory. **(C)** Smoothness per acceleration (correction sensitivity) computed to indicate smoothness achieved as a function of correction effort. **(D)** Correction sensitivity positively related to implicit agency across subjects and control modes (F-stat = 4.40,  $p = 0.039$ ).

**TABLE 3A |** Mean value comparisons for performance efficiency metrics across control modes.

Metric	Control mode					ANOVA		
	Baseline	Slow	Fast	Noisy	Auto	F-Stat	p-val	$\eta^2$
Smoothness (unitless, $10^{-5}$ )	$2.5 \pm 0.34$	$2.6 \pm 0.46$	$2.1 \pm 0.29$	$2.3 \pm 0.21$	$1.4 \pm 0.41$	27.8	7.6E-14	0.61
Total Acceleration ( $\text{cm/sec}^2$ )	$5.0 \pm 0.2$	$7.6 \pm 0.4$	$4.1 \pm 0.2$	$5.4 \pm 0.3$	$6.2 \pm 0.4$	260.3	3.2E-41	0.94
Correction Sensitivity ( $\text{sec}^2/\text{m}$ , $10^{-4}$ )	$3.6 \pm 0.5$	$2.5 \pm 0.3$	$3.7 \pm 0.4$	$3.1 \pm 0.3$	$2.0 \pm 0.3$	55.0	6.7E-21	0.76

**TABLE 3B |** Post hoc comparisons (p-value) between control modes for smoothness.

Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	–	0.81	<b>2.7E-02</b>	0.69	<b>9.9E-09</b>
Slow	–	–	<b>9.3E-04</b>	0.12	<b>9.9E-09</b>
Fast	–	–	–	0.42	<b>9.2E-06</b>
Noisy	–	–	–	–	<b>1.8E-08</b>

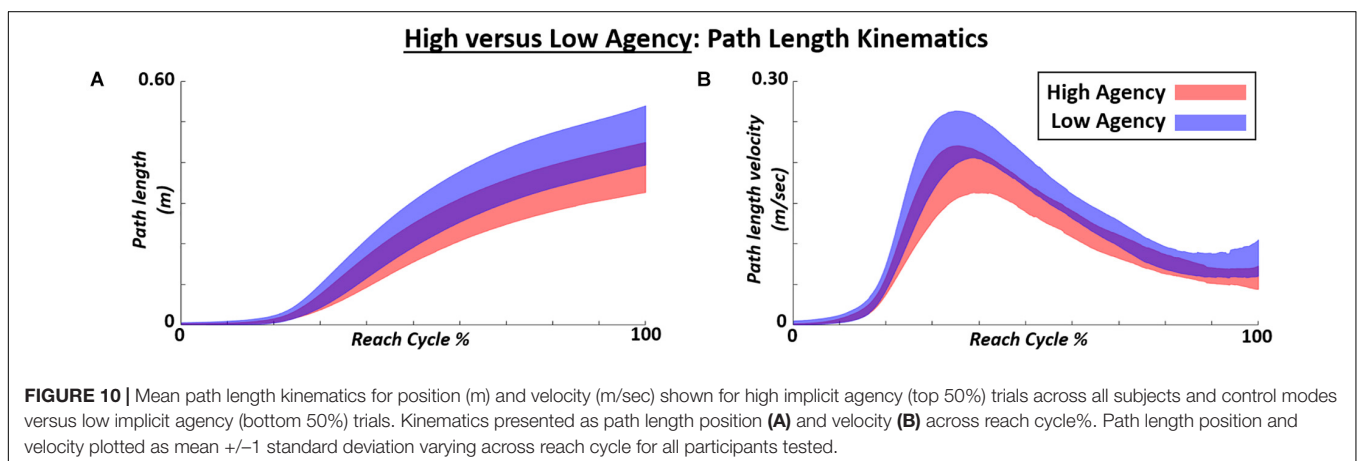
**TABLE 3C |** Post hoc comparisons (p-value) between control modes for total acceleration.

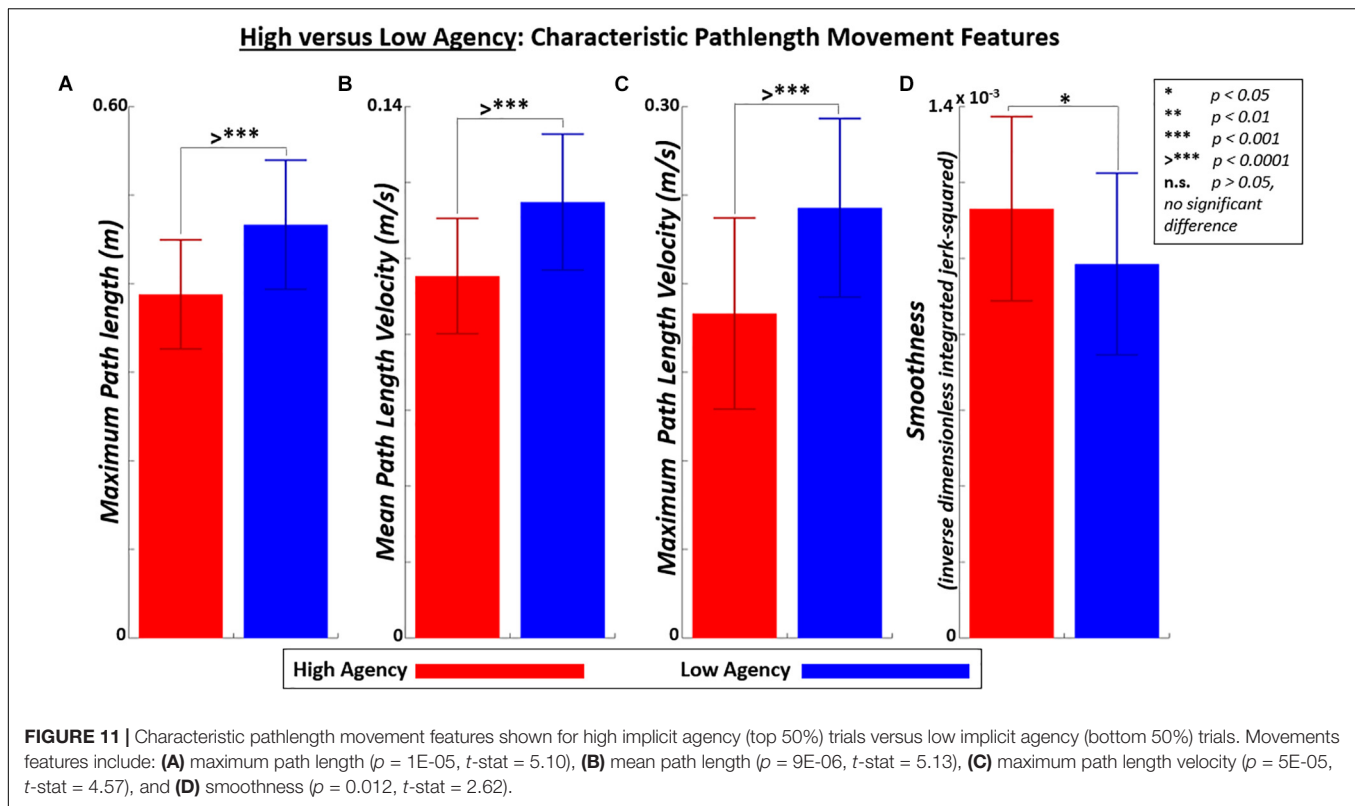
Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	–	<b>9.9E-09</b>	<b>1.0E-08</b>	<b>1.7E-02</b>	<b>9.9E-09</b>
Slow	–	–	<b>9.9E-09</b>	<b>9.9E-09</b>	<b>9.9E-09</b>
Fast	–	–	–	<b>9.9E-09</b>	<b>9.9E-09</b>
Noisy	–	–	–	–	<b>2.8E-08</b>

**TABLE 3D |** Post hoc comparisons (p-value) between control modes for correction efficiency (smoothness over acceleration).

Control mode	Control mode				
	Baseline	Slow	Fast	Noisy	Auto
Baseline	–	<b>9.9E-09</b>	0.96	<b>3.5E-03</b>	<b>9.9E-09</b>
Slow	–	–	<b>9.9E-09</b>	<b>2.5E-04</b>	<b>1.5E-02</b>
Fast	–	–	–	<b>3.4E-04</b>	<b>9.9E-09</b>
Noisy	–	–	–	–	<b>1.0E-08</b>

All post hoc comparisons made with Bonferonni correction. Significant post hoc p-values (<0.05) bolded.





(Wulf et al., 2010). To this end, the flexibility and accessibility of VR environments can be well leveraged to adapt rehabilitation platforms that co-maximize agency and movement performance. While implicating greater user agency over a device with higher functional performance is intuitive, the agency-performance link for movement has not been clearly established previously. This study relates agency, the autonomous sense of control, to functional performance across several control modes that can be standardly adapted for rehabilitation devices or rehabilitation training paradigms.

This study demonstrated a significant positive relationship ( $p < 0.001$ , **Figure 6**) between grasp agency and reaching performance across five distinct modes of control. This study also indicated how agency of grasp action is reduced in the presence of a preceding reach (**Figure 4**) compared to agency of movement initiation (Haggard et al., 2002). This is an important result since it suggests how agency of complex task action is modulated due to intermediate movement stages, which are further modified in this study with each control mode. The tested control modes were chosen to reflect control features (speed, noise mitigation, automaticity) commonly tuned for a movement device. The overall positive relationship between performance and agency is driven by the relatively high-agency, high-performing “Baseline” case and the relatively low-agency, low-performing “Slow” case. The “Fast” case yielded moderate-agency and moderate-performance. In total, these observations suggest that control sensitivity of speed may be a key tuning parameter for a device to co-maximize both user agency and functional performance.

Given high agency and performance for “Baseline,” it may be especially important to tune motion of a device to best match that of intact or restored proprioception and kinesthesia (Marasco et al., 2018). “Baseline” may also best facilitate the positive contributions of embodiment onto agency (Caspar et al., 2015). The “Slow” condition demonstrated the lowest agency, which indicates that experiencing slower device speeds relative to one’s intent significantly reduced the sense of control. For “Slow,” participants were required (unintentional) to move their own hands faster to compensate for the visual lags they observed for the virtual hand. While greater *intentional* effort can produce greater agency (Minohara et al., 2016), greater unintentional effort may reduce perceived efficacy of user control, especially if it promotes feelings of inability to initiate faster speeds (Kawabe, 2013). In this study, slower and faster speed control of the virtual hand required participants to actually reach longer and shorter, respectively. This limitation was required to employ changes in control speed while ensuring the task pathlength of the virtual hand was constant.

The remaining control mode cases were “Noisy” and “Auto,” which were categorically different from the other three which can be related by speed. These cases generally produced intermediate agency and performance relative to “Baseline” and “Slow.” “Noisy” may have been cognitively distracting in this study, but previous work has suggested that sensory noise (Collins et al., 2003) can improve motor function or indicate natural tremor (Allum, 1984; Riviere and Thakor, 1996) to better reflect human motion. For noise to be a cognitive or performance enhancer, there may be additional considerations

beyond the scope of this study such as identifying a custom resonant frequency for each person. “Auto” would expectantly reduce agency given its intended feature to remove control from the user. It has been shown that increased automation can reduce sense of agency during aircraft control (Berberian et al., 2012), and that intentional binding is sensitive to degrees of automaticity. Our study similarly uses intentional binding to indicate a reduction in agency with increased automation of a movement device. Bang-bang (abrupt switch between on-off states) control is an underlying principle in automating natural movement (Ben-Itzhak and Karniel, 2008) and powered movement assistance (Farris et al., 2013). To facilitate greater user agency over a rehabilitation device, the level of proportional control (Lenzi et al., 2012) must be optimized.

While agency is a measure of subjective perception, its implicit quantification through intentional binding and positive relationship to performance suggests its plausible incorporation in engineering better movement systems. For comparison to explicit agency (Moore et al., 2012), participants provided Likert-scale survey responses, but only after each trial block, as in Berberian et al. (2012). Since performance and implicit measures were taken after each trial, no conclusions between explicit agency and performance were made in this study. While implicit and explicit measures of agency may expectedly be related, they indicate agency at different levels. With implicit agency there is low-level and non-conceptual formation of being an agent, while explicit attribution of agency involves higher-order judgment (Moore et al., 2012). There has been compelling suggestion that there are separable implicit and explicit learning systems in dissociating their effects. Perruchet et al. (2006) demonstrated how, for a probabilistic learning task pairing two events, greater prediction strength was observed with implicit learning which relied more on recency effect. In our study, there appeared to be an inverse relationship between implicit and explicit measures of agency (Figure 7C), indicating separate levels of perceived learning. There also appears to be larger shifts toward implicit agency with the “Fast” case but toward explicit agency with “Slow.” This result suggests that perceptions of probabilistic learning and conscious judgment are sensitive to speed in this study and should be considered accordingly for potential device adaptation.

We next investigated metrics for movement control efficiency (increased smoothness, decreased acceleration, change in smoothness per change in acceleration) across control modes and their dependence on agency. Since agency is affected by perception of outcome and effort, efficiency of a movement device is implicated with agency and should be considered in optimizing user-device integration. Against the hypothesis that agency produces better movement characteristics, the “Slow” case, which demonstrated the lowest agency, also exhibited the highest smoothness (Figure 9A). Further inspection showed how this smoothness came at a cost of higher corrective accelerations to modulate biomechanical control (Winter, 2009), otherwise minimized for a constant velocity task. When smoothness is normalized

by total accelerations, a metric for correction sensitivity was inferred. For “Slow,” this sensitivity is significantly lower than “Baseline” or “Fast,” just as with agency and tracking performance. Across all five control modes, there is an apparent positive relationship ( $p < 0.05$  on linear regression slope) between correction sensitivity and agency (Figure 9D). Participants were not aware of considering any efficiency metrics, but control modes producing higher agency may produce performance benefits at multiple levels (execution and efficiency).

Finally, the effects of generally high (top 50%) agency were also observed on general path length kinematics (Figure 10) and specific pathlength characteristics (Figure 11). High agency generated reduced path length (primary performance task objective), reduced mean and peak path length velocity (closer to target constant velocity of  $0.08 \text{ m/s}^2$ ), and smoother (more efficient) movement. While this study primarily aimed to demonstrate performance and agency modulation across control modes as motivation for device adaptation, a general positive relationship between agency and movement performance was also apparent.

## CONCLUSION

In conclusion, this study has demonstrated clear dependence between implicit agency, based on time-interval estimation, and reaching performance across varying control modes. This dependence is apparent across conditions of speed changes, inclusion of noise, and adding a measure of automation. This suggests the potential for adapting control of devices, such as those for movement assistance, to co-maximize cognitive agency and performance. While performance indicates greater functional abilities, higher agency facilitates cognitive integration between user and device for ease-of-use and more natural control. Agency may also be key in accelerating learning and clinical retention of rehabilitation devices. Implicit measures of agency based on intentional binding are potentially reliable foundations for observing positive agency-performance dependencies.

This study was conducted in VR to ensure the pathlength for the reach-to-grasp task was visually similar while systematically varying control modes. It remains unclear how VR is best employed to identify optimal control modes for real-world devices. The objective with this work was to demonstrate the positive relationship between implicit agency and performance and their dependencies across control modes. This finding should then inspire practical methods that robustly and automatically adapt device control for each user toward greater agency and performance. Virtual reality could be a highly efficient medium in which to identify initial user-fitted control parameters. Those parameters may then be further refined based on real-world observations. Similar approaches have been utilized whereby computational models indicate basic operating characteristics of a control system (Nataraj et al., 2010, 2012c) prior to implementation in a clinical setting (Nataraj et al., 2012a,b).



In the future, alternative measures of agency such as neurophysiological recordings may be more robust for control system adaptation. Characterizing agency according to patterns in muscle electromyography (EMG) or brain electroencephalography (EEG) would be practically beneficial. These recordings often serve as command inputs to control systems for movement devices. Furthermore, neurophysiological recordings would not necessitate conscious user responses during adaptation of device control. Reducing such user onus could mitigate cognitive fatigue although that was not readily apparent in this study. Meanwhile, implicit agency through time-interval estimates could be critical in identifying what neurophysiological patterns best represent cognitive states of high agency. Changes in EEG readiness potential have been shown with greater agency in relation to the intent to initially move (Jo et al., 2014). However, spectral coherence changes in EEG and EMG during high agency movement remains unclear. Ultimately, clear biomarkers for high agency and performance would be invaluable in optimizing user-device interfaces for movement through better musculoskeletal control systems (Nataraj et al., 2010, 2012b).

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Stevens Institute of Technology IRB.

## REFERENCES

- Agostini, V., and Knaflitz, M. (2012). An algorithm for the estimation of the signal-to-noise ratio in surface myoelectric signals generated during cyclic movements. *IEEE Trans. Biomed. Eng.* 59, 219–225. doi: 10.1109/TBME.2011.2170687
- Allum, J. (1984). "Segmental reflex, muscle mechanical and central mechanisms underlying human physiological tremor," in *Movement Disorders: Tremor*, eds L. J. Findley and R. Capildeo (Berlin: Springer), 135–155. doi: 10.1007/978-1-349-06757-2\_8
- Bar-Cohen, Y. (2003). Haptic devices for virtual reality, telepresence, and human-assistive robotics. *Biol. Inspired Intell. Robots* 122:73.
- Ben-Itzhak, S., and Karniel, A. (2008). Minimum acceleration criterion with constraints implies bang-bang control as an underlying principle for optimal trajectories of arm reaching movements. *Neural Comput.* 20, 779–812. doi: 10.1162/neco.2007.12-05-077
- Berberian, B., and Sarrazin, J.-C., Le Blaye, P., and Haggard, P. (2012). Automation technology and sense of control: a window on human agency. *PLoS One* 7:e34075. doi: 10.1371/journal.pone.0034075
- Bhadra, N., Peckham, P. H., Keith, M. W., Kilgore, K. L., Montague, F., Gazdik, M., et al. (2002). Implementation of an implantable joint-angle transducer. *J. Rehabil. Res. Dev.* 39, 411–422.
- Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn. Sci.* 6, 237–242. doi: 10.1016/s1364-6613(02)01907-1
- Blaya, J. A., and Herr, H. (2004). Adaptive control of a variable-impedance ankle-foot orthosis to assist drop-foot gait. *IEEE Trans. Neural Syst. Rehabil. Eng.* 12, 24–31. doi: 10.1109/tnsre.2003.823266

The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

RN: designing and developing the experiment, analyzing the data, writing and revising the manuscript, and directing the project. SS: recruiting participants, performing the data collections, and revising the manuscript. AS: recruiting participants and performing the data collections. ML: revising the manuscript.

## FUNDING

This work was made possible by support from the Schaefer School of Engineering and Science, at the Stevens Institute of Technology and research grant (PC 53-19) from the New Jersey Health Foundation.

## ACKNOWLEDGMENTS

The authors would like to acknowledge Felix Chen, an electrical and computer engineering graduate student at Stevens, for initiating real-time marker streaming in the virtual reality protocol.

- Boostani, R., and Moradi, M. H. (2003). Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiol. Meas.* 24:309. doi: 10.1088/0967-3334/24/2/307
- Caldwell, D. G., Favede, C., and Tsagarakis, N. (1998). "Dextrous exploration of a virtual world for improved prototyping," in *1998 IEEE International Conference on Robotics and Automation (Cat. 8CH36146)*, (Piscataway, NJ: IEEE), 298–303.
- Caldwell, D. G., Kocak, O., and Andersen, U. (1995). "Multi-armed dextrous manipulator operation using glove/exoskeleton control and sensory feedback," in *1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, (Piscataway, NJ: IEEE), 567–572.
- Caspar, E. A., Cleeremans, A., and Haggard, P. (2015). The relationship between human agency and embodiment. *Conscious. Cogn.* 33, 226–236. doi: 10.1016/j.concog.2015.01.007
- Childress, D. S. (1973). Powered limb prostheses: their clinical significance. *IEEE Trans. Biomed. Eng.* BME-20, 200–207. doi: 10.1109/tbme.1973.324273
- Collins, J. J., Priplata, A. A., Gravelle, D. C., Niemi, J., Harry, J., and Lipsitz, L. A. (2003). Noise-enhanced human sensorimotor function. *IEEE Eng. Med. Biol. Mag.* 22, 76–83. doi: 10.1109/memb.2003.1195700
- Coyle, D., Moore, J., Kristensson Fletcher, O., and Blackwell, A. (2012). "I did that! Measuring users' experience of agency in their own actions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, eds J. A. Konstan, E. H. Chi, and K. Hook (New York, NY: ACM), 2025–2034.
- Davoodi, R., Urata, C., Hauschild, M., Khachani, M., and Loeb, G. E. (2007). Model-based development of neural prostheses for movement. *IEEE Trans. Biomed. Eng.* 54, 1909–1918. doi: 10.1109/tbme.2007.902252

- Desmurget, M., and Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends Cogn. Sci.* 4, 423–431. doi: 10.1016/s1364-6613(00)01537-0
- Dewey, J. A., and Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One* 9:e110118. doi: 10.1371/journal.pone.0110118
- Evans, N., Gale, S., Schurger, A., and Blanke, O. (2015). Visual feedback dominates the sense of agency for brain-machine actions. *PLoS One* 10:e0130019. doi: 10.1371/journal.pone.0130019
- Farris, R. J., Quintero, H. A., Murray, S. A., Ha, K. H., Hartigan, C., and Goldfarb, M. (2013). A preliminary assessment of legged mobility provided by a lower limb exoskeleton for persons with paraplegia. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 482–490. doi: 10.1109/TNSRE.2013.2268320
- Flash, T., and Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *J. Neurosci.* 5, 1688–1703. doi: 10.1523/jneurosci.05-07-01688.1985
- Frith, C. D., Blakemore, S. J., and Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 1771–1788. doi: 10.1098/rstb.2000.0734
- Frith, C. D., and Haggard, P. (2018). Volition and the brain – revisiting a classic experimental study. *Trends Neurosci.* 41, 405–407. doi: 10.1016/j.tins.2018.04.009
- Ghez, C., Gordon, J., and Ghilardi, M. F. (1995). Impairments of reaching movements in patients without proprioception. II. Effects of visual information on accuracy. *J. Neurophysiol.* 73, 361–372. doi: 10.1152/jn.1995.73.1.361
- Haggard, P. (2017). Sense of agency in the human brain. *Nat. Rev. Neurosci.* 18, 196–207. doi: 10.1038/nrn.2017.14
- Haggard, P. (2019). The neurocognitive bases of human volition. *Annu. Rev. Psychol.* 70, 9–28. doi: 10.1146/annurev-psych-010418-103348
- Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.* 5, 382–385. doi: 10.1038/nn827
- Hartigan, C., Kandilakis, C., Dalley, S., Clausen, M., Wilson, E., Morrison, S., et al. (2015). Mobility outcomes following five training sessions with a powered exoskeleton. *Top. Spinal Cord Inj. Rehabil.* 21, 93–99. doi: 10.1310/sci2102-93
- Heo, P., and Gu, G. M., Lee, S.-J., Rhee, K., and Kim, J. (2012). Current hand exoskeleton technologies for rehabilitation and assistive engineering. *Int. J. Precis. Eng. Manuf.* 13, 807–824.
- Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., et al. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485:372. doi: 10.1038/nature11076
- Hogan, N., and Sternad, D. (2009). Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J. Motor Behav.* 41, 529–534. doi: 10.3200/35-09-004-RC
- Hughes, A. M., Burrage, J. H., Demain, S. H., Ellis-Hill, C., Meagher, C., Tedesco-Triccas, L., et al. (2014). Translation of evidence-based assistive technologies into stroke rehabilitation: users' perceptions of the barriers and opportunities. *BMC Health Serv. Res.* 14:124. doi: 10.1186/1472-6963-14-124
- Jo, H.-G., Wittmann, M., Hinterberger, T., and Schmidt, S. (2014). The readiness potential reflects intentional binding. *Front. Hum. Neurosci.* 8:421. doi: 10.3389/fnhum.2014.00421
- Johannes, M. S., Bigelow, J. D., Burck, J. M., Harshbarger, S. D., Kozlowski, M. V., and Van Doren, T. (2011). An overview of the developmental process for the modular prosthetic limb. *Johns Hopkins APL Tech. Digest* 30, 207–216.
- Kawabe, T. (2013). Inferring sense of agency from the quantitative aspect of action outcome. *Conscious. Cogn.* 22, 407–412. doi: 10.1016/j.concog.2013.01.006
- Kotecha, A., Zhong, J., Stewart, D., and da Cruz, L. (2014). The argus II prosthesis facilitates reaching and grasping tasks: a case series. *BMC Ophthalmol.* 14:71. doi: 10.1186/1471-2415-14-71
- Kuiken, T. A., Li, G., Lock, B. A., Lipschutz, R. D., Miller, L. A., Stubblefield, K. A., et al. (2009). Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms. *JAMA* 301, 619–628. doi: 10.1001/jama.2009.116
- Lenzi, T., De Rossi, S. M. M., Vitiello, N., and Carrozza, M. C. (2012). Intention-based EMG control for powered exoskeletons. *IEEE Trans. Biomed. Eng.* 59, 2180–2190. doi: 10.1109/TBME.2012.2198821
- Li, G., Schultz, A. E., and Kuiken, T. A. (2010). Quantifying pattern recognition—Based myoelectric control of multifunctional transradial prostheses. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 185–192. doi: 10.1109/TNSRE.2009.2039619
- Limerick, H., Coyle, D., and Moore, J. W. (2014). The experience of agency in human-computer interactions: a review. *Front. Hum. Neurosci.* 8:643. doi: 10.3389/fnhum.2014.00643
- Lin, K.-C., Wu, C.-Y., Wei, T.-H., Gung, C., Lee, C.-Y., and Liu, J.-S. (2007). Effects of modified constraint-induced movement therapy on reach-to-grasp movements and functional performance after chronic stroke: a randomized controlled study. *Clin. Rehabil.* 21, 1075–1086. doi: 10.1177/0269215507079843
- Loureiro, R. C., and Harwin, W. S. (2007). “Reach & grasp therapy: design and control of a 9-DOF robotic neuro-rehabilitation system,” in *2007 IEEE 10th International Conference on Rehabilitation Robotics*, (Piscataway, NJ: IEEE), 757–763.
- Lucas, L., DiCicco, M., and Matsuoka, Y. (2004). An EMG-controlled hand exoskeleton for natural pinching. *J. Robot. Mechatron.* 16, 482–488. doi: 10.20965/jrm.2004.p0482
- Marasco, P. D., Hebert, J. S., Sensinger, J. W., Shell, C. E., Schofield, J. S., Thumser, Z. C., et al. (2018). Illusory movement perception improves motor control for prosthetic hands. *Sci. Transl. Med.* 10:eaa06990. doi: 10.1126/scitranslmed.aaa06990
- Minohara, R., Wen, W., Hamasaki, S., Maeda, T., Kato, M., Yamakawa, H., et al. (2016). Strength of intentional effort enhances the sense of agency. *Front. Psychol.* 7:1165. doi: 10.3389/fpsyg.2016.01165
- Moore, J., and Haggard, P. (2008). Awareness of action: inference and prediction. *Conscious. Cogn.* 17, 136–144. doi: 10.1016/j.concog.2006.12.004
- Moore, J. W. (2016). What is the sense of agency and why does it matter? *Front. Psychol.* 7:1272. doi: 10.3389/fpsyg.2016.01272
- Moore, J. W., and Fletcher, P. C. (2012). Sense of agency in health and disease: a review of cue integration approaches. *Conscious. Cogn.* 21, 59–68. doi: 10.1016/j.concog.2011.08.010
- Moore, J. W., Middleton Haggard, D., and Fletcher, P. C. (2012). Exploring implicit and explicit aspects of sense of agency. *Conscious. Cogn.* 21, 1748–1753. doi: 10.1016/j.concog.2012.10.005
- Moore, J. W., and Obhi, S. S. (2012). Intentional binding and the sense of agency: a review. *Conscious. Cogn.* 21, 546–561. doi: 10.1016/j.concog.2011.12.002
- Moore, J. W., Wegner, D. M., and Haggard, P. (2009). Modulating the sense of agency with external cues. *Conscious. Cogn.* 18, 1056–1064. doi: 10.1016/j.concog.2009.05.004
- Nataraj, R., Audu, M. L., Kirsch, R. F., and Triolo, R. J. (2010). Comprehensive joint feedback control for standing by functional neuromuscular stimulation—a simulation study. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18, 646–657. doi: 10.1109/TNSRE.2010.2083693
- Nataraj, R., Audu, M. L., and Triolo, R. J. (2012a). Center of mass acceleration feedback control of functional neuromuscular stimulation for standing in the presence of internal postural perturbations. *J. Rehabil. Res. Dev.* 49:889. doi: 10.1682/jrrd.2011.07.0127
- Nataraj, R., Audu, M. L., and Triolo, R. J. (2012b). Center of mass acceleration feedback control of standing balance by functional neuromuscular stimulation against external postural perturbations. *IEEE Trans. Biomed. Eng.* 60, 10–19. doi: 10.1109/tbme.2012.2218601
- Nataraj, R., Audu, M. L., and Triolo, R. J. (2012c). Comparing joint kinematics and center of mass acceleration as feedback for control of standing balance by functional neuromuscular stimulation. *J. Neuroeng. Rehabil.* 9:25. doi: 10.1186/1743-0003-9-25
- Nataraj, R., Evans, J., Seitz, W. H. Jr., and Li, Z.-M. (2014a). Effects of carpal tunnel syndrome on reach-to-pinch performance. *PLoS One* 9:e92063. doi: 10.1371/journal.pone.0092063
- Nataraj, R., Pasluosta, C., and Li, Z.-M. (2014b). Online kinematic regulation by visual feedback for grasp versus transport during reach-to-pinch. *Hum. Mov. Sci.* 36, 134–153. doi: 10.1016/j.humov.2014.05.007
- Nataraj, R., and van den Bogert, A. J. (2017). Simulation analysis of linear quadratic regulator control of sagittal-plane human walking—implications for exoskeletons. *J. Biomech. Eng.* 139:101009. doi: 10.1115/1.4037560
- Oishi, H., Tanaka, K., and Watanabe, K. (2018). Feedback of action outcome retrospectively influences sense of agency in a continuous action task. *PLoS One* 13:e0202690. doi: 10.1371/journal.pone.0202690

- Peckham, P. H., and Knutson, J. S. (2005). Functional electrical stimulation for neuromuscular applications. *Annu. Rev. Biomed. Eng.* 7, 327–360. doi: 10.1146/annurev.bioeng.6.040803.140103
- Perruchet, P., Cleeremans, A., and Destrebecqz, A. (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *J. Exp. Psychol. Learn. Mem. Cogn.* 32:955. doi: 10.1037/0278-7393.32.5.955
- Perry, J. C., Rosen, J., and Burns, S. (2007). Upper-limb powered exoskeleton design. *IEEE/ASME Trans. Mechatron.* 12, 408–417. doi: 10.1109/IEMBS.2011.6090139
- Phillips, B., and Zhao, H. (1993). Predictors of assistive technology abandonment. *Assist. Technol.* 5, 36–45. doi: 10.1080/10400435.1993.10132205
- Popovic, M. B. (2003). Control of neural prostheses for grasping and reaching. *Med. Eng. Phys.* 25, 41–50. doi: 10.1016/s1350-4533(02)00187-x
- Riviere, C. N., and Thakor, N. V. (1996). Modeling and canceling tremor in human-machine interfaces. *IEEE Eng. Med. Biol. Mag.* 15, 29–36. doi: 10.1109/51.499755
- Ronsse, R., Vitiello, N., Lenzi, T., Van Den Kieboom, J., Carrozza, M. C., and Ijspeert, A. J. (2010). “Adaptive oscillators with human-in-the-loop: proof of concept for assistance and rehabilitation,” in *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechanics*, (Piscataway, NJ: IEEE), 668–674.
- Ronsse, R., Vitiello, N., Lenzi, T., Van Den Kieboom, J., Carrozza, M. C., and Ijspeert, A. J. (2011). Human-robot synchrony: flexible assistance using adaptive oscillators. *IEEE Trans. Biomed. Eng.* 58, 1001–1012. doi: 10.1109/tbme.2010.2089629
- Rosen, J., Brand, M., Fuchs, M. B., and Arcan, M. (2001). A myosignal-based powered exoskeleton system. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 31, 210–222. doi: 10.1109/3468.925661
- Saleh, S., Fluet, G., Qiu, Q., Merians, A., Adamovich, S. V., and Tunik, E. (2017). Neural patterns of reorganization after intensive robot-assisted virtual reality therapy and repetitive task practice in patients with chronic stroke. *Front. Neurol.* 8:452. doi: 10.3389/fneur.2017.00452
- Saunders, J. A., and Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp. Brain Res.* 152, 341–352. doi: 10.1007/s00221-003-1525-2
- Shah, A., Sanford, S., Chen, Y., and Nataraj, R. (2018). “Role of cognitive agency in reach-to-grasp movement performance,” in *Paper Presented at the American Society Of Biomechanics*, Mayo Clinic, Rochester, MN.
- Shepherd, R. B. (2001). Exercise and training to optimize functional motor performance in stroke: driving neural reorganization? *Neural Plast.* 8, 121–129. doi: 10.1155/np.2001.121
- Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *J. Neuroeng. Rehabil.* 1:10.
- Taylor, D. M., Tillery, S. I. H., and Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science* 296, 1829–1832. doi: 10.1126/science.1070291
- Terenzi, S. (1998). Automatic tuning of myoelectric prostheses. *Development* 35, 294–304.
- Timmermans, A. A., Seelen, H. A., Willmann, R. D., and Kingma, H. (2009). Technology-assisted training of arm-hand skills in stroke: concepts on reacquisition of motor control and therapist guidelines for rehabilitation technology design. *J. Neuroeng. Rehabil.* 6:1. doi: 10.1186/1743-0003-6-1
- Todorov, E. (2004). Optimality principles in sensorimotor control. *Nat. Neurosci.* 7:907. doi: 10.1038/nn1309
- Todorov, E., Erez, T., and Tassa, Y. (2012). “Mujoco: a physics engine for model-based control,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Piscataway, NJ: IEEE), 5026–5033.
- van Vliet, P. M., and Sheridan, M. R. (2007). Coordination between reaching and grasping in patients with hemiparesis and healthy subjects. *Arch. Phys. Med. Rehabil.* 88, 1325–1331. doi: 10.1016/j.apmr.2007.06.769
- Wege, A., Kondak, K., and Hommel, G. (2005). “Mechanical design and motion control of a hand exoskeleton for rehabilitation,” in *IEEE International Conference Mechatronics and Automation, 2005*, (Piscataway, NJ: IEEE), 155–159.
- Wen, W., Yamashita, A., and Asama, H. (2015). The sense of agency during continuous action: performance is more important than action-feedback association. *PLoS One* 10:e0125226. doi: 10.1371/journal.pone.0125226
- Wing, S. A., Weber, D. J., and Santello, M. (2003). The role of vision on hand preshaping during reach to grasp. *Exp. Brain Res.* 152, 489–498. doi: 10.1007/s00221-003-1571-9
- Winter, D. A. (2009). *Biomechanics and Motor Control of Human Movement*. Hoboken, NJ: John Wiley & Sons.
- Wolpaw, J. R., and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17849–17854. doi: 10.1073/pnas.0403504101
- Wulf, G., Shea, C., and Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. *Med. Educ.* 44, 75–84. doi: 10.1111/j.1365-2923.2009.03421.x
- Yang, Y.-R., Wang, R.-Y., Lin, K.-H., Chu, M.-Y., and Chan, R.-C. (2006). Task-oriented progressive resistance strength training improves muscle strength and functional performance in individuals with stroke. *Clin. Rehabil.* 20, 860–870.

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

Copyright © 2020 Nataraj, Sanford, Shah and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Using More Ecological Paradigms to Investigate Working Memory: Strengths, Limitations and Recommendations

Lison Fanuel<sup>1,2</sup>, Gaën Plancher<sup>1\*†</sup> and Pascale Piolino<sup>3,4,5†</sup>

<sup>1</sup> Cognitive Mechanisms Research Laboratory, Université Lyon 2, Bron, France, <sup>2</sup> Lyon Neuroscience Research Center (CRNL), INSERM U1028, CNRS UMR5292, Université Lyon 1, Université de Lyon, Lyon, France, <sup>3</sup> Laboratoire Mémoire, Cerveau et Cognition, MC<sup>2</sup>Lab 7536, Université de Paris, Paris, France, <sup>4</sup> Institut de Psychologie, Université de Paris, Boulogne Billancourt, France, <sup>5</sup> Institut Universitaire de France, Paris, France

**Keywords:** memory, working memory, virtual reality, naturalistic events, ecological environment

## OPEN ACCESS

### Edited by:

Valerio Rizzo,  
University of Palermo, Italy

### Reviewed by:

José Manuel Reales,  
National University of Distance  
Education (UNED), Spain  
Fabio Solari,  
University of Genoa, Italy

### \*Correspondence:

Gaën Plancher  
gaen.plancher@univ-lyon2.fr

<sup>†</sup>These authors share last authorship

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 15 October 2019

**Accepted:** 06 April 2020

**Published:** 04 May 2020

### Citation:

Fanuel L, Plancher G and Piolino P  
(2020) Using More Ecological  
Paradigms to Investigate Working  
Memory: Strengths, Limitations  
and Recommendations.  
Front. Hum. Neurosci. 14:148.  
doi: 10.3389/fnhum.2020.00148

Working memory (WM) is essential to daily-life activities as it allows maintaining information in the short-term while processing concurrent information (Baddeley and Hitch, 1974). For example, one must maintain which ingredient is already in the plate while following a recipe. WM is a complex cognitive function involving multiple processes (e.g., encoding, maintenance, retrieval processes). The present paper focuses on the utility of virtual reality (VR) in investigating maintenance in WM, but the relevance of VR studies also applies to other WM-related mechanisms.

Recent models proposed an attention-based mechanism supporting maintenance of domain-general information: attentional refreshing (or refreshing; Camos et al., 2009; Camos and Barrouillet, 2014; Camos, 2017). Refreshing is described as a brief thought to an information that is no longer perceptually present (Johnson, 1992) and received growing attention in both WM and episodic memory (EM). The WM field provides convincing evidence of an involvement of refreshing in maintenance of visual, spatial, verbal information, as well as in the binding between these information (Hudjetz and Oberauer, 2007; Camos et al., 2009; Vergauwe et al., 2009, 2010, 2012). Studies using delayed recall suggest that memory performance depends on the time available for refreshing (Camos and Portrat, 2015; Souza and Oberauer, 2017; Jarjat et al., 2018) and that refreshing plays a role in construction of episodic traces (Johnson et al., 2002; Loaiza and McCabe, 2013). So far, studies used very simple to-be-remembered material such as letters or spatial locations (e.g., Camos et al., 2009; Vergauwe et al., 2009, 2010; Camos and Portrat, 2015). As refreshing is involved in maintenance of domain-general information and construction of EM, it should play a significant part in maintenance and long-term retention of rich and complex information. Because WM is central in daily-life activities, future research should design more ecological experiments to better understand the role of WM and refreshing in naturalistic situations.

VR seems to be a useful tool to investigate memory functioning in daily-life-like environments. VR allows creating naturalistic situation and increasing their ecological validity as compared to classical experimental or neuropsychological tests (Plancher and Piolino, 2017). Ecological validity refers to the extent to which experimental conditions are similar to a real-world setting (Bohil et al., 2011). Accordingly, a VR experience can provide complex and rich information involving multiple senses (vision, audition, proprioception, etc.) and spatiotemporal features. VR also enables interaction with the environment, for example by controlling displacements, which increase the feeling of immersion in this environment (Mestre and Fuchs, 2006). Besides improving ecological validity, controlled environments can be created to assess multiple features of memory traces—the content of the memory trace (what) and its spatial and temporal location (where and when)—as well as the binding between these features (Plancher et al., 2010). VR is thus a good compromise between memory assessment of daily-life-like experience and experimental control.



While VR was extensively used to better understand EM (Plancher and Piolino, 2017; La Corte et al., 2019), and executive functions (Negut, 2014; Neguț et al., 2016), only few studies investigated WM mechanisms with this method. Meilinger et al. (2008) investigated the involvement of WM in a wayfinding task. In comparison to a control condition without concurrent processing, a concurrent task (e.g., indicating the spatial location of a sound) negatively affected wayfinding of the routes previously seen. Both verbal and spatial concurrent tasks (continuously repeating a syllable sequence or tapping a spatial sequence, respectively), impaired memory performance for the landmark location and only the spatial concurrent task impaired memory performance for the route (Gras et al., 2013). More recently, Plancher et al. (2018) investigated the role of WM in construction of EM traces using a VR paradigm. While driving into a virtual town, participants had to memorize the encountered scenes as detailed as possible including the elements constituting the scene (*what*), the spatial location (*where*) and the temporal context of the scene (*when*). The recall of the spatial or temporal context associated to each element provided a binding score. As compared to a condition without concurrent processing, a verbal concurrent task (memorizing the number of garbage containers) only impaired memory performance of *what* information and a visuospatial concurrent task (memorizing the spatial position of containers) impaired memory performance of *what*, *when*, and *what-where-when* binding information. These results suggest that construction of memory traces rely on verbal and visuospatial maintenance mechanisms and were interpreted as reflecting an involvement of both phonological loop (i.e., verbal-specific WM mechanism, Baddeley and Hitch, 1974) and refreshing in the construction of *what* and an involvement of refreshing in the construction of *when* and binding components of EM traces.

Typically, the involvement of refreshing in maintenance is investigated using complex span tasks where to-be-processed items are interleaved in-between each to-be-memorized information (Barrouillet et al., 2004, 2007). Following the assumption that maintenance and processing compete for one limited resource (i.e., attention), increasing the amount of attentional resources required by the processing task leave less attentional resources available for attentional maintenance (i.e., refreshing, Barrouillet et al., 2007). Attentional sharing between maintenance and processing is proposed to rely on time: when time is occupied by a processing task, attentional maintenance cannot take place, and vice versa (Barrouillet et al., 2007). Varying the amount of time required for processing a concurrent task (i.e., its cognitive load) results in manipulating the amount of time available for refreshing. Poorer WM performance under higher cognitive load is taken as evidence of an involvement of refreshing in WM maintenance (Barrouillet et al., 2004, 2007; Vergauwe et al., 2009, 2010).

To understand the involvement of refreshing in maintenance of rich and complex information, we suggest adapting complex span tasks to VR paradigms. The task concurrent to maintenance should be distinct from the memorization task and allow to measure response times. Thereby, it will be possible to manipulate the cognitive load of the concurrent task

and investigate whether and how refreshing is involved in maintenance of the different features of a complex memory trace (*what*, *when*, *where*) and the binding of these features. A passive exploration of the environment would allow determining the time-course of the task and controlling temporal parameters and require no control nor planning for traveling. Motor and planning actions required by active traveling can constitute an attentional cost (Plancher et al., 2013) and have an uncontrolled detrimental effect on WM performance. However, a passive exploration result in a simple video experience. Active navigation seems more useful to enrich the EM trace (Plancher et al., 2012, 2013; Sauzéon et al., 2012; Jebara et al., 2014). Immersion and real time interaction with the environment are necessary for self-experience and bodily representation and modulate the sense of presence in the present that is central in daily-life experience (Nash et al., 2000; Makowski et al., 2017). Self-experience and bodily representation reinforce EM performance (Bergouignan et al., 2014; Repetto et al., 2016; Tuena et al., 2017, 2019; Blanke et al., 2018) and might also influence maintenance in WM. For a more immersive experience and a better understanding of the involvement of refreshing in daily-life situations, future studies should systematically use an active condition. Contrasting different levels of immersion (from computer screens to head-mounted displays or cave automatic virtual environments) and interaction (from joystick to motion capture) with a passive condition would enable determining the minimum conditions for studying WM in an ecological context and explore how embodiment impacts refreshing.

To study WM in conditions as close as possible from real-life using VR, we suggest designing a virtual environment where the participant is freely exploring and encounter events of different nature (e.g., visual, auditory, proprioceptive, spatial, or any combination). To enhance ecological validity, events should occur at a non-isochronous pace. Temporal parameters related to to-be-memorized and to-be-processed events (number of events, presentation duration, inter-stimuli intervals) should be fixed and participants' behavior (e.g., response times) should be measured. Studies of interdependence between WM processes and other cognitive functions will also benefit from VR paradigms. Long-term memory and semantic representation seem to contribute to refreshing and WM (Loaiza et al., 2015; Loaiza and Camos, 2018). Refreshing is involved in construction of EM traces (Johnson et al., 2002; Loaiza and McCabe, 2013) and might play a part in prospective memory (Marsh and Hicks, 1998). Carefully timed VR experiments manipulating WM parameters will contribute to a detailed insight of these cognitive processes and their relationship with WM.

Developing virtual-reality-based paradigms to investigate WM would be useful to identify the neural basis of WM mechanisms in ecological situations. To date, as for behavioral studies, neurophysiological studies of maintenance in WM (Vogel and Machizawa, 2004; Guimond et al., 2011; Lefebvre et al., 2013; Grimault et al., 2014) and refreshing have used very simple stimuli (Johnson et al., 2005, 2015). To our knowledge, no study combined WM daily-life-like paradigms and neurophysiological measures. Yet, VR paradigms can be combined easily with neurophysiological recording like

eye-tracking (e.g., Whitmire et al., 2016) or electrodermal and cardiac responses (e.g., Parsons et al., 2011; Armougum et al., 2019). Electroencephalography (e.g., Jaiswal et al., 2010; Bohil et al., 2011) and fMRI (e.g., Kalpouzos et al., 2010) recordings are possible during exposure in experiments with limited movements. Neurophysiological measures during retrieval depending on WM manipulation during exposure (e.g., the amount of time available for refreshing) could provide a further insight on the implication of WM processes in episodic construction.

VR is a powerful experimental tool that allow creating multimodal and naturalistic environments to assess memory for rich and complex information while keeping a strong experimental control. VR will allow testing theoretical assumptions with enhanced ecological validity and interactive fidelity and explore new hypothesis as whether and how traveling affects WM. In addition, VR allows us to examine some of the basic properties of the situated and embodied approach through exact control of methodological factors and go further into the understanding of the role of presence and consciousness in memory. Yet, using VR paradigms in WM studies require to compromise between a strong control of temporal parameters and immersion through active exploring. We recommend fixing temporal parameters of the experiment and to measure behavior, especially response times, as precisely as possible. Future studies will need to determine the minimum immersion and interaction conditions to explore WM with VR to limit the attentional cost of active navigation and facilitate the combined use of VR and neurophysiological measurements. Moreover, some other technical aspects of VR such as the risk of cybersickness, photorealistic level of environment, type of interaction and way of navigation, level and mode of the embodiment need further investigations to test their impact on WM as suggested by some authors in the domain of memory studies (Smith, 2019).

Future studies should also assess metric properties of VR-based measurements of WM to ensure a good construct validity and reliability. Besides the good equivalence between cognitive performance and physiological responses in the virtual and the real world (Sorita et al., 2013; Armougum et al., 2019),

construct validity of VR-based neuropsychological assessments seem suitable (Neguț et al., 2015), comforting the feasibility of developing VR-based assessment of cognitive functions. Given that VR is becoming more accessible at low cost, future studies will be able to multiply on a larger number of subjects to obtain reliable and reproducible data.

WM is impaired in various populations such as healthy aging, age-related dementia (e.g., Baddeley et al., 1986; Huntley and Howard, 2010) or schizophrenia (e.g., Lee and Park, 2005). In some populations, WM deficit is proposed to be due to an impairment of refreshing (e.g., Hoareau et al., 2016; Fanuel et al., 2018 in healthy aging; Grillon et al., 2013 in schizophrenia). Developing naturalistic tools to investigate WM functioning seem very useful to characterize the WM deficits in these populations. Previous studies suggest that WM training involving multi-modal stimuli, demanding high cognitive engagement and targeting WM domain-general mechanisms are more likely to yield WM and general cognitive enhancement (Morrison and Chein, 2011). VR-based trainings of WM thus seem a promising approach for enhancing both WM and broader cognitive functions. It is a crucial point nowadays to examine the VR acceptability for fragile populations.

## AUTHOR CONTRIBUTIONS

All authors conceptualized the manuscript. LF wrote the original draft. GP and PP reviewed and edited its intellectual content.

## FUNDING

This work was supported by a grant from Région Rhône-Alpes. It was conducted within the framework of the LabEx Cortex (Construction, Function and Cognitive Function and Rehabilitation of the Cortex, ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

## REFERENCES

- Armougum, A., Orriols, E., Gaston-Bellegarde, A., Joie-La Marle, C., and Piolino, P. (2019). Virtual reality: a new method to investigate cognitive load during navigation. *J. Environ. Psychol.* 65:101338. doi: 10.1016/j.jenvp.2019.101338
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1
- Baddeley, A. D., Logie, R., Bressi, S., Sala, S. D., and Spinnler, H. (1986). Dementia and working memory. *Q. J. Exp. Psychol. Sec. A* 38, 603–618. doi: 10.1080/14640748608401616
- Barrouillet, P., Bernardin, S., and Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *J. Exp. Psychol. Gen.* 133, 83–100. doi: 10.1037/0096-3445.133.1.83
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., and Camos, V. (2007). Time and cognitive load in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 570–585. doi: 10.1037/0278-7393.33.3.570
- Bergouignan, L., Nyberg, L., and Ehrsson, H. H. (2014). Out-of-body-induced hippocampal amnesia. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4421–4426. doi: 10.1073/pnas.1318801111
- Bréchet, L., Mange, R., Herbelin, B., Gauthier, B., Serino, A., and Blanke, O. (2018). Viewing one's body during encoding boosts episodic memory. *bioRxiv [Preprint]*, bioRxiv:318956. doi: 10.1101/318956
- Bohil, C. J., Alicea, B., and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 753–762. doi: 10.1038/nrn3122
- Camos, V. (2017). Domain-specific versus domain-general maintenance in working memory: reconciliation within the time-based resource sharing model. *Psychol. Learn. Motiv.* 67, 135–171. doi: 10.1016/bs.plm.2017.03.005
- Camos, V., and Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: the executive and phonological loops. *Front. Hum. Neurosci.* 8:900. doi: 10.3389/fnhum.2014.00900
- Camos, V., Lagner, P., and Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *J. Mem. Lang.* 61, 457–469. doi: 10.1016/j.jml.2009.06.002

- Camos, V., and Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychon. Bull. Rev.* 22, 1029–1034. doi: 10.3758/s13423-014-0772-5
- Fanuel, L., Plancher, G., Monsaingeon, N., Tillmann, B., and Portrat, S. (2018). Temporal dynamics of maintenance in young and old adults. *Ann. N. Y. Acad. Sci.* 1424, 137–148. doi: 10.1111/nyas.13640
- Gras, D., Gyselinck, V., Perrusset, M., Orriols, E., and Piolino, P. (2013). The role of working memory components and visuospatial abilities in route learning within a virtual environment. *J. Cogn. Psychol.* 25, 38–50. doi: 10.1080/20445911.2012.739154
- Grillon, M.-L., Oppenheim, C., Varoquaux, G., Charbonneau, F., Devauchelle, A.-D., Krebs, M.-O., et al. (2013). Hyperfrontality and hypoconnectivity during refreshing in schizophrenia. *Psychiatry Res.* 211, 226–233. doi: 10.1016/j.psychres.2012.09.001
- Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., et al. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: evidence from magnetoencephalography. *NeuroImage* 94, 96–106. doi: 10.1016/j.neuroimage.2014.03.020
- Guimond, S., Vachon, F., Nolden, S., Lefebvre, C., Grimault, S., and Jolicoeur, P. (2011). Electrophysiological correlates of the maintenance of the representation of pitch objects in acoustic short-term memory: Pitch and ASTM. *Psychophysiology* 48, 1500–1509. doi: 10.1111/j.1469-8986.2011.01234.x
- Hoareau, V., Lemaire, B., Portrat, S., and Plancher, G. (2016). Reconciling two computational models of working memory in aging. *Top. Cogn. Sci.* 8, 264–278. doi: 10.1111/tops.12184
- Hudjetz, A., and Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: testing four models of the complex span paradigm. *Mem. Cogn.* 35, 1675–1684. doi: 10.3758/BF03193501
- Huntley, J. D., and Howard, R. J. (2010). Working memory in early Alzheimer's disease: a neuropsychological review. *Int. J. Geriatr. Psychiatr.* 25, 121–132. doi: 10.1002/gps.2314
- Jaiswal, N., Ray, W., and Slobounov, S. (2010). Encoding of visual-spatial information in working memory requires more cerebral efforts than retrieval: evidence from an EEG and virtual reality study. *Brain Res.* 1347, 80–89. doi: 10.1016/j.brainres.2010.05.086
- Jarjat, G., Hoareau, V., Plancher, G., Hot, P., Lemaire, B., and Portrat, S. (2018). What makes working memory traces stable over time?: working memory conditions for long-term recall. *Ann. N. Y. Acad. Sci.* 1424, 149–160. doi: 10.1111/nyas.13668
- Jebara, N., Orriols, E., Zaoui, M., Berthoz, A., and Piolino, P. (2014). Effects of enactment in episodic memory: a pilot virtual reality study with young and elderly adults. *Front. Aging Neurosci.* 6:338. doi: 10.3389/fnagi.2014.00338
- Johnson, M. K. (1992). MEM: mechanisms of recollection. *J. Cogn. Neurosci.* 4, 268–280. doi: 10.1162/jocn.1992.4.3.268
- Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., and Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: prefrontal correlates of refreshing a just-activated representation. *Cogn. Affect. Behav. Neurosci.* 5, 339–361. doi: 10.3758/CABN.5.3.339
- Johnson, M. K., Reeder, J. A., Raye, C. L., and Mitchell, K. J. (2002). Second thoughts versus second looks: an age-related deficit in reflectively refreshing just-activated information. *Psychol. Sci.* 13, 64–67. doi: 10.1111/1467-9280.00411
- Johnson, M. R., McCarthy, G., Muller, K. A., Brudner, S. N., and Johnson, M. K. (2015). Electrophysiological correlates of refreshing: event-related potentials associated with directing reflective attention to face, scene, or word representations. *J. Cogn. Neurosci.* 27, 1823–1839. doi: 10.1162/jocn\_a\_00823
- Kalpourzou, G., Eriksson, J., Sjölie, D., Molin, J., and Nyberg, L. (2010). Neurocognitive systems related to real-world prospective memory. *PLoS ONE* 5:e13304. doi: 10.1371/journal.pone.0013304
- La Corte, V., Sperduti, M., Abichou, K., and Piolino, P. (2019). Episodic memory assessment and remediation in normal and pathological aging using virtual reality: a mini review. *Front. Psychol.* 10:173. doi: 10.3389/fpsyg.2019.00173
- Lee, J., and Park, S. (2005). Working memory impairments in schizophrenia: a meta-analysis. *J. Abnorm. Psychol.* 114, 599–611. doi: 10.1037/0021-843X.114.4.599
- Lefebvre, C., Vachon, F., Grimault, S., Thibault, J., Guimond, S., Peretz, I., et al. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. *Neuropsychologia* 51, 2939–2952. doi: 10.1016/j.neuropsychologia.2013.08.003
- Loaiza, V. M., and Camos, V. (2018). The role of semantic representations in verbal working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 863–881. doi: 10.1037/xlm0000475
- Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., and McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychon. Bull. Rev.* 22, 274–280. doi: 10.3758/s13423-014-0673-7
- Loaiza, V. M., and McCabe, D. P. (2013). The influence of aging on attentional refreshing and articulatory rehearsal during working memory on later episodic memory performance. *Aging Neuropsychol. Cogn.* 20, 471–493. doi: 10.1080/13825585.2012.738289
- Makowski, D., Sperduti, M., Nicolas, S., and Piolino, P. (2017). “Being there” and remembering it: presence improves memory encoding. *Consci. Cogn.* 53, 194–202. doi: 10.1016/j.concog.2017.06.015
- Marsh, R. L., and Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 336–348. doi: 10.1037/0278-7393.24.2.336
- Meilinger, T., Knauff, M., and Bühlhoff, H. H. (2008). Working memory in wayfinding—a dual task experiment in a virtual city. *Cogn. Sci.* 32, 755–770. doi: 10.1080/03640210802067004
- Mestre, D., and Fuchs, P. (2006). “Immersion et présence,” in *Le traité de la réalité virtuelle*. (pp. 309–338). Available online at: [http://www.ism.univ-amu.fr/mestre/projects/virtual%20reality/Pres\\_2005.pdf](http://www.ism.univ-amu.fr/mestre/projects/virtual%20reality/Pres_2005.pdf)
- Morrison, A. B., and Chein, J. M. (2011). Does working memory training work? the promise and challenges of enhancing cognition by training working memory. *Psychon. Bull. Rev.* 18, 46–60. doi: 10.3758/s13423-010-0034-0
- Nash, E. B., Edwards, G. W., Thompson, J. A., and Barfield, W. (2000). A review of presence and performance in virtual environments. *Int. J. Hum. Comput. Interact.* 12, 1–41. doi: 10.1207/S15327590IJHC1201\_1
- Negut, A., Matu, S.-A., Sava, F. A., and David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. *Clin. Neuropsychol.* 30, 165–184. doi: 10.1080/13854046.2016.1144793
- Negut, A. (2014). Cognitive assessment and rehabilitation in virtual reality: theoretical review and practical implications. *Roman. J. Appl. Psychol.* 16, 1–7. doi: 10.1155/2014/594540
- Negut, A., Matu, S.-A., Sava, F. A., and David, D. (2015). Convergent validity of virtual reality neurocognitive assessment: a meta-analytic approach. *Transylv. J. Psychol.* 16, 31–54.
- Parsons, T. D., Courtney, C. G., Arizmendi, B., and Dawson, M. (2011). *Virtual Reality Stroop Task for Neurocognitive Assessment*. Newport Beach, CA: Medicine Meets Virtual Reality.
- Plancher, G., Barra, J., Orriols, E., and Piolino, P. (2013). The influence of action on episodic memory: a virtual reality study. *Q. J. Exp. Psychol.* 66, 895–909. doi: 10.1080/17470218.2012.722657
- Plancher, G., Gyselinck, V., Nicolas, S., and Piolino, P. (2010). Age effect on components of episodic memory and feature binding: a virtual reality study. *Neuropsychologia* 24, 379–390. doi: 10.1037/a0018680
- Plancher, G., Gyselinck, V., and Piolino, P. (2018). The integration of realistic episodic memories relies on different working memory processes: evidence from virtual navigation. *Front. Psychol.* 9:47. doi: 10.3389/fpsyg.2018.00047
- Plancher, G., Tirard, A., Gyselinck, V., Nicolas, S., and Piolino, P. (2012). Using virtual reality to characterize episodic memory profiles in amnesic mild cognitive impairment and Alzheimer's disease: influence of active and passive encoding. *Neuropsychologia* 50, 592–602. doi: 10.1016/j.neuropsychologia.2011.12.013
- Plancher, G., and Piolino, P. (2017). “Virtual reality for assessment of episodic memory in normal and pathological aging,” in *The Role of Technology in Clinical Neuropsychology*, eds T. Parsons and R. Kane (Oxford: Oxford University Press).
- Repetto, C., Serino, S., Macedonia, M., and Riva, G. (2016). Virtual reality as an embodied tool to enhance episodic memory in elderly. *Front. Psychol.* 7:1839. doi: 10.3389/fpsyg.2016.01839
- Sauzéon, H., Arvind Pala, P., Larrue, F., Wallet, G., Déjos, M., Zheng, X., et al. (2012). The use of virtual reality for episodic memory assessment: effects of active navigation. *Exp. Psychol.* 59, 99–108. doi: 10.1027/1618-3169/a000131

- Smith, S. A. (2019). Virtual reality in episodic memory research: a review. *Psychon. Bull. Rev.* 26, 1213–1237. doi: 10.3758/s13423-019-01605-w
- Sorita, E., N'Kaoua, B., Larrue, F., Criquillon, J., Simion, A., Sauzéon, H., et al. (2013). Do patients with traumatic brain injury learn a route in the same way in real and virtual environments? *Disabil. Rehabil.* 35, 1371–1379. doi: 10.3109/09638288.2012.738761
- Souza, A. S., and Oberauer, K. (2017). Time to process information in working memory improves episodic memory. *J. Mem. Lang.* 96, 155–167. doi: 10.1016/j.jml.2017.07.002
- Tuena, C., Serino, S., Dutriaux, L., Riva, G., and Piolino, P. (2019). Virtual enactment effect on memory in young and aged populations: a systematic review. *J. Clin. Med.* 8:620. doi: 10.3390/jcm8050620
- Tuena, C., Serino, S., Gaston-Bellegarde, A., Orriols, E., Makowski, D., Riva, G., et al. (2017). How virtual embodiment affects episodic memory functioning: a proof-of-concept study. *Ann. Rev. Cyberther. Telemed.* 15, 98–103.
- Vergauwe, E., Barrouillet, P., and Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: a time-based resource-sharing account. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1012–1028. doi: 10.1037/a0015859
- Vergauwe, E., Barrouillet, P., and Camos, V. (2010). Do mental processes share a domain-general resource? *Psychol. Sci.* 21, 384–390. doi: 10.1177/0956797610361340
- Vergauwe, E., Dewaele, N., Langerock, N., and Barrouillet, P. (2012). Evidence for a central pool of general resources in working memory. *J. Cogn. Psychol.* 24, 359–366. doi: 10.1080/20445911.2011.640625
- Vogel, E. K., and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751. doi: 10.1038/nature02447
- Whitmire, E., Trutoiu, L., Cavin, R., Perek, D., Scally, B., Phillips, J., et al. (2016). “EyeContact: scleral coil eye tracking for virtual reality,” in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, 184–191. doi: 10.1145/2971763.2971771

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

Copyright © 2020 Fanuel, Plancher and Piolino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality

Silvia Cavedoni<sup>1\*</sup>, Alice Chirico<sup>2</sup>, Elisa Pedrolì<sup>1,3</sup>, Pietro Cipresso<sup>1,2</sup> and Giuseppe Riva<sup>1,2</sup>

<sup>1</sup> Applied Technology for Neuro-Psychology Lab, Istituto Auxologico Italiano, Milan, Italy, <sup>2</sup> Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy, <sup>3</sup> Faculty of Psychology, eCampus University, Novedrate, Italy

## OPEN ACCESS

### Edited by:

Soledad Ballesteros,  
National University of Distance  
Education (UNED), Spain

### Reviewed by:

Vitoantonio Bevilacqua,  
Politecnico di Bari, Italy  
Junichi Chikazoe,  
National Institute for Physiological  
Sciences, Japan

### \*Correspondence:

Silvia Cavedoni  
silvia.cavedoni@outlook.com

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 14 April 2020

**Accepted:** 02 June 2020

**Published:** 24 July 2020

### Citation:

Cavedoni S, Chirico A, Pedrolì E,  
Cipresso P and Riva G (2020) Digital  
Biomarkers for the Early Detection  
of Mild Cognitive Impairment: Artificial  
Intelligence Meets Virtual Reality.  
Front. Hum. Neurosci. 14:245.  
doi: 10.3389/fnhum.2020.00245

Elderly people affected by Mild Cognitive Impairment (MCI) usually report a perceived decline in cognitive functions that deeply impacts their quality of life. This subtle waning, although it cannot be diagnosable as dementia, is noted by caregivers on the basis of their relative's behaviors. Crucially, if this condition is also not detected in time by clinicians, it can easily turn into dementia. Thus, early detection of MCI is strongly needed. Classical neuropsychological measures – underlying a categorical model of diagnosis – could be integrated with a dimensional assessment approach involving Virtual Reality (VR) and Artificial Intelligence (AI). VR can be used to create highly ecologically controlled simulations resembling the daily life contexts in which patients' daily instrumental activities (IADL) usually take place. Clinicians can record patients' kinematics, particularly gait, while performing IADL (Digital Biomarkers). Then, Artificial Intelligence employs Machine Learning (ML) to analyze them in combination with clinical and neuropsychological data. This integrated computational approach would enable the creation of a predictive model to identify specific patterns of cognitive and motor impairment in MCI. Therefore, this new dimensional cognitive-behavioral assessment would reveal elderly people's neural alterations and impaired cognitive functions, typical of MCI and dementia, even in early stages for more time-sensitive interventions.

**Keywords:** gait analysis, kinematic, Mild Cognitive Impairment, Virtual Reality, Machine Learning, elderly, digital biomarkers, Artificial Intelligence

## INTRODUCTION

A categorical approach to diagnosing dementia struggles to capture subclinical conditions, such as Mild Cognitive Impairment (MCI). Crucially, MCI can either revert to normal cognition, stabilize, or slowly evolve toward other forms of dementia (Chiu, 2005; Walters, 2011; Morris, 2012; Díaz-Mardomingo et al., 2017; Vanacore et al., 2017). This construct indicates people affected by an in-between condition between normal aging and early dementia (Petersen, 2004; Albert et al., 2011; Mckhann et al., 2011; Seo et al., 2017) and is usually segmented into single- or multiple-domain amnesic (aMCI) and non-amnesic (naMCI) subtypes, depending on whether impairments concern only memory or other cognitive functions, e.g., executive and visuo-spatial

abilities (Petersen, 2004; Apostolova and Cummings, 2008; Albert et al., 2011; Hughes et al., 2011; Michaud et al., 2017; Facal et al., 2019). Both patients and their caregivers can observe and report clear signals of this subtle waning, undiagnosable as dementia. Frequently, elderly people express concern over their perceived worsening in one or more cognitive domains, such as memory or language (Petersen et al., 1999, 2018). This waning has a great impact on their quality of life, reducing their ability to autonomously carry out activities. A key aspect concerns the possibility of detecting an initial cognitive decline at the behavioral level with a slowdown in execution of the instrumental activities of daily life (IADL), such as grocery shopping and medication and financial management (Kim and Kim, 2009; Gold and Gold, 2012).

Changes associated with subclinical forms of dementia manifest themselves through behavioral alterations. Usually, caregivers are the first ones to notice these altered behaviors, as shown by Van Vliet et al. (2011). The authors explored the barriers hindering a timely diagnosis of dementia, focusing on interviews conducted with caregivers of relatives that were later diagnosed with early-onset dementia (EOD). Caregivers frequently reported behavioral changes in relatives with EOD, either alone or associated with neuropsychiatric symptoms (NPS), such as apathy or depression, and personality changes. Behavioral impairment then evolved toward a decline in IADL and involved cognitive impairment, particularly memory loss (Van Vliet et al., 2011). The broader detrimental impact of behavioral changes generated familial/marital conflicts and reduced job productivity, leading to a decreased income or even dismissal (Van Vliet et al., 2011). Though valuable, this anecdotal information rarely becomes part of a (categorical) diagnosis based on medical and neuropsychological assessment. Over time, caregivers have been considered a source of information that is not always reliable, given their tendency to over- or underestimate elderly people's deficits, possibly due to knowledge gaps (Akl et al., 2015; Jekel et al., 2015). Caregivers might be absent or suffer from physical or psychological conditions exacerbated by their relative's worsening (Okonkwo et al., 2008; Van Vliet et al., 2011; Pfeifer et al., 2013; Akl et al., 2015; Jekel et al., 2015). They might explain the elderly person's decline and behavioral, cognitive, and personality changes rather as a result of aging. Sometimes, caregivers are not aware of the symptoms because of their relative's ability to cover them up, denying their impairments or developing subsequent compensatory strategies to disguise the difficulties. This, in turn, delays the consultation of a practitioner and the diagnostic process as well (Okonkwo et al., 2008; Van Vliet et al., 2011; Roehr et al., 2019). Early detection of MCI, resulting in time-sensitive interventions, is still an open issue.

In this regard, two components appear relevant. Firstly, there is a need to rely more on rigorous and systematic behavioral analysis for early detection of MCI. Secondly, there is a need to integrate this new practice into current ones, i.e., neuropsychological evaluation. Including these data jointly in MCI assessment can allow a more sensitive measurement of the deficit by placing it on a continuum, reflecting a dimensional approach accounting for several other subclinical conditions,

including Subjective Cognitive Decline (SCD; Roehr et al., 2019) or Pre-Mild Cognitive Impairment (Pre-MCI; Crocco et al., 2018; Grassi et al., 2018).

This is far more crucial when considering that MCI can turn into dementia if the elderly person does not receive a timely diagnosis (Chiu, 2005), which should be built upon finer discrimination among the early stages of MCI and the collection of behavioral data, moving beyond a categorical, dichotomous approach rooted in previous diagnostic models, such as DSM-IV-TR or ICD-10 (American Psychiatric Association, 2000; World Health Organization, 2007; Negu et al., 2016), and the distinction between aMCI and naMCI.

The exclusive implementation of neuropsychological assessment tools cannot provide information on the finer behavioral aspects of the early stages of MCI and, despite their widespread use and efficacy, they fail to predict an individual's behavior in real life, and there is a need to improve their ecological validity, sensitivity, and specificity (Rizzo et al., 2004; Negu et al., 2016; Plancher and Piolino, 2017; Kim et al., 2019). The available objective methods for assessing MCI are frequently based on informant-reports or conducted in isolated and artificial situations, thus opening the possibility for evaluation biases. A resounding change might be fostered by a novel approach assembling in new ways existing technologies and data analysis methodologies that allow a refined assessment and the creation of a continuum for MCI following a dimensional approach. These technologies aim to integrate rather than replace existing neuropsychological evaluation or caregiver/informant reports in order to obtain a more complete and dynamic picture of the strengths and critical aspects of the elderly person as they evolve over time.

This perspective aims to propose the development of a new integrated, multimethod, dimensional approach for early detection of MCI on the basis of behavioral data that incorporates existing, consolidated technologies, such as gait kinematic analysis, Virtual Reality (VR), and Machine Learning (ML), in the conventional assessment of MCI. The outcome would be a finer, continuous, time-sensitive assessment of MCI, in line with a dimensional approach compliant with new DSM-5 guidelines (American Psychiatric Association, 2013). Moreover, it would draw on recent empirical evidence and scientific groundwork, helping the clinician to tailor the rehabilitation to the needs of the individual. This positive contribution would improve their quality of life, decreasing both health care assistance costs and hospitalization rates, thus opening up new possibilities for primary and secondary prevention. Moreover, it would facilitate the communication between practitioners and researchers, providing a solid foundation and fostering mutual exchange.

## A NEW INTEGRATED APPROACH TO MCI ASSESSMENT

We suggest that Virtual Reality would provide the most suitable context (i.e., answering the question *Where?*) for the assessment

of key behavioral variables indicating MCI onset (i.e., *What?*), which can be analyzed in a systematic and accurate way in relation to neuropsychological and clinical data by means of Machine Learning (ML) (i.e., *How?*). We expand on all of these aspects in the following sections.

## “Where” Does the Assessment Take Place? Virtual Reality

Usually, the assessment of cognitive functions does not take place in daily-life contexts, potentially hindering an ecological evaluation of the individual's impairment (Rizzo et al., 2004; de los Reyes-Guzmán et al., 2014). A promising integration of conventional practices could rely on novel dimensional assessment techniques, based on realistic immersive simulations of daily situations, e.g., Virtual Reality (VR) – a 3D computer-generated environment with some degree of immersion and interactivity, along with a sense of being really present in it (Riva and Mantovani, 2014; Riva et al., 2018, 2019a; Moreno et al., 2019). VR has developed into a key technology that is able to resemble even complex daily situations and interactions in a safe and controlled setting, due to the feeling of *immersion* (i.e., the number of senses stimulated within the environment, together with the closeness of the stimuli employed in simulations to reality) (Slater, 2009; Plancher and Piolino, 2017; Cipresso, 2018), the *sense of presence* within the environment (i.e., the feeling of being really “there” in the simulated environment, along with the ability to realize our intentions within it), and the possibility to *interact* with objects (Biocca, 1997; Heeter, 2000; Bailenson et al., 2006; Sundar et al., 2010; Negu et al., 2016; Plancher and Piolino, 2017; Cipresso, 2018; Kim et al., 2019). Depending on the degree of immersion of the system employed, VR allows a realistic experience through the use of multi-sensorial displays (i.e., visual, auditory) along with tracking devices that detect any movement of the individual and deliver the recorded data to the visualization system for a real-time update of the virtual environment (Chirico et al., 2016; Plancher and Piolino, 2017; Cipresso, 2018). The most immersive 3D VR environments can provide a high sense of presence also by isolating individuals, facilitating natural interactions and exchanges that resemble equivalent ones in daily life (Gold and Gold, 2012; Allain et al., 2014; Riva and Mantovani, 2014; Chirico et al., 2016; Riva et al., 2018).

The main features of VR allow the creation of ecological, safe, standardized settings and exert a strict experimental control over stimulus delivery and measurement (Rizzo et al., 2004; Gold and Gold, 2012; Allain et al., 2014; Negu et al., 2016; Plancher and Piolino, 2017). This, in turn, has supported its deployment for both clinical and non-clinical samples of elderly people and young adults (García-Betances et al., 2015; De Tommaso et al., 2016; Plancher and Piolino, 2017). Within medical and neuropsychological settings, VR has been extensively applied as an assessment and a rehabilitation tool for elderly people suffering from consequences of a traumatic brain injury (Aida et al., 2018; Alashram et al., 2019; Maggio et al., 2019), for post-stroke patients (Henderson et al., 2007; Saposnik and Levin, 2011; Laver et al., 2017), and for spatial memory and balance

(Allain et al., 2014; Serino et al., 2017; Gerber et al., 2018; Soares et al., 2018), among other applications (see Plancher and Piolino, 2017; Moreno et al., 2019). Crucially, VR allows the therapy to be tailored in a controlled way, according to each disease starting from a continuous assessment of the individual's behaviors. Only recently, VR has been employed to assess IADL in MCI patients while including kinematic measures that integrate a neuropsychological evaluation (Seo et al., 2017). As previously mentioned, an initial cognitive decline can be behaviorally manifested by a slowdown in the execution of IADL (Kim and Kim, 2009; Millán-Calenti et al., 2010; Gold and Gold, 2012), which implies a neurological and cognitive alteration that is partially reflected in indexes such as bodily movements or gait. Previous studies have examined these behavioral alterations of IADL in order to refine MCI assessment and have already delivered promising results (Schröter et al., 2003; Montero-Odasso et al., 2009; de los Reyes-Guzmán et al., 2014). Motion detectors, applied to the elderly person's leg joints allow gait kinematics and their impairments to be tracked during the performance of IADL within a VR environment. This could consolidate preliminary findings of specific motor alterations that integrate neuropsychological and cognitive evaluation to identify MCI. In this perspective, the preliminary work of Seo et al. (2017) is the closest application of the technologies proposed to refine MCI assessment, although gait analysis was not included. The recording of kinematic measures from the performance of IADL within an immersive VR environment potentially adds more discriminative value in distinguishing MCI individuals from the healthy control group (Seo et al., 2017). Including an evaluation where the elderly person him/herself performs IADL might be essential for establishing more precise criteria (Díaz-Mardomingo et al., 2017; Seo et al., 2017). Several authors have tried to refine early MCI detection by combining two out of the three variables considered in this paper: either behavioral alteration (IADL, gait) within a VR environment (Lee et al., 2003; Seo et al., 2017; Kim et al., 2019; Eraslan Boz et al., 2019), gait kinematics extracted and analyzed by means of ML, which will be discussed further (Begg and Kamruzzaman, 2005; Pogorelc et al., 2012; Zhang and Wang, 2012; Eskofier et al., 2013; Akl et al., 2015; Costa et al., 2016; Mannini et al., 2016; Caldas et al., 2017; Farah et al., 2017; Ur Rehman et al., 2019), or ML techniques for predicting MCI evolution (Filipovych and Davatzikos, 2011; Williams and Weakley, 2013; Moradi et al., 2014, 2015; Bratić et al., 2018; Grassi et al., 2018, 2019; Graham et al., 2020). Thus, to our best knowledge, this is the first paper proposing an integration of VR, gait kinematics, and ML in order to refine early detection of MCI following a dimensional approach in line with the most recent diagnostic systems and possibly providing information on disease progression. However relevant, traditional neuropsychological assessment does not provide this extent of information and could serve as a starting point that should be integrated with further information in order to detect a subclinical condition otherwise undiagnosable following a categorical approach. Crucially, some anecdotal evidence and more systematic but scattered evidence from kinematic analysis

of specific movements (i.e., *What?*) suggest the feasibility and the relevance of an approach based on assessment of behavioral variables for early detection of MCI. We present preliminary evidence in this regard in the following.

## “What” Variables Are Included in the Assessment? Gait Kinematics

Is it possible to give relevance to behavioral data reported by the caregivers, relying on the anecdotal description of the elderly person's daily functioning and their IADL performance, in a scientific and rigorous manner?

A potential solution is to analyze the elderly individual's movements (kinematics) while performing IADL. Kinematic analysis automatically records movements in a controlled setting and assesses the underlying cognitive impairment. Preliminary studies proved the feasibility of tracking the elderly person's head, dominant hand, or gait during the performance of IADL to refine the assessment of MCI and other cognitive conditions (Schröter et al., 2003; de los Reyes-Guzmán et al., 2014; Akl et al., 2015; Seo et al., 2017). Among these indexes, gait kinematic analysis has progressively received more attention, despite the paucity of MCI-focused studies. The work of Martín-Gonzalo et al. (2019) thoroughly explains the contribution of considering gait alterations, beginning in early cognitive decline, to an improved understanding of neurocognitive disorders. In fact, gait kinematics are strongly related to neurophysiological alterations (Persad et al., 2008; Maquet et al., 2010; Martín-Gonzalo et al., 2019), brain volume changes in specific areas (Tian et al., 2017; Allali et al., 2019; Martín-Gonzalo et al., 2019), and subsequent cognitive decline, predicting future risks of impairment (Martín-Gonzalo et al., 2019). Kinematics assesses the sequential configuration of the leg joints required to maintain the body's center of gravity above the stance base while a person is moving forward. Compared to healthy subjects, the gait of a person suffering from MCI shows decreased velocity, longer stride time, increased stride-to-stride variability (Hausdorff, 2007; Bahureksa et al., 2017; Byun et al., 2018; Martín-Gonzalo et al., 2019), and spatiotemporal complexity (Ihlen et al., 2016; Martín-Gonzalo et al., 2019).

A gait cycle is defined by ongoing changes in the sequential configurations of the joints allowed by muscle activation, which is controlled by neural mechanisms depending on the integrity of somatosensory, motor, and cognitive integration cerebral networks (Perry and Burnfield, 2010; Caldas et al., 2017; Costilla-Reyes et al., 2020). Successful locomotion is indeed a dual task requiring the ability to simultaneously perform a cognitive task that could interfere with gait performance, particularly in elderly people (Pedroli et al., 2018; Costilla-Reyes et al., 2020). A decrease in attentional and executive functioning is physiological in aging and could impact this simultaneous execution (Hsu et al., 2012; Montero-Odasso et al., 2012; Wang et al., 2015; Gwak et al., 2018; Pedroli et al., 2018). In order to maintain walking capacity, damage to cerebral networks involved in gait leads to an adaptation of the nervous system, generating new signals reflecting the damage. Brain

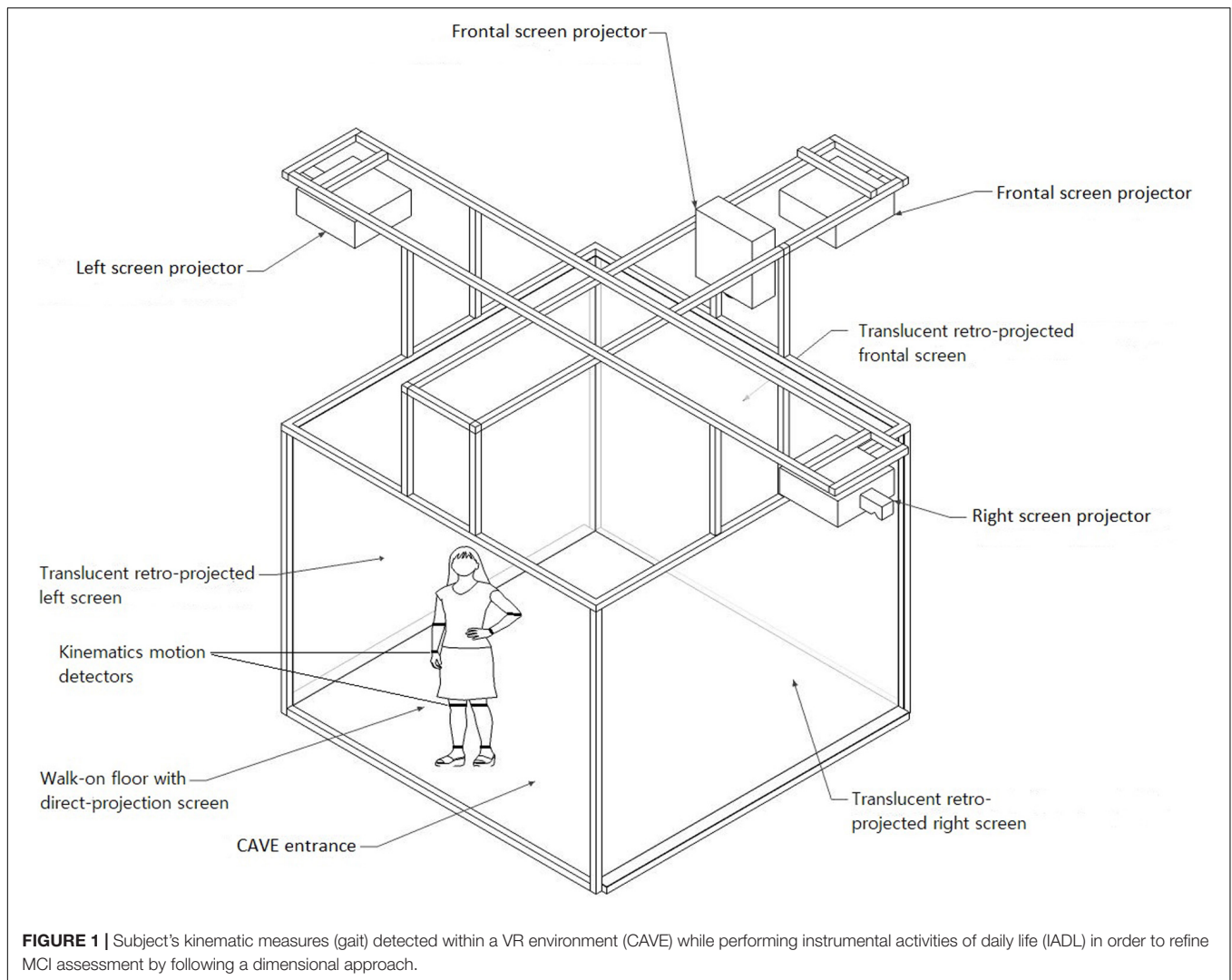
signals to the muscles controlling joint movement may become discontinuous and uncoordinated: this generates noise that could be consequent to the failure of some neuronal networks and produces configurations that respond to intentional cognitive directives, such as changing gait pace, little or not at all (Martín-Gonzalo et al., 2019). Indeed, kinematic data provide additional, crucial information that increases the sensitivity and specificity of MCI assessment. Paper-and-pencil neuropsychological tests are not suitable for the detection of gait features and its alterations, which appear relevant for more precise identification of MCI individuals.

To date, gait analysis has been studied within a context with little ecological validity: the walking task is generally an end in itself and is not recorded while the subject is completing a complex activity. Even the extraction of gait kinematics from videos or home-based motion sensors could provide only partial, bi-dimensional information or could be less sensitive in detecting real-time movement adjustment (Akl et al., 2015; Prakash et al., 2015; Neverova, 2016). The use of VR enables continuous, tridimensional tracking of the ongoing events within a highly immersive, safe, and standardized environment, enhancing the methodological strength of the procedure as well.

This introduces the need for a highly ecological and immersive context that allows the elderly person's kinematics while performing IADL to be observed and detected. A plausible solution comes from the implementation of Virtual Reality (VR), as shown in **Figure 1** (Pedroli et al., 2018). The technological equipment illustrated in **Figure 1** is a four-walled Cave Automated Virtual Environment (CAVE), available at Istituto Auxologico Italiano, which is routinely used for cognitive and motor rehabilitation of elderly people. This highly immersive technology is equipped with eight ( $4 \times 2$ ) Vicon Bonita 10 cameras (Opti-Tracking system, 1MP) and different Hi-res Hi-FOV head-tracked 3D HMDs and also with a wide range of physiological and motion measures for quantifying embodiment in VR and movements within the environment. A virtual representation of, e.g., a city or a supermarket can be projected on the four walls, and subjects can actively navigate and interact with the environment. This setting was used by Seo et al. (2017), which, as previously mentioned, is the most similar procedure to the one that we propose.

For the aims of this perspective, the most important feature of VR is its ability to detect both real-time behaviors (e.g., specific bodily movements such as those of the head and upper limbs and gait) and physiological indexes (e.g., skin conductance, heart rate). The great amount of data collected with VR and kinematics implies the need for a computational method of analysis that is able to extract meaning from a large amount of data. Thus, Machine Learning appears to be a viable solution, as shown by previous research employing this technique to discriminate between normal and pathological gait alteration and for diagnostic purposes as well (Pogorelec et al., 2012; Zhang and Wang, 2012; Eskofier et al., 2013; Costa et al., 2016; Mannini et al., 2016; Caldas et al., 2017; Farah et al., 2017; Ur Rehman et al., 2019).





## “How” to Analyze Them? Machine Learning

The massive amount of kinematic information extrapolated from motion detectors, complemented by neuropsychological and neuropsychiatric symptoms and signs, needs a similarly powerful technology in order to process it and convert it into an output intelligible for both clinicians and patients. The employment of kinematic measures and VR within a healthcare setting, i.e., a hospital, inevitably involves the use of a large amount of electronic health records (EHR) of patients' evolution over time. Despite the challenges related to the use of EHRs, several prediction algorithms and models have been developed from their use (Häyrynen et al., 2008; Miotto et al., 2016, 2017; Goldstein et al., 2017; Graham et al., 2020). Among other advantages, EHR-based predictors consider various metrics of multiple individuals, observed at different time points: this makes use of a higher frequency of data recording, facilitating the prediction of possible near-term evolution; they also reflect real life more closely than cohort studies (Goldstein et al., 2017).

The most suitable technique capable of administering a volume of complex and extensive information may be Machine Learning (ML). This scientific discipline stems from Artificial Intelligence (AI), i.e., a computer science field performing tasks capable of emulating human performance, generally learning to understand complex data, an endeavor that requires human intelligence (Bawack, 2019; Wang, 2019; Graham et al., 2020). ML algorithms have progressively gained popularity for several reasons, including their ability to automatically learn the inherent structure of a dataset (Kononenko, 2001; Abu-Mostafa et al., 2012; Facal et al., 2019) without requiring *a priori* hypotheses about relationships between variables (Miotto et al., 2017; Vieira et al., 2017; Graham et al., 2019, 2020). Conversely, ML algorithms can discover and predict data trends and patterns by building on existing information and highlight unexpected relationships between variables (Vieira et al., 2017; Graham et al., 2019, 2020). This “learning by processing” approach generates increasingly accurate predictive models and, so far, has demonstrated enormous potential for supporting individual prognosis, risk estimation, and classification learning

for diagnosis (Lehmann et al., 2007; Patel et al., 2015; Vieira et al., 2017; Dwyer et al., 2018; Facal et al., 2019). Inevitably, ML techniques work with high-dimensional data, which require a pre-processing step to remove redundant information, reduce data dimensionality, and improve learning accuracy and data comprehensibility (Khalid et al., 2014). This can be achieved by means of (i) feature selection (i.e., the selection of the best and most optimal features from a larger set of those useful for discriminating between classes to increase accuracy and generalizability); and (ii) feature extraction or dimensionality reduction (i.e., the transformation of original features to generate other, more significant features and reduce complexity) by means of Principal Component Analysis (PCA) or Independent Component Analysis (ICA), among other approaches (Khalid et al., 2014; Dwyer et al., 2018). The application of ML for healthcare purposes has been further developed into two main sub-classes, supervised (SL) and unsupervised (UL) techniques. SL jointly employs pre-labeled data, e.g., MCI versus healthy subjects, and additional features derived from clinical or neuroimaging sources to determine which feature predicts the pre-labeled data the most (Dwyer et al., 2018; Graham et al., 2020). SL operates with probabilistic and non-probabilistic classifiers (Naïve Bayes and Support Vector Machine, respectively), as well as with decision tree, linear, and logistic regression (Dhall and Kaur, 2020). UL techniques, instead, sets unlabeled and unstructured data, e.g., clinical notes, as a starting point to seek relationships or patterns and to learn general representations that enable the automatic extraction of information when building predictors (Miotto et al., 2017; Dwyer et al., 2018; Graham et al., 2020). The algorithms employed by UL include K-means clustering, PCA, and Artificial Neural Networks (ANN) (Dhall and Kaur, 2020).

However, at the time of data collection, it is unclear whether MCI subjects will progress toward other forms of dementia (e.g., AD) or convert back to normal cognition, and this evolution could become more evident over the years. This challenges data labeling: thus, researchers tackling MCI detection have employed semi-supervised learning (SSL) techniques capable of combining labeled and unlabeled data to improve the classification procedure (Zhu, 2008; Filipovych and Davatzikos, 2011; Moradi et al., 2015; Dwyer et al., 2018; Van Engelen and Hoos, 2020). A semi-supervised approach, therefore, allows cases to be managed by providing only partial data labels (Filipovych and Davatzikos, 2011). Several studies have employed MCI data as unlabeled data and have shown an improvement in the predictive performance of the model (Batmanghelich et al., 2011; Filipovych and Davatzikos, 2011; Ye et al., 2011; Moradi et al., 2015): this approach could be particularly feasible for the purpose of the integrated dimensional approach offered in the present perspective.

The ability to process raw data, the need for manual engineering of features, and the extensive expertise needed to perform the analyses represent the main limitations of conventional (shallow) ML techniques (Lecun et al., 2015; Vieira et al., 2017; Zhang et al., 2020). This has led to the

dissemination of deep learning (DL) algorithms, including Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) (Dhall and Kaur, 2020). DL outperforms ML in many ways, showing best-in-class performance and increased complexity in the computed function and addressing problems in multiple domains such as language and speech (Zhang et al., 2020). Moreover, it eliminates the need for manual feature engineering, reducing possible human biases and removing the need for advanced expertise (Zhang et al., 2020). DL is capable of learning data representation in an unprocessed or raw form, and its high performance and expressive power in one specific domain can be transferred to other contexts, providing a flexible adaptation to problems (Bengio, 2009; Lecun et al., 2015; Miotto et al., 2017; Vieira et al., 2017; Chauhan et al., 2019; Esteva et al., 2019; Costilla-Reyes et al., 2020; Zhang et al., 2020). Despite all the advantages, it is crucial to consider that DL techniques require very large datasets to perform, which may be too hard to achieve, expensive, or time-consuming to obtain; thus, ML may be more feasible and efficient (Zhang et al., 2020).

To date, advanced statistical ML and pattern recognition techniques have proved their usefulness in outlining neurodegenerative patterns of mild symptoms manifesting during the early stages of diseases, and MCI is no exception (Davatzikos et al., 2008, 2010; Vemuri et al., 2009; Wee et al., 2014). ML has been repeatedly applied to diagnostic transitions from MCI to other forms of dementia, e.g., AD, employing different types of information: mostly neuroimaging data (e.g., MRI, PET scan, Diffusion Tensor Imaging) (Batmanghelich et al., 2011; Filipovych and Davatzikos, 2011; Ye et al., 2011; Zhang and Shen, 2011, 2012; O'Dwyer et al., 2012; Shaffer et al., 2013; Moradi et al., 2015; Bratić et al., 2018), cerebrospinal fluid biomarkers (Davatzikos et al., 2010; Fjell et al., 2010; Zhang and Shen, 2011; Shaffer et al., 2013; Bratić et al., 2018), demographic and cognitive data (Moradi et al., 2015; Bratić et al., 2018), and gait kinematics (Mannini et al., 2016; Farah et al., 2017; Gwak et al., 2018). Broad variations in studies' results have been reported, as has the lack of a gold-standard ML algorithm to predict disease progression (Grassi et al., 2018, 2019; Chiu et al., 2019; Facal et al., 2019; Mallo et al., 2019). Specifically, Grassi and colleagues (Grassi et al., 2018, 2019) have recently developed clinically translatable ML algorithms to identify which subjects with pre-MCI and MCI will convert to AD (Grassi et al., 2018, 2019). ML likewise appears promising for precision medicine: given the patients' extreme heterogeneity of symptoms, medication response, and prognosis, the implementation of ML to create computational models of disease development tackles patients' diverseness (Fisher et al., 2019). Over the years, researchers have devised a number of disease progression models for both MCI and AD, relying on clinical and imaging data (Mueller et al., 2005; Ito et al., 2011; Rogers et al., 2012; Moradi et al., 2015; Miotto et al., 2017; Samper-Gonzalez et al., 2017; Fisher et al., 2019). Previous applications of ML to clinical data have proven useful in predicting a single outcome (e.g., the likelihood of conversion from MCI to AD) (Fisher et al., 2019). From a clinical point

of view, however, it would be important to predict the disease progression and trajectory for everyone, which is difficult with current data-driven modeling approaches.

In their latest work, Graham et al. (2020) support the employment of AI and ML for ranking those variables crucial for MCI assessment and cognitive impairment. The authors show that clinical and psychometric assessments appear promising for identifying individuals at high risk for cognitive impairment (Lins et al., 2017; Senanayake et al., 2017; Moreira and Namen, 2018), which could be better identified by means of brain imaging and neuropsychological data as well (Fan et al., 2018; Iizuka et al., 2019). Even more importantly, Graham et al. (2020) report on several studies employing novel techniques to detect cognitive impairment in MCI subjects as well, such as home-installed motion sensors (Akl et al., 2015) and multi-modal wearable activity devices (Gwak et al., 2018), therefore including behavioral data in ML analysis (Graham et al., 2020). However useful, for providing real-world behavioral data in an ecological context, the employment of motion sensors alone has shown substantial heterogeneity (Graham et al., 2020); therefore, VR appears a promising integrative solution achieved by simulating a supervised and controlled real-life-like environment.

## DISCUSSION

Although both elderly people and caregivers notice and report their concerns regarding behavioral, personality, and cognitive changes, MCI is a subclinical condition that remains undiagnosed by an official categorical system while progressively compromising the independent functioning of the elderly person. Although a possible regression to normal cognition is desirable, more often, MCI evolves toward other forms of dementia. A delayed diagnosis entails the worsening of the individual's conditions, greatly reducing the extent of possible interventions and making primary and secondary prevention essential (Van Vliet et al., 2011; Jekel et al., 2015; Roehr et al., 2019). However, MCI assessment should necessarily move beyond a stringent categorical approach in favor of a dimensional one able to include finer discrimination among early stages of MCI, thus reflecting the complexity of this construct. So far, its assessment has followed a dichotomous view, relying on neuropsychological instruments to test MCI's presence or absence. Despite their proven efficacy, a dimensional approach would integrate them by implementing existing technologies and data analysis methodologies, placing MCI on a continuum. With this in mind, this perspective aimed primarily to move forward, proposing a novel assessment that could enable a more accurate prevision of the trajectory of MCI decline, employing Virtual Reality (VR) for a continuous dimensional assessment of MCI behaviors in ecological and realistic tailored, safe, and controlled simulated contexts.

Since the individual's altered behavior reflects impaired cognition (Martín-Gonzalo et al., 2019), this proposal would allow early detection of MCI, enabling timely rehabilitative

interventions. Specifically, gait kinematics is a behavioral index whose analysis has proved sensitive to cerebral and cognitive alterations capable of discerning patients with cognitive decline from healthy individuals (Martín-Gonzalo et al., 2019). Nevertheless, few studies have specifically employed gait measurement as a possible marker to refine MCI assessment, and even then mainly in unfamiliar contexts, thus hindering ecological validity (Jekel et al., 2015; Seo et al., 2017). So far, MCI assessment has relied on neuropsychological measures rather than behavioral ones, despite the importance of the latter in revealing initial cognitive decline. When available, these behavioral data are generally based on informant-report questionnaires or reported as anecdotal information lacking scientific rigor (Van Vliet et al., 2011; de los Reyes-Guzmán et al., 2014; Kim et al., 2019). Behavioral data appear to provide a relevant contribution to MCI assessment: further research should deepen and consolidate the preliminary, promising evidence reported (Seo et al., 2017; Martín-Gonzalo et al., 2019).

It appears evident that a mere conventional neuropsychological assessment, however relevant, cannot provide such a high degree of information, giving rise to the necessity of integrating paper-and-pencil instruments and anecdotal evidence with behavioral alterations evaluated within a highly ecological and standardized setting, such as VR. A plausible, practical implementation of the approach could be structured as follows. During a first brief clinical interview, the practitioner could collect anamnestic and quantitative information from (i) the elderly person, relying on neuropsychological/neuropsychiatric and cognitive measures as well; and (ii) the caregiver, which could fill in informant-report IADL measures. A second appointment would be dedicated to VR-based assessment: the elderly person could perform IADL (e.g., money withdrawal, grocery shopping) within the CAVE virtual environment, while kinematic information of their performance would be simultaneously collected. These data could be provided by kinematics motion detectors placed on the individual's joints, as illustrated in Figure 1. The entire VR-kinematic assessment would last a maximum of 20 min to possibly avoid cybersickness, i.e., a form of motion sickness that

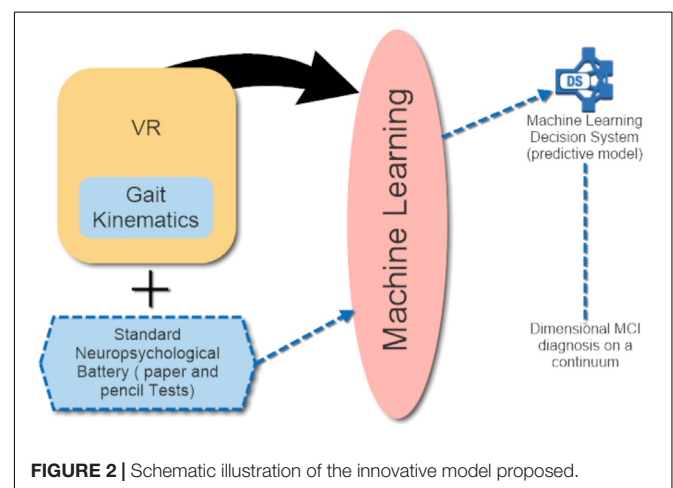


FIGURE 2 | Schematic illustration of the innovative model proposed.



includes nausea, headaches, and disorientation, among other symptoms (Laviola, 2000; Davis et al., 2014). Cybersickness is a common side effect of VR and could interfere with the completion of quantitative measures: thus, whenever it is necessary to complete paper-and-pencil assessment in the second appointment, this should be done before the VR procedure starts. VR would allow the clinician to closely observe the real life-like behavior of the individual and employ motion detectors, which extrapolate a large amount of data computed by means of Artificial Intelligence (AI) and, specifically, Machine Learning (ML). As mentioned in the ML section, recently developed, clinically translatable ML algorithms could help to identify MCI subjects who will convert to AD (Grassi et al., 2018, 2019). Thus, these algorithms could be tested and implemented in the assessment procedure illustrated in the previous section after collecting data from both the quantitative evaluation and the VR procedure within the CAVE. This could generate an accurate, predictive model proposing a gradient of behavioral and cognitive decline: a subclinical condition such as MCI could not be detected promptly by a categorical approach. A schematic illustration of this model is depicted in **Figure 2**.

The first and foremost added value of this approach lies in moving one step forward toward refined MCI early detection by integrating (i) behavioral (gait kinematics, IADL), neuropsychological/neuropsychiatric and cognitive information; (ii) a highly ecological and standardized setting, such as VR; and (iii) a powerful method capable of analyzing an extensive amount of data and predicting MCI progression over time. This is the first dimensional approach jointly considering all of the mentioned sources of information, whether previous studies considered only two out of three variables at the same time. The main focus is the relevance of building an innovative assessment procedure that is data-fusion-based and capable of identifying a subclinical condition that is otherwise undetected. Many ML and DL algorithms exist to analyze the extensive amount of data collected and, except for some that were recently tested (Grassi et al., 2018, 2019), there is no consensus regarding a gold-standard algorithm to predicting MCI diagnostic transition. Moreover, several open-source libraries for ML can provide information regarding the most feasible programming language (e.g., Python) and algorithms to use (Rathi, 2019).

We are aware that the integration of kinematic analysis, VR, and ML, could be very expensive and may not be available in a clinical setting, such as in a hospital. In addition, there may be risk of initial acceptance resistance by elderly individuals and healthcare providers due to the novelty of the equipment. However, the implementation of this approach would offer a crucial benefit by enabling the dimensional assessment of a subclinical condition otherwise undiagnosable, and the trained models, enriched by data of numerous patients, would easily overcome the initial expense. Moreover, a hospital would be the only setting where biological and neuroimaging data (e.g., MRI, PET scan, cerebrospinal fluid biomarkers) can be collected. Although the method proposed, so far, does not include them, these types of information could eventually be added to ML

analyses, since they have been previously indicated as plausible contributors to MCI assessment (Fjell et al., 2010; Batmanghelich et al., 2011; Filipovych and Davatzikos, 2011; Ye et al., 2011; Zhang and Shen, 2011, 2012; O'Dwyer et al., 2012; Shaffer et al., 2013; Moradi et al., 2015; Bratić et al., 2018; Chiu et al., 2019). The employment of ML and DL methods usually requires a large sample size, which may not always be feasible in the healthcare setting. However, this limitation could be settled by developing multi-centric studies, providing an adequate sample size of patients and sharing data (Vieira et al., 2017). The dimensional approach also needs to be applied carefully in order to avoid hypervigilance for the slightest cognitive and behavioral age-related alteration, which might lead to excessive diagnosis and false-positives (Vanacore et al., 2017). Diagnosis communication must be carefully handled, given the potential harm of anxiety about a condition that may not progress (Chiu, 2005; Díaz-Mardomingo et al., 2017; Vanacore et al., 2017) prognostic possibilities can be discussed and planned accordingly. Strengthening or rehabilitative interventions could foster regression to normal cognition or decelerate the progression toward other clinical conditions.

In summary, while the majority of the literature has studied the application of several combinations of VR, gait kinematic analysis, and ML, this is the first paper to integrate all of these three methods and techniques in order to refine early detection of MCI and possibly predict its evolution over time. VR allows the collection of “Digital Biomarkers” – physiological/behavioral data – by means of digital technologies, used as an indicator of biologic processes or responses to therapeutic interventions (Coravos et al., 2019) directly connected to brain functioning. On the other side, AI, by applying ML techniques to the individual's digital biomarkers, allows the creation of a predictive model – following a dimensional approach to MCI – able to identify specific behavioral cognitive patterns within an ecological and safe environment, for accurate early detection of MCI and its potential evolutionary trajectory (Riva et al., 2019b).

## AUTHOR CONTRIBUTIONS

SC, AC, and PC developed the new model integrating gait kinematics, Virtual Reality, and Machine Learning. PC and GR supervised the sections of Virtual Reality and Machine Learning. EP supervised the section regarding MCI. SC wrote the manuscript under the final supervision of AC, EP, GR, and PC. All authors have approved the final version of the manuscript.

## FUNDING

This research was funded by “Future Home for Future Communities” (“Accordo Quadro di Collaborazione tra Regione Lombardia e Consiglio Nazionale delle Ricerche” – Convenzione Operativa n 19365/RCC) and by the Italian Ministry of Health research project “High-End and Low-End Virtual Reality Systems for the Rehabilitation of Frailty in the Elderly” (PE-2013-0235594).



## REFERENCES

- Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. (2012). *Learning From Data*, Vol. 4. Pasadena, CA: AML Book.
- Aida, J., Chau, B., and Dunn, J. (2018). Immersive virtual reality in traumatic brain injury rehabilitation: a literature review. *Neuro Rehabil.* 42, 441–448. doi: 10.3233/NRE-172361
- Akl, A., Taati, B., and Mihailidis, A. (2015). Autonomous unobtrusive detection of mild cognitive impairment in older adults. *IEEE Trans. Biomed. Eng.* 62, 1383–1394. doi: 10.1109/TBME.2015.2389149
- Alashram, A. R., Annino, G., Padua, E., Romagnoli, C., and Mercuri, N. B. (2019). Cognitive rehabilitation post traumatic brain injury: a systematic review for emerging use of virtual reality technology. *J. Clin. Neurosci.* 66, 209–219. doi: 10.1016/j.jocn.2019.04.026
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008
- Allain, P., Foloppe, D. A., Besnard, J., Yamaguchi, T., Etcharry-Bouyx, F., Le Gall, D., et al. (2014). Detecting everyday action deficits in Alzheimer's disease using a nonimmersive virtual reality kitchen. *J. Int. Neuropsychol. Soc.* 20, 468–477. doi: 10.1017/S1355617714000344
- Allali, G., Montembeault, M., Saj, A., Wong, C. H., Cooper-Brown, L. A., Bherer, L., et al. (2019). Structural brain volume covariance associated with gait speed in patients with amnesic and non-amnesic mild cognitive impairment: a double dissociation. *J. Alzheimers Dis.* 71, S29–S39. doi: 10.3233/JAD-190038
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR)*, 4th Edn. Washington, DC: American Psychiatric Association.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596
- Apostolova, L. G., and Cummings, L. (2008). Neuropsychiatric manifestations in mild cognitive impairment: a systematic review of the literature. *Dement Geriatr. Cogn. Disord.* 25, 115–126. doi: 10.1159/000112509
- Bahureksa, L., Najafi, B., Saleh, A., Sabbagh, M., Coon, D., Mohler, M. J., et al. (2017). The impact of Mild Cognitive Impairment on gait and balance: a systematic review and meta-analysis of studies using instrumented assessment. *Gerontology* 63, 67–83. doi: 10.1159/000445831
- Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence* 15, 359–372. doi: 10.1162/pres.15.4.359
- Batmanghelich, K. N., Ye, D. H., Pohl, K. M., Taskar, B., and Davatzikos, C. (2011). "Disease classification and prediction via semi-supervised dimensionality reduction," in *Proceedings of the International Symposium on Biomedical Imaging*, Chicago, IL, doi: 10.1109/ISBI.2011.5872590
- Bawack, R. E. (2019). "Artificial intelligence in practice: implications for information systems research," in *Proceedings of the Twenty-fifth Americas Conference on Information Systems*, Cancun, ME.
- Begg, R., and Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *J. Biomech.* 38, 401–408. doi: 10.1016/j.jbiomech.2004.05.002
- Bengio, Y. (2009). Learning deep architectures for AI. *Found. Trends Mach. Learn.* 1, 1–27. doi: 10.1561/2200000006
- Biocca, F. (1997). The cyborg's dilemma: progressive embodiment in virtual environments [1]. *J. Comput. Commun.* 3:3428. doi: 10.1111/j.1083-6101.1997.tb00070.x
- Bratić, B., Kurbalija, V., Ivanović, M., Oder, I., and Bosnić, Z. (2018). Machine learning for predicting cognitive diseases: methods, data sources and risk factors. *J. Med. Syst.* 42:243. doi: 10.1007/s10916-018-1071-x
- Byun, S., Han, J. W., Kim, T. H., Kim, K., Kim, T. H., Park, J. Y., et al. (2018). Variability can predict the risk of cognitive decline in cognitively normal older people. *Dement. Geriatr. Cogn. Disord.* 45, 251–261. doi: 10.1159/000489927
- Caldas, R., Mundt, M., Potthast, W., Buarque de Lima, N. F., and Markert, B. (2017). A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait Posture* 57, 204–210. doi: 10.1016/j.gaitpost.2017.06.019
- Chauhan, S., Vig, L., De Filippo De Grazia, M., Corbetta, M., Ahmad, S., and Zorzi, M. (2019). A comparison of shallow and deep learning methods for predicting cognitive performance of stroke patients from mri lesion images. *Front. Neuroinform.* 13:53. doi: 10.3389/fninf.2019.00053
- Chirico, A., Yaden, D. B., Riva, G., and Gaggioli, A. (2016). The potential of virtual reality for the investigation of awe. *Front. Psychol.* 7:1766. doi: 10.3389/fpsyg.2016.01766
- Chiu, E. (2005). Limitations in the current classification system for dementia. *Int. Psychogeriatr.* 17(Suppl. 1), 17–26. doi: 10.1017/S1041610205001924
- Chiu, P., Tang, H., Wei, C., Zhang, C., and Hung, G. (2019). NMD-12: a new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. *PLoS One* 14:e0213430. doi: 10.1371/journal.pone.0213430
- Cipresso, P. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front. Psychol.* 9:2086. doi: 10.3389/fpsyg.2018.02086
- Coravos, A., Khozin, S., and Mandl, K. D. (2019). Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digit. Med.* 2:14. doi: 10.1038/s41746-019-0090-4
- Costa, L., Gago, M. F., Yelshyna, D., Ferreira, F., Silva, H. D., Rocha, L. A., et al. (2016). Application of machine learning in postural control kinematics for the diagnosis of Alzheimer's Disease. *Comput. Intell. Neurosci.* 2016, 1–2. doi: 10.1155/2016/3891253
- Costilla-Reyes, O., Vera-Rodriguez, R., Alharthi, A. S., Yunas, S. U., and Ozanyan, K. B. (2020). "Deep learning in gait analysis for security and healthcare," in *Deep Learning: Algorithms and Applications. Studies in Computational Intelligence*, eds W. Pedrycz and S. M. Chen (Cham: Springer), 865. doi: 10.1007/978-3-030-31760-7\_10
- Crocco, E. A., Loewenstein, D. A., Curiel, R. E., Alperin, N., Czaja, S. J., and Harvey, P. D. (2018). A novel cognitive assessment paradigm to detect Pre-mild cognitive impairment (PreMCI) and the relationship to biological markers of Alzheimer's disease. *J. Psychiatr. Res.* 96, 33–38. doi: 10.1016/j.jpsychires.2017.08.015.A
- Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., and Trojanowski, J. Q. (2010). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* 32, e19–e27. doi: 10.1016/j.neurobiolaging.2010.05.023
- Davatzikos, C., Fan, Y., Wu, X., Shen, D., and Resnick, S. M. (2008). Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging* 29, 514–523. doi: 10.1016/j.neurobiolaging.2006.11.010
- Davis, S., Nesbitt, K., and Nalivaiko, E. (2014). "A Systematic review of cybersickness," in *ACM International Conference Proceeding Series*, (New York, NY: ACM), doi: 10.1145/2677758.2677780
- de los Reyes-Guzmán, A., Dimbwadyo-Terrer, I., Trincado-Alonso, F., Monasterio-Huelin, F., Torricelli, D., and Gil-Agudo, A. (2014). Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: a review. *Clin. Biomech.* 29, 719–727. doi: 10.1016/j.clinbiomech.2014.06.013
- De Tommaso, M., Ricci, K., Delussi, M., Montemurro, A., Vecchio, E., Brunetti, A., et al. (2016). Testing a novel method for improving wayfinding by means of a P3b virtual reality visual paradigm in normal aging. *Springerplus* 5:1297. doi: 10.1186/s40064-016-2978-7
- Dhall, D., and Kaur, R. J. M. (2020). "Machine learning: a review of the algorithms and its application," in *Proceedings of the ICRIC 2019. Lecture Notes in Electrical Engineering*, Vol. 597, eds P. Singh, A. Kar, Y. Singh, and M. T. S. Kolekar (Cham: Springer), doi: 10.1007/978-3-030-29407-6\_5
- Díaz-Mardomingo, M., García-Herranz, S., Rodríguez-Fernández, R., Venero, C., and Peraíta, H. (2017). Problems in classifying mild cognitive impairment (mci): one or multiple syndromes? *Brain Sci.* 7:111. doi: 10.3390/brainsci7090111
- Dwyer, D. B., Falkai, P., and Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* 14, 91–118. doi: 10.1146/annurev-clinpsy-032816-045037
- Eraslan Boz, H., Limoncu, H., Zygouris, S., Tsolaki, M., Giakoumis, D., Votis, K., et al. (2019). A new tool to assess amnesic mild cognitive impairment in

- Turkish older adults: virtual supermarket (VSM). *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 1–15. doi: 10.1080/13825585.2019.1663146
- Eskofier, B. M., Federolf, P., Kugler, P. F., and Nigg, B. M. (2013). Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. *Comput. Methods Biomech. Biomed. Engin.* 16, 435–442. doi: 10.1080/10255842.2011.624515
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., et al. (2019). A guide to deep learning in healthcare. *Nat. Med.* 25, 24–29. doi: 10.1038/s41591-018-0316-z
- Facal, D., Valladares—Rodríguez, S., Lojo—Seoane, C., Pereiro, A. X., Anido—Rifon, L., Juncos—Rabadán, O., et al. (2019). Machine learning approaches to studying the role of cognitive reserve in conversion from mild cognitive impairment to dementia. *Int. J. Geriatr. Psychiatry* 34, 941–949. doi: 10.1002/gps.5090
- Fan, M., Yang, A. C., Fuh, J. L., and Chou, C. A. (2018). Topological pattern recognition of severe Alzheimer's disease via regularized supervised learning of EEG complexity. *Front. Neurosci.* 12:685. doi: 10.3389/fnins.2018.00685
- Farah, J. D., Baddour, N., and Lemaire, E. D. (2017). "Gait phase detection from thigh kinematics using machine learning techniques," in *Proceedings of the 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, Rochester, MN, 263–268. doi: 10.1109/MeMeA.2017.7985886
- Filipovych, R., and Davatzikos, C. (2011). Semi-supervised pattern classification of medical images: Application to mild cognitive impairment (MCI). *Neuroimage* 1109–1119. doi: 10.1016/j.neuroimage.2010.12.066
- Fisher, C. K., Smith, A. M., Walsh, J. R., Against, C., and Diseases, M. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Sci. Rep.* 9:13622. doi: 10.1038/s41598-019-49656-2
- Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., et al. (2010). CSF biomarkers in prediction of cerebral and clinical change in Mild Cognitive Impairment and Alzheimer's Disease. *J. Neurosci.* 30, 2088–2101. doi: 10.1523/JNEUROSCI.3785-09.2010
- García-Betances, R. I., Jiménez-Mixco, V., Arredondo, M. T., and Cabrera-Umpiérrez, M. F. (2015). Using virtual reality for cognitive training of the elderly. *Am. J. Alzheimers Dis. Dement.* 30, 49–54. doi: 10.1177/1533317514545866
- Gerber, S., Müri, R., Mosimann, U., Nef, T., and Urwyler, P. (2018). "Virtual reality for activities of daily living training in neurorehabilitation: a usability and feasibility study in healthy participants," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine 446 and Biology Society. IEEE Engineering in Medicine and Biology Society Conference*, Piscataway, NJ: IEEE, 1–4. doi: 10.1109/EMBC.2018.8513003
- Gold, D. A., and Gold, D. A. (2012). An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J. Clin. Exp. Neuropsychol.* 34, 37–41. doi: 10.1080/13803395.2011.614598
- Goldstein, B. A., Navar, A. M., Pencina, M. J., and Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J. Am. Med. Informatics Assoc.* 24, 198–208. doi: 10.1093/jamia/ocw042
- Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., et al. (2019). Artificial intelligence for mental health and mental illnesses: an overview. *Curr. Psychiatry Rep.* 21:116. doi: 10.1007/s11920-019-1094-0
- Graham, S. A., Leeabcd, E. E., Jestaebcd, D. V., Pattenabc, R. V., Twamleyabcd, E. W., Nebekerbcf, C., et al. (2020). Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: a conceptual review. *Psychiatry Res.* 284:112732. doi: 10.1016/j.psychres.2019.112732
- Grassi, M., Perna, G., Caldirola, D., Schruers, K., Duara, R., and Loewenstein, D. A. (2018). A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment. *J. Alzheimers Dis.* 61, 1555–1573. doi: 10.3233/JAD-170547
- Grassi, M., Rouleaux, N., Caldirola, D., Loewenstein, D., Schruers, K., Perna, G., et al. (2019). A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* 10:756. doi: 10.3389/fneur.2019.00756
- Gwak, M., Sarrafzadeh, M., and Woo, E. (2018). Support for a clinical diagnosis of mild cognitive impairment using photoplethysmography and gait sensors. *Proc. APSIPA Annu. Summit Conf.* 2018, 671–678.
- Hausdorff, J. M. (2007). Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking. *Hum. Mov. Sci.* 26, 555–589. doi: 10.1016/j.humov.2007.05.003
- Häyriinen, K., Saranto, K., and Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int. J. Med. Inform.* 77, 291–304. doi: 10.1016/j.ijmedinf.2007.09.001
- Heeter, C. (2000). Interactivity in the context of designed experiences. *J. Interact. Advert.* 1, 3–14. doi: 10.1080/15252019.2000.10722040
- Henderson, A., Korner-Bitensky, N., and Levin, M. (2007). Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. *Top. Stroke Rehabil.* 14, 52–56. doi: 10.1310/tsr1402-52
- Hsu, C. L., Nagamatsu, L. S., Davis, J. C., and Liu-Ambrose, T. (2012). Examining the relationship between specific cognitive processes and falls risk in older adults: a systematic review. *Osteoporos. Int.* 23, 2409–2424. doi: 10.1007/s00198-012-1992-z
- Hughes, T. F., Snitz, B. E., and Ganguli, M. (2011). Should mild cognitive impairment be subtyped? *Curr. Opin. Psychiatry* 24, 237–242. doi: 10.1097/YCO.0b013e328344696b
- Ihlen, E. A., Weiss, A., Bourke, A., Helbostad, J. L., and Hausdorff, J. M. (2016). The complexity of daily life walking in older adult community-dwelling fallers and non-fallers. *J. Biomech.* 49, 1420–1428. doi: 10.1016/j.jbiomech.2016.02.055
- Iizuka, T., Fukasawa, M., and Kameyama, M. (2019). Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. *Sci. Rep.* 9:8944. doi: 10.1038/s41598-019-45415-5
- Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., et al. (2011). Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. *Alzheimers Dement* 7, 151–160. doi: 10.1016/j.jalz.2010.03.018
- Jekel, K., Damian, M., Wattmo, C., Hausner, L., Bullock, R., Connelly, P. J., et al. (2015). Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. *Alzheimers Res. Ther.* 18:17. doi: 10.1186/s13195-015-0099-0
- Khalid, S., Khalil, T., and Nasreen, S. (2014). "A survey of feature selection and feature extraction techniques in machine learning," in *Proceedings of the 2014 Science and Information Conference, SAI 2014*, London, doi: 10.1109/SAI.2014.6918213
- Kim, J., and Kim, S. (2009). Impairment of instrumental activities of daily living in patients with mild cognitive impairment. *Psychiatry Investig* 6, 180–184. doi: 10.4306/pi.2009.6.3.180
- Kim, O., Pang, Y., and Kim, J. (2019). The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. *BMC Psychiatry* 19:219. doi: 10.1186/s12888-019-2180-x
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artif. Intell. Med.* 23, 89–109. doi: 10.1016/S0933-3657(01)00077-X
- Laver, K. E., Lange, B., George, S., Deutsch, J. E., Saposnik, G., Crotty, M., et al. (2017). Virtual reality for stroke rehabilitation. *Cochr. Database Syst. Rev.* 11:CD008349. doi: 10.1002/14651858.CD008349.pub4
- Laviola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* 32:344. doi: 10.1145/333329.333344
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521:436–484. doi: 10.1038/nature14539
- Lee, J. H., Ku, J., Cho, W., Hahn, W. Y., Kim, Y., Lee, S.-M., et al. (2003). A virtual reality system for the assessment and rehabilitation of the activities of daily living. *Cyberpsychol. Behav.* 6, 383–388. doi: 10.1089/109493103322278763
- Lehmann, C., Koenig, T., Jelic, V., Prichep, L., John, R. E., Wahlund, L. O., et al. (2007). Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J. Neurosci. Methods* 161, 342–350. doi: 10.1016/j.jneumeth.2006.10.023
- Lins, A. J. C. C., Muniz, M. T. C., Garcia, A. N. M., Gomes, A. V., Cabral, R. M., and Bastos-Filho, C. J. A. (2017). Using artificial neural networks to select the parameters for the prognostic of mild cognitive impairment and dementia in elderly individuals. *Comp. Methods Progr. Biomed.* 2, 93–104. doi: 10.1016/j.cmpb.2017.09.013

- Maggio, M. G., De Luca, R., Molonia, F., Porcari, B., Destro, M., Casella, C., et al. (2019). Cognitive rehabilitation in patients with traumatic brain injury: a narrative review on the emerging use of virtual reality. *J. Clin. Neurosci.* 61, 1–4. doi: 10.1016/j.jocn.2018.12.020
- Mallo, S. C., Valladares-Rodríguez, S., Facal, D., Lojo-Seoane, C., Fernández-Iglesias, M. J., Pereiro, A. X., et al. (2019). Neuropsychiatric symptoms as predictors of conversion from MCI to dementia: a machine learning approach. *Int. Psychogeriatr.* 32, 381–392. doi: 10.1017/S1041610219001030
- Mannini, A., Trojaniello, D., Cereatti, A., and Sabatini, A. M. (2016). A machine learning framework for gait classification using inertial sensors: application to elderly, post-stroke and huntington's disease patients. *Sens. (Switzerland)* 16:134. doi: 10.3390/s16010134
- Maquet, D., Lekeu, F., Warzee, E., Gillain, S., Wojtasik, V., Salmon, E., et al. (2010). Gait analysis in elderly adult patients with Mild Cognitive Impairment and patients with Mild Alzheimer's disease: simple versus dual task: a preliminary report. *Clin. Physiol. Funct. Imaging* 30, 51–56. doi: 10.1111/j.1475-097X.2009.00903.x
- Martín-Gonzalo, J. A., Pulido-Valdeolivas, I., Wang, Y., Wang, T., Chiclana-Actis, T., Algarra-Lucas, M. D. A., et al. (2019). Permutation entropy and irreversibility in gait kinematic time series from patients with mild cognitive decline and early Alzheimer's dementia. *Entropy* 21:868. doi: 10.3390/e21090868
- Mckhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- Michaud, T. L., Su, D., Siahpush, M., and Murman, D. L. (2017). The risk of incident mild cognitive impairment and progression to dementia considering mild cognitive impairment subtypes. *Dement. Geriatr. Cogn. Dis. Extra* 7, 15–29. doi: 10.1159/000452486
- Millán-Calenti, J. C., Tubío, J., Pita-Fernández, S., González-Abraldes, I., Lorenzo, T., Fernández-Aruty, T., et al. (2010). Prevalence of functional disability in activities of daily living (ADL), instrumental activities of daily living (IADL) and associated factors, as predictors of morbidity and mortality. *Arch. Gerontol. Geriatr.* 50, 306–310. doi: 10.1016/j.archger.2009.04.017
- Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016). Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* 6:26094. doi: 10.1038/srep26094
- Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2017). Deep learning for healthcare: review opportunities and challenges. *Brief Bioinform* 19, 1236–1246. doi: 10.1093/bib/bbx044
- Montero-Odasso, M., Casas, A., Hansen, K. T., Bilski, P., Gutmanis, I., Wells, J. L., et al. (2009). Quantitative gait analysis under dual-task in older people with mild cognitive impairment: a reliability study. *J. Neuroeng. Rehabil.* 6, 35. doi: 10.1186/1743-0003-6-35
- Montero-Odasso, M., Verghese, J., Beauchet, O., and Hausdorff, J. M. (2012). Gait and cognition: a complementary approach to understanding brain function and the risk of falling. *J. Am. Geriatr. Soc.* 60, 2127–2136. doi: 10.1111/j.1532-5415.2012.04209.x
- Moradi, E., Pepe, A., Gaser, C., Huttunen, H., and Tohka, J. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* 104, 398–412. doi: 10.1016/j.neuroimage.2014.10.002
- Moradi, E., Tohka, J., and Gaser, C. (2014). "Semi-supervised learning in MCI-to-ad conversion prediction - When is unlabeled data useful?" in *Proceedings of the 2014 International Workshop on Pattern Recognition in Neuroimaging*, Tubingen, doi: 10.1109/PRNI.2014.6858535
- Moreira, L. B., and Namen, A. A. (2018). A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Comput. Methods Programs Biomed.* 165, 139–149. doi: 10.1016/j.cmpb.2018.08.016
- Moreno, A., Wall, K. J., Thangavelu, K., Craven, L., Ward, E., and Dissanayaka, N. N. (2019). A systematic review of the use of virtual reality and its effects on cognition in individuals with neurocognitive disorders. *Alzheimers Dement. Transl. Res. Clin. Interv.* 5, 834–850. doi: 10.1016/j.trci.2019.09.016
- Morris, J. C. (2012). Revised criteria for mild cognitive impairment may compromise the diagnosis of alzheimer disease dementia. *Arch. Neurol.* 69, 700–708. doi: 10.1001/archneurol.2011.3152
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* 1, 55–66. doi: 10.1016/j.jalz.2005.06.003
- Negu, A., Matu, S. A., Sava, F. A., and David, D. (2016). Virtual reality measures in neuropsychological assessment: a meta-analytic review. *Clin. Neuropsychol.* 30, 165–184. doi: 10.1080/13854046.2016.1144793
- Neverova, N. (2016). *Deep Learning for Human Motion Analysis*. Ph.D. Thesis, Christian Wolf, Graham Taylor.
- O'Dwyer, L., Lamberton, F., Bokde, A. L. W., Ewers, M., Faluy, Y. O., Tanner, C., et al. (2012). Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment. *PLoS One* 7:e32441. doi: 10.1371/journal.pone.0032441
- Okonkwo, O. C., Wadley, V. G., Griffith, H. R., Belue, K., Lanza, S., Zamrini, E. Y., et al. (2008). Awareness of deficits in financial abilities in patients with Mild Cognitive Impairment: going beyond self-informant discrepancy. *Am. J. Geriatr. Psychiatry* 16, 650–659. doi: 10.1097/JGP.0b013e31817e8a9d
- Patel, M. J., Andreescu, C., Price, J. C., Edelman, K. L., Reynolds, C. F., and Aizenstein, H. J. (2015). Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. *Int. J. Geriatr. Psychiatry* 30, 1056–1067. doi: 10.1002/gps.4262
- Pedroli, E., Greci, L., Colombo, D., Serino, S., Cipresso, P., Arlati, S., et al. (2018). Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: positive bike. *Sensors* 18:2343. doi: 10.3390/s18072343
- Perry, J., and Burnfield, J. (2010). GAIT normal and pathological function. *J. Sports Sci. Med.* 9:353.
- Persad, C. C., Jones, J. L., Ashton-Miller, J. A., Alexander, N. B., and Giordani, B. (2008). Executive function and gait in older adults with cognitive impairment. *J. Gerontol. A. Biol. Sci. Med. Sci.* 63, 1350–1355. doi: 10.1093/gerona/63.12.1350
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x
- Petersen, R. C., Lopez, O., Armstrong, M. J., and Getchius, T. S. D. (2018). Practice guideline update summary: mild cognitive impairment report of the guideline development, dissemination, and implementation. *Neurology* 90, 126–135. doi: 10.1212/WNL.0000000000004826
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild Cognitive Impairment: clinical characterization and outcome. *Arch. Neurol.* 56, 303–308. doi: 10.1001/archneur.56.3.303
- Pfeifer, L., Drobetz, R., Fankhauser, S., Mortby, M. E., Maercker, A., and Forstmeier, S. (2013). Caregiver rating bias in mild cognitive impairment and mild Alzheimer's disease: Impact of caregiver burden and depression on dyadic rating discrepancy across domains. *Int. Psychogeriatrics* 25, 1345–1355. doi: 10.1017/S1041610213000562
- Plancher, G., and Piolino, P. (2017). "Virtual Reality for assessment of episodic memory in normal and pathological aging," in *The Role of Technology in Clinical Neuropsychology*, eds R. L. Kane and T. D. Parsons (Oxford: Oxford University Press).
- Pogorelec, B., Bosnić, Z., and Gams, M. (2012). Automatic recognition of gait-related health problems in the elderly using machine learning. *Multimed. Tools Appl.* 58, 333–354. doi: 10.1007/s11042-011-0786-1
- Prakash, C., Gupta, K., Mittal, A., Kumar, R., and Laxmi, V. (2015). Passive marker based optical system for gait kinematics for lower extremity. *Proc. Comput. Sci.* 45:116. doi: 10.1016/j.procs.2015.03.116
- Rathi, S. (2019). *Top 8 Python Libraries for Machine Learning & Artificial Intelligence*. Available online at: <https://hackernoon.com/top-8-python-libraries-for-machine-learning-and-artificial-intelligence-y08id3031> (accessed May 22, 2020).
- Riva, G., and Mantovani, F. (2014). "Extending the self through the tools and the others: a general framework for presence and social presence in mediated interactions," in *Interacting with Presence: HCI and the Sense of Presence in Computer-Mediated Environments*, eds G. Riva and J. Waterworth (Berlin: DeGruyter Open Ltd), 9–31. doi: 10.2478/9783110409697.1
- Riva, G., Wiederhold, B. K., Chirico, A., Di Lernia, D., Mantovani, F., and Gaggioli, A. (2018). Brain and virtual reality: what do they have in common and how to exploit their potential. *Annu. Rev. CyberTherapy Telemed.* 16, 3–7.



- Riva, G., Wiederhold, B. K., Di Lerna, D., Chirico, A., Riva, E. F. M., Mantovani, F., et al. (2019a). Virtual reality meets artificial intelligence: the emergence of advanced digital therapeutics and digital biomarkers. *Annu. Rev. CyberTherapy Telemed.* 18, 3–7.
- Riva, G., Wiederhold, B. K., and Mantovani, F. (2019b). Neuroscience of virtual reality: from virtual exposure to embodied medicine. *Cyberpsychol. Behav. Soc. Netw.* 22, 82–96. doi: 10.1089/cyber.2017.29099.gri
- Rizzo, A. A., Schultheis, M., Kerns, K. A., and Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychol. Rehabil.* 14, 207–239. doi: 10.1080/09602010343000183
- Roehr, S., Riedel-Heller, S. G., Kaduszkiewicz, H., Wagner, M., Fuchs, A., van der Leeden, C., et al. (2019). Is function in instrumental activities of daily living a useful feature in predicting Alzheimer's disease dementia in subjective cognitive decline? *Int. J. Geriatr. Psychiatry* 34, 193–203. doi: 10.1002/gps.5010
- Rogers, J. A., Polhamus, D., Gillespie, W. R., Ito, K., Romero, K., Qiu, R., et al. (2012). Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. *J. Pharmacokinet. Pharmacodyn.* 39, 479–498. doi: 10.1007/s10928-012-9263-3
- Samper-Gonzalez, J., Burgos, N., Fontanella, S., Bertin, H., Habert, M.-O., Durrleman, S., et al. (2017). “Yet another ADNI machine learning paper? Paving the Way towards fully- reproducible research on classification of alzheimer's disease,” in *Machine Learning in Medical Imaging. MLMI 2017. Lecture Notes in Computer Science*, Vol. 10541, eds Q. Wang, Y. Shi, and H. I. Suk (Cham: Springer), doi: 10.1007/978-3-319-67389-9\_7
- Saposnik, G., and Levin, M. (2011). Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. *Stroke* 42, 1380–1386. doi: 10.1161/STROKEAHA.110.605451
- Schröter, A., Mergl, R., Bürger, K., Hampel, H., Möller, H. J., and Hegerl, U. (2003). Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dement. Geriatr. Cogn. Disord.* 15, 132–142. doi: 10.1159/000068484
- Senanayake, U., Sowmya, A., Dawes, L., Kochan, N., Wen, W., and Sachdev, P. (2017). “Deep learning approach for classification of mild cognitive impairment subtypes,” in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, (Valletta, MA: ICPRAM), doi: 10.5220/0006246306550662
- Seo, K., Kim, J., kwan Oh, D. H., Ryu, H., and Choi, H. (2017). Virtual daily living test to screen for mild cognitive impairment using kinematic movement analysis. *PLoS One* 12:e0181883. doi: 10.1371/journal.pone.0181883
- Serino, S., Pedrolì, E., Tuena, C., De Leo, G., Stramba-Badiale, M., Goulene, K., et al. (2017). Novel Virtual reality-based training protocol for the enhancement of the ‘mental frame syncing’ in individuals with alzheimer's disease: a development-of-concept trial. *Front. Aging Neurosci.* 9:240. doi: 10.3389/fnagi.2017.00240
- Shaffer, J. L., Petrella, J. R., Sheldon, F. C., Choudhury, K. R., Calhoun, V. D., Coleman, R. E., et al. (2013). Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* 266, 583–591. doi: 10.1148/radiol.12120010
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138
- Soares, J., Amorim, C. De, Leite, R. C., Brizola, R., and Yonamine, C. Y. (2018). Virtual reality therapy for rehabilitation of balance in the elderly: a systematic review and META-analysis. *Adv. Rheumatol.* 18, 1–10. doi: 10.1186/s42358-018-0013-0
- Sundar, S. S., Xu, Q., and Bellur, S. (2010). “Designing interactivity in media interfaces: a communications perspective,” in *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, Atlanta, GE, doi: 10.1145/1753326.1753666
- Tian, Q., Chastan, N., Bair, W. N., Resnick, S. M., Ferrucci, L., and Studenski, S. A. (2017). The brain map of gait variability in aging, cognitive impairment and dementia: a systematic review. *Neurosci. Biobehav. Rev.* 74, 149–162. doi: 10.1016/j.neubiorev.2017.01.020
- Ur Rehman, R. Z., Del Din, S., Shi, J. Q., Galna, B., Lord, S., Yarnall, A. J., et al. (2019). Comparison of walking protocols and gait assessment systems for machine learning-based classification of parkinson's disease. *Sensors (Switzerland)* 19:5363. doi: 10.3390/s19245363
- Van Engelen, J. E., and Hoos, H. H. (2020). A survey on semi-supervised learning. *Mach. Learn.* 109, 373–440. doi: 10.1007/s10994-019-05855-6
- Van Vliet, D., De Vugt, M. E., Bakker, C., Koopmans, R. T., Pijnenburg, Y. A., Vernooij-Dassen, M. J., et al. (2011). Caregivers' perspectives on the pre-diagnostic period in early onset dementia: a long and winding road. *Int. Psychogeriatrics* 23, 1393–1404. doi: 10.1017/S1041610211001013
- Vanacore, N., Di Pucchio, A., Lacorte, E., Bacigalupo, I., Mayer, F., Grande, G., et al. (2017). Dal mild cognitive impairment alla demenza: qual è il ruolo della sanità pubblica? *Recenti Prog Med* 208, 211–215.
- Vemuri, P., Wiste, H. J., Weigand, S. D., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., et al. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology* 73, 294–301. doi: 10.1212/WNL.0b013e3181af79fb
- Vieira, S., Pinaya, W. H. L., and Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci. Biobehav. Rev.* 74(Pt A), 58–75. doi: 10.1016/j.neubiorev.2017.01.002
- Walters, G. D. (2011). Dementia: continuum or distinct entity? *Psychol. Aging* 25, 534–544. doi: 10.1037/a0018167
- Wang, P. (2019). On defining artificial intelligence. *J. Artif. Gen. Intell.* 10:2478. doi: 10.2478/jagi-2019-0002
- Wang, X., Pi, Y., Chen, P., Liu, Y., Wang, R., and Chan, C. (2015). Cognitive motor interference for preventing falls in older adults: a systematic review and meta-analysis of randomized controlled trials. *Age Ageing* 44, 205–212. doi: 10.1093/ageing/afu175
- Wee, C., Zhang, D., Zhou, L., and Yap, P. (2014). “Machine learning techniques for AD / MCI diagnosis and prognosis,” in *Machine Learning in Healthcare Informatics. Intelligent Systems Reference Library*, Vol. 56, eds S. Dua, U. Acharya, and P. Dua (Berlin: Springer).
- Williams, J., and Weakley, A. (2013). *[THIS SPACE MUST BE KEPT BLANK] Machine Learning Techniques for Diagnostic Differentiation of Mild Cognitive Impairment and Dementia*. Pullman, WA: Washington State University.
- World Health Organization (2007). *International Classification of Diseases, (ICD-10)*, 10th Edn. Geneva: World Health Organization.
- Ye, D. H., Pohl, K. M., and Davatzikos, C. (2011). “Semi-supervised pattern classification: application to structural MRI of Alzheimer's disease,” in *Proceedings of the International Workshop on Pattern Recognition in NeuroImaging, PRNI 2011*, Seoul, doi: 10.1109/PRNI.2011.12
- Zhang, D., and Shen, D. (2011). “Semi-supervised multimodal classification of Alzheimer's disease,” in *Proceedings of the International Symposium on Biomedical Imaging*, Chicago, IL, doi: 10.1109/ISBI.2011.5872715
- Zhang, D., and Shen, D. (2012). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* 7:e33182. doi: 10.1371/journal.pone.0033182
- Zhang, S., Zhang, S., Wang, B., and Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access* 8, 29857–29881. doi: 10.1109/ACCESS.2020.2972859
- Zhang, Y., and Wang, Z. (2012). “Gait recognition using principal curves and neural networks,” in *Advances in Neural Networks - ISNN 2006. ISNN 2006. Lecture Notes in Computer Science*, Vol. 3972, eds J. Wang, Z. Yi, J. M. Zurada, B. L. Lu, and H. Yin (Berlin: Springer), doi: 10.4156/ijact.vol4.issue 2.268
- Zhu, X. (2008). Semi-supervised learning literature survey contents. *Sci. York* 10:2352.

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

Copyright © 2020 Cavedoni, Chirico, Pedrolì, Cipresso and Riva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Being in the Past and Perform the Future in a Virtual World: VR Applications to Assess and Enhance Episodic and Prospective Memory in Normal and Pathological Aging

Azzurra Rizzo<sup>†</sup>, Giuditta Gambino<sup>†</sup>, Pierangelo Sardo<sup>\*</sup> and Valerio Rizzo<sup>\*</sup>

Department of Biomedicine, Neuroscience and Advanced Diagnostic, Università Degli Studi di Palermo, Palermo, Italy

## OPEN ACCESS

### Edited by:

Soledad Ballesteros,  
National University of Distance  
Education (UNED), Spain

### Reviewed by:

Nicola Cellini,  
University of Padua, Italy  
José Manuel Reales,  
National University of Distance  
Education (UNED), Spain

### \*Correspondence:

Pierangelo Sardo  
pierangelo.sardo@unipa.it  
Valerio Rizzo  
valeriorizzophd@gmail.com

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Cognitive Neuroscience, a section of  
the journal  
Frontiers in Human Neuroscience

**Received:** 14 November 2019

**Accepted:** 03 July 2020

**Published:** 04 August 2020

### Citation:

Rizzo A, Gambino G, Sardo P and  
Rizzo V (2020) Being in the Past and  
Perform the Future in a Virtual World:  
VR Applications to Assess and  
Enhance Episodic and Prospective  
Memory in Normal and  
Pathological Aging.  
*Front. Hum. Neurosci.* 14:297.  
doi: 10.3389/fnhum.2020.00297

The process of aging commonly features a gradual deterioration in cognitive performance and, in particular, the decline of memory. Despite the increased longevity of the world's population, the prevalence of neurodegenerative conditions, such as dementia, continues to be a major burden on public health, and consequently, the latest research has been focused on memory and aging. Currently, the failure of episodic and Prospective memory (PM) is one of the main complaints in the elderly, considered among the early symptoms of dementia. It is therefore increasingly important to define more clearly the boundaries between normal and pathological aging. Recently, researchers have begun to build and apply Virtual Environments (VE) to the explicit purpose of better understanding the performance of episodic and PM in complex and realistic contexts, with the perspective of further developing effective training procedures that depend on reliable cognitive assessment methods. Virtual technology offers higher levels of realism than "pen and paper" testing and at the same time more experimental control than naturalistic settings. In this mini-review article, we examine the outcomes of recently available studies on virtual reality technology applications developed for the assessment and improvement of episodic and/or PM. To consider the latest technology, we selected 29 articles that have been published in the last 10 years. These documents show that VR-based technologies can provide a valid basis for screening and treatment and, through increased sensory stimulation and enriched environments reproducing the scenarios of everyday life, could represent effective stimulating experiences even in pathological aging.

**Keywords:** aging, pathological aging, virtual reality, episodic memory (EM), prospective memory (PM), assessment, cognitive training, cognitive impairment

## INTRODUCTION

Memory has been for centuries an intriguing field of brain research since it is a biologically essential function to the survival of almost all species (Bisaz et al., 2014). Memory is defined as the ability to acquire, process, store, and retrieve information (Fietta and Fietta, 2011). The remembering process is not a monolithic entity, but memory can be categorized and sub-categorized following many domains (Squire and Zola, 1996; Purves et al., 2001; Plescia et al., 2014). Conceivably due to the

increasing human life expectancy and the growing incidence of severe diseases that can induce neuronal degeneration and alter neuronal excitability (Carletti et al., 2016, 2017; Jaul and Barron, 2017; Park and Festini, 2017), memory has been considered a core feature to study upon normal and pathological aging processes. Most people report some early age-related memory impairments since the age of 60, especially in longitudinal studies (Nilsson, 2003; Rönnlund et al., 2005). Even earlier, by the age of 30, the decline of some cognitive functions was evidenced in cross-cutting studies (Park et al., 2002). Nevertheless, specific memory abilities that do not entail the conscious recollection of previously experienced material seem not to be altered with normal and pathological aging (Mitchell and Bruss, 2003; Ballesteros and Reales, 2004). Classically, aging has been linked to neuronal loss independently of the brain region involved (Coleman and Flood, 1987). However, research involving healthy adults indicates that normal aging is always associated with morphological alterations in neurons belonging to structures involved in cognition (Tisserand and Jolles, 2003). Not all cognitive abilities are affected by aging, but impaired memory skills are generally reported by the elderly and give rise to bitter complaints (Craik, 2008).

Memory has been classified by time direction (Maylor, 1993): retrospective memory refers to the ability to retrieve past information. Focusing on episodic memory (EM) refers to long-term memories including specific information such as time, location, or perceptual details as well as the connection of multidimensional information (Tulving, 2002). Currently, the alteration of EM is considered the main early symptom of dementia (Gold and Budson, 2008), though, it is also common in healthy aging, therefore its failure is not specific of pathological aging (Nilsson, 2003; Rönnlund et al., 2005; Craik, 2008). Prospective memory (PM), on the other hand, is the ability to remember to execute previously planned actions and can be defined as “remembering to remember,” thus referring to the future, for example when you have to remember to take a drug at a certain time. Some studies suggest that the onset of pathological aging determines more difficulty in PM (Huppert et al., 2000). However, potential errors in PM may be associated with considerable risks (Smith et al., 2000; Maylor et al., 2002), for example forgetting to turn off the gas.

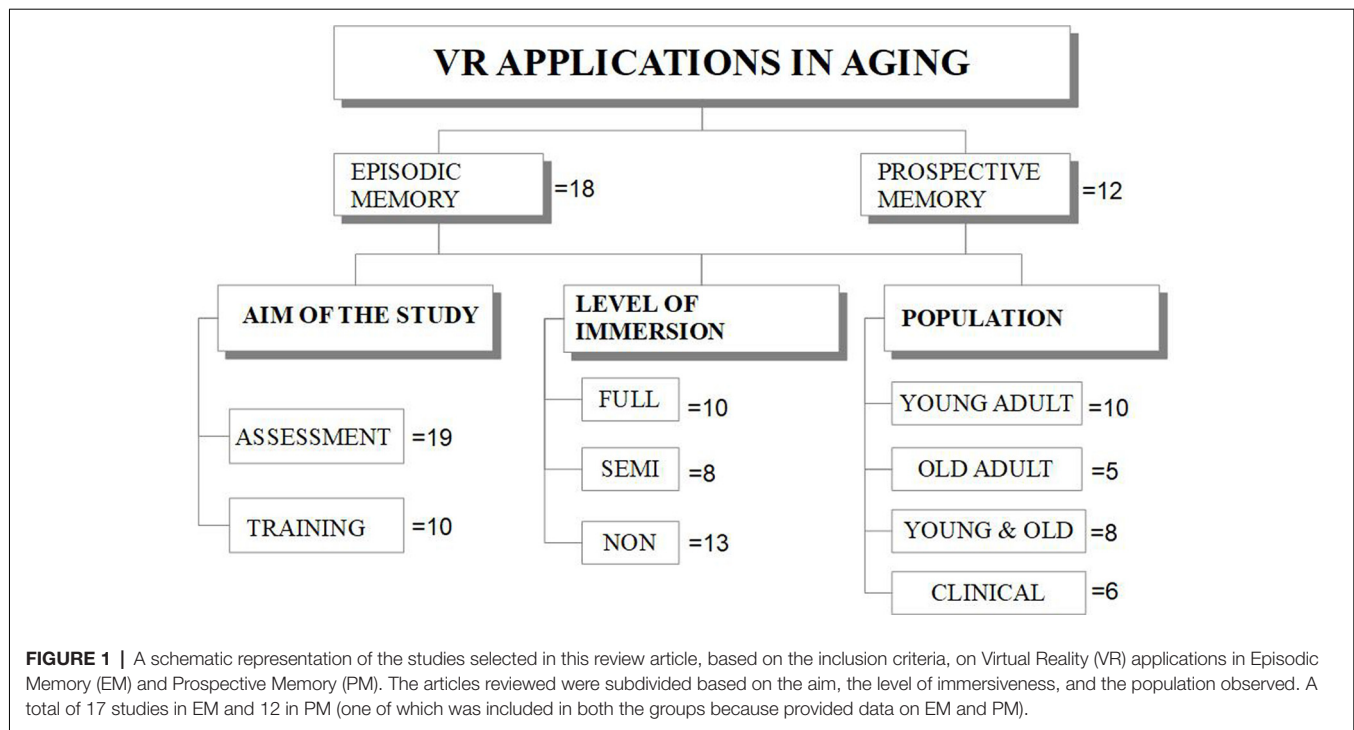
Considering this context, it is essential to differentiate normal and pathological aging, that is when aging brings about complications due to the presence of diseases such as Alzheimer’s disease (AD), Mild Cognitive Impairment (MCI), Parkinson’s Disease (PD), as well as other dementias (Hedden and Gabrieli, 2004; Craik, 2008) or diabetes type 2 (Redondo et al., 2016). Indeed, the progressive impairment in executive functions and memory processes in healthy adults could be exacerbated by the concomitant presence of common chronic diseases. Questions arise about if it is possible and how to inhibit pathological and non-pathological memory loss, especially considering the recent discovery of the ability of the nervous system to reconstruct cellular synapses upon interaction with enriched environments (Barak et al., 2013). A huge number of studies on the field have led to a better molecular understanding of different types of memory. Although, the most reliable way to assess

memory processes in normal and pathological aging is still intensely debated.

Classic memory tests using paper and pencil or computer systems for the evaluation of EM usually require older adults to remember static stimuli, therefore, they may not provide sufficient detail for predicting patients’ daily difficulties in different dynamic environments. Some studies have argued that neuropsychological assessments should provide a good degree of similarity to daily life tasks since the lack of ecological validity can negatively affect predictions about patient’s memory failures (Schultheis et al., 2002; Parsons and Rizzo, 2008). Episodic retrieval, for example, requires information about central and perceptual details, space-time contextual elements, and the binding of this multidimensional information (Abichou et al., 2019). Similarly, in everyday life, people present motivational aspects and adopt strategies for coding daily intentions that are difficult to probe through classic tests. Moreover, these tests measure memory components in isolation and failing to offer a comprehensive understanding of their operation (Tulving, 2002). Even naturalistic observation is not always an effective solution due to a series of difficulties including problems of standardization, control of the stimuli and distractors, economic costs to physically build the observation environments, as well as security problems.

In this regard, an outstanding advantage could be posed by the usage of virtual reality (VR) that evaluates memory consolidation by interacting with an enriched everyday environment, which guarantees both laboratory analytical control and precise assessments of how memory and other cognitive processes operate. The main goal of VR is to allow the patient to undertake specific tasks through artificial sensory stimulation and the illusion of being in an interactive environment perceived as a real place (Mantovani and Riva, 1999; Riva et al., 2007; LaValle, 2019). This experimental application could be reconducted to the heterogeneous family of Embodied Cognition theoretical approaches claiming that the physical properties of the human body, especially perceptual and motor systems, must be considered essential factors for the development and functioning of a cognitive system and could modulate learning and memory formation (Madan and Singhal, 2012).

The concepts of immersion and presence, related to VR, can better describe the experience from the user’s physical and psychological point of view. The immersion refers to the physical configuration of the interface of a VR application: the number and range of sensory and motors channels connected to the system determine the “immersiveness,” ranging from Non-Immersive (NI) systems on desktop computers to fully immersive systems. This distinction is based on how much the user can perceive the outside world during the virtual simulation (LaValle, 2019). The fully immersive types are indeed characterized by the use of a head-mounted display (HMD) in which a high-fidelity graphic screen is mounted in front of one’s eyes with separate lenses for each eye. The interaction in this type of virtual reality is controlled by tracking the movement of the head in combination with a computer system, therefore when users move their head to look around, they consequently move their visual field within the virtual environment 360 degrees. The



presence, on the other hand, is defined as “being in there” and occurs when the subject experiences an illusion of non-mediation in his space of action and acts as he would if the medium were not present (Mantovani and Riva, 1999; Riva et al., 2007). Presence is a subjective “response” to a system that has a certain level of immersion (Sanchez-Vives and Slater, 2005) people react and act on it as if they were real. This response is on many levels, ranging from unconscious physiological processes (cerebral, cardiac, skin, etc.) through to deliberate volitional behavior (Slater et al., 2009). In particular, the ability to induce the sense of presence seems to have positive effects on attention and involvement and consequently is very relevant in the evaluation of memory (Sutcliffe et al., 2005; Makowski et al., 2017).

In this review article, we aimed to explore the most recent evidence about the ability to evaluate EM and PM, as well as to stimulate improvements in both normal and pathological aging, through representative and significant examples of applications in VR.

## METHODOLOGY

Initially, a systematic online bibliography search was carried out through the following profile databases: Web of Knowledge, ScienceDirect, PubMed, and Google Scholar, on the date of June 2019. We used the following core search terms and their combinations: VR or virtual environment, prospective memory or PM, EM or EM; and the following as additional search terms with “Xor” combinations: assessment, cognitive training, aging, pathological aging, cognitive impairment. Also, to get a broader and more complete view of the topic we have included studies on young adult subjects. Second, a selection of relevant articles was

limited to the period 2009–2019 to obtain information mainly about outcomes referring to the latest technology.

The references selected were included in the review in case the following criteria were met as shown in **Figure 1**: research on the impairment of EM or PM in the aging; description of VR methodologies for assessment or training; a clear description of VR tools that determine the related level of immersion. Overall, 29 studies have been identified and summarized in this mini-review article and classified as in **Table 1**.

## VR FOR EPISODIC MEMORY IN AGING

Within Virtual Environments (VE), participants can be immersed in scenarios that represent different everyday situations such as virtual apartments (Sauzéon et al., 2012), grocery stores (Parsons and Barnett, 2017; Plechatá et al., 2019; Corriveau-Lecavalier et al., 2020) or city (Plancher et al., 2012, 2013, 2018; Abichou et al., 2019). This gives the chance to implement simple tasks to assess the versatile nature of EM in ecological situations in a rich and specific space-time context.

Plancher et al. (2018) analyzed the role of working memory (WM) while building an episodic trace, through an on-screen projected urban virtual environment. They reported that the memory of central information was altered by simultaneous tasks and that the memory of the temporal context and binding was compromised only upon the performance of a competing visuospatial activity. To the purpose of testing WM’s key role in consolidating the EM, participants were asked to explore the environment using a steering wheel, a gas pedal, and a brake pedal. At the same time, a secondary numerical task interfering with the phonological cycle (e.g., storing the number

**TABLE 1 |** Virtual reality (VR) studies in aging classified according to contribution in Episodic Memory (EM) or Prospective Memory (PM), kind of intended purpose divided into Assessment (AS) or Training (TR), level of immersion divided into Full-Immersive (FI) or Semi-Immersive (SI) or Non-Immersive (NI); Experimental subjects (ES); the presence of Information about Cybersickness (IC); type of Environment; the number of Training Session (TS); navigation Time and Main results.

	Articles included	FI	SI	NI	AS	TR	ES	IC	Environment	TS	Time	Main results
<b>EM</b>	Sauzéon et al. (2012)						Young adults	X	Apartament		N/A	Active exploration increases hippocampal activity and improves spatial learning
	Plancher et al. (2012)						Older adults and AD patients					Interaction and planning affect spatial memory/worst memory performance for factual information
	Plancher et al. (2013)		X		X		Young adults		Urban		N/A	Interaction and planning affect spatial memory/worst memory performance for factual information
	Taillade et al. (2013)						Young and healthy older adults	X	Urban		10–15'	Worse spatial learning in motor control associated with executive functions
	Jebara et al. (2014)						Young and healthy older adults		Urban		N/A	Better recall of prospective intentions rather than retrospective in mild AD
	Plancher et al. (2018)						Young adults		Urban		20–25'	Concurrent activity (WM) negative impacts on long-term memory of central information
	Clemenson and Stark (2015)						Young adults		Angry bird Game; Super Mario 3d world	28	N/A	3D spatial aspect improves hippocampal dependent behavior
	Makowski et al. (2017)						Young adults		Avangers movie	1	N/A	Emotions and higher levels of presence associated with better memory
	Serino et al. (2017)		X		X		Older adults and AD patients		Urban	10	20'	Spatial performance improvement in AD/spatial decision making improvement in healthy subjects
	West et al. (2017)						Healthy older adults		Super Mario 64; Super Mario Galaxy	120	30'	3d spatial aspect improves hippocampaldependent behavior
	Bakdash et al. (2008)						Young adults		Urban		20'	Decision-making positively affect EM encoded vs. control
	Bergouignan et al. (2014)						Young adults		Room		N/A	Depersonalization states compromise the coding of EM
	Parsons and Barnett (2017)		X		X		Young and healthy older adults	X	Grocery store		15–20'	Virtual Shop construct validity

(Continued)



TABLE 1 | Continued

	Articles included	FI	SI	NI	AS	TR	ES	IC	Environment	TS	Time	Main results
PM	Corriveau-Lecavalier et al. (2020)						Young and healthy older adults	X	Grocery store		N/A	Construct validity of the virtual shop/higher levels of motivation than traditional memory test
	Optale et al. (2010)	X				X	Healthy older adults		Familiar setting (family home, park, ecc)	36 + 24	15'	Improvement in general cognitive functioning and in long-term memory
	Serino et al. (2015)			X	X		Healthy older adults and AD/aMCI patients		Room		N/A	Deficiencies of aMCI and AD patients in storing hetero-centric independent representational memories/deficiencies of AD patients in synchrony
	Abichou et al. (2019)						Young and healthy older adults	X	Urban		N/A	EM consolidation after a period of sleep
	Plechata et al. (2019)	X		X	X		Young and healthy older adults	X	Supermarket		10'	Elderly lower performance in HMD
	Nolin et al. (2013)						MCI patients and older adults		Apartment		N/A	Positive correlation between PM Virtual task and MoCA test
	Dong et al. (2017)						Young adults		Shopping streets		35'	Better correlation of virtual test with PM daily memory compared to desktop test
	Parsons and Barnett (2017)		X		X		Young and healthy older adults	X	Grocery store		15–20'	Virtual Shop construct validity
	Ouellet et al. (2018)						Young and healthy older adults	X	Grocery store		15'	Ecological and construct validity/Greater difficulty in virtual pointer technique in older adults compared to young adult
	Trawley et al. (2014)						Young adults		Office building	1	12'	Cue saliency increases attentional load on PM
	Yip and Man (2013)						ABI patients	X	Urban	12	35- 40'	Improvements in immediate recall of PM tasks performed by participants
	Debarnot et al. (2015)			X		X	Healthy older adults		Urban	1	30'	Improvement of the excitatory stimulation (Itbs) of Frontopolar cortex in event-based PM
	Mioni et al. (2015)						PD patients and older adults		Virtual day	2	15–20'	Emotionally-related improvement in PM performance

(Continued)

TABLE 1 | Continued

Articles included	FI	SI	NI	AS	TR	ES	IC	Environment	TS	Time	Main results
Rose et al. (2015)						Healthy older adults		Virtual day	20	40–60'	Better scores in performing realistic PM tasks and IADL
Gonneaud et al. (2014)		X		X		Healthy older adults		Urban		N/A	Link-EB task induce better performance than no link-EB and TB tasks
Lecouvey et al. (2019)						Older adults and mild AD patients					Better recall of prospective vs. retrospective components of intentions in AD patients
Dong et al. (2016)	X		X	X		Young adults		Shopping streets		N/A	Immersive VR induces prefrontal cortex activity in the BA10

of garbage containers in the path) and a secondary visuospatial task (e.g., memorizing the spatial model of containers) were applied to predict that secondary activities performed during learning would interfere with coding and resulted in altered memory performances.

One of the main advantages of VEs is, indeed, the ability to precisely model and control the environment itself according to the requirements decided by the experimenter, avoiding possible problems of building real scenarios (Sauzéon et al., 2012). Due to the extreme adaptability of this technique, EM in VE has already been tested in clinical contexts (Plancher et al., 2012; García-Betances et al., 2015; Serino et al., 2015, 2017). To compare VR memory tasks with traditional neuropsychological tools for its evaluation, Plancher et al. (2012) conducted a study on healthy participants, patients with amnesic MCI (aMCI) and with mild Alzheimer's. The experimental groups were asked to store as much information as possible during active and passive browsing conditions. The virtual task allowed characterizing the different cognitive profiles of the three populations and the authors found that spatial allocentric memory assessments discriminated against patients with aMCI from controls. Nevertheless, after active exploration of the VE, all participants, including patients with aMCI and AD, retrieved significantly better both central and allocentric spatial information and the process of binding. As pointed out by the authors, these results about active exploration are particularly promising because they provide support for the feasibility of VR as an effective non-pharmacological tool to promote neuroplasticity and neural reorganization in patients with AD.

Preclinical studies on aging have shown that immersion in enriched environments may drive long-term enhancement of the activity of the hippocampus and changes in memory-associated brain regions inducing structural changes in animal models (Harvey et al., 2009). Clemenson and Stark (2015) discovered that young adults trained with Super Mario 3D showed better spatial and EM performance dependent on hippocampus activity, compared to people trained in a 2D-controlled game. More recently, West et al. (2017) proposed the same 3D platform training applied to an elderly population reporting increases in gray matter thickness in brain regions known to be implicated in cognitive-related decline. Also, it was suggested that a greater feeling of presence improves the effectiveness of VR applications (Optale et al., 2010). It was found that higher levels of presence were associated with better factual memory and the impact of the emotional stimulus was mediated by a sense of presence (Makowski et al., 2017).

As pointed out by Repetto et al. (2016), besides environmental enrichment in VR research, two main aspects may be advantageous in the context of EM's study. The first is that VR allows exploration from an egocentric point of view (Bergouignan et al., 2014; Serino et al., 2015, 2017); i.e., Bergouignan et al. (2014), using an out-of-body-induced illusion, have reported that an accurate EM encoding is favored by the perception of the world from the perspective of one's own body. Second, VR enables active exploration of the environment. However, comparisons of active and passive navigation showed contradictory results with both negative (Taillade et al., 2013)

and positive effects of active navigation (Sauzéon et al., 2012; Plancher et al., 2013).

Employing a similar paradigm to Plancher et al. (2013), Jebara et al. (2014) assessed the performance of a sample population of young adults and seniors in a virtual city projected on a screen. Four interaction conditions were included: “passive” (passengers in a virtual car cannot choose directions and route), “itinerary control” (passengers can choose), “low control” (driver move the car on rails) and “high control” (driver choose also direction). Better scores in EM (*what—where—when and binding*) in both young and old groups were obtained only in the conditions of choosing the route and low navigation control. This suggests that EM performance benefits from multimodal coding, through the enrichment of motor interaction and that it has been improved by active navigation when it is not too expensive in terms of attention efforts. According to some authors (Bakdash et al., 2008; Sauzéon et al., 2012; Jebara et al., 2014), active navigation may require additional cognitive resources that are not fully available for the coding process. Consequently, inconsistent results on memory performance related to the active-passive navigation may be due to differences in the manipulation of sensorimotor stimulation and its confusing effects on cognitive activity, as shown by several studies reporting worse memory performance caused by split attention (e.g., Craik et al., 1996). Plancher et al., 2013, have shown that driving in a VR while encoding information can be considered as a double task in which motor control can impact factual memory.

However, although these results suggest that active navigation VR training may have a beneficial effect on EM, it should be noted that older adults perform worse than young people, particularly in binding scores. This age-related effect noted in the low control condition encourages greater attention from research on the elderly regarding the complexity of the motor task which would risk having diametrically opposite effects on memory.

Lastly, as the use of HMD spreads, the effects of active and passive navigation must also be investigated in fully immersive environments especially for elder persons who are not familiar with technology and not used to handle it.

## VR FOR PROSPECTIVE MEMORY IN AGING

A growing number of studies on memory have focused on its prospective side, but uncertainties remain regarding the characteristics of PM impairments. It is still to be fully unveiled the influence of the executive functions, the life-span development of prospective remembering and the age effects, the underlying mechanism involved in event-based or in time-based PM task and the role of motivational aspect (Kliegel and Martin, 2003). Compared to traditional laboratory paradigms, virtual reality creates realistic tasks for the evaluation of the PM, increasing the variety of possible actions to perform. This allows measurements of multiple cognitive processes involved and to systematically control interactive stimuli with immediate feedback on performance through sensory modalities. Nolin et al. (2013) exploited an eMagin Z800 immersion target on a population of older adults with MCI vs. healthy controls

in an urban environment. This VR-based evaluative approach in PM tasks could be more sensitive to the effects of MCI than traditional neuropsychological ones such as the Rivermead Behavioral Memory Test (RBMT; Wilson et al., 1985). Indeed, although RBMT was widely used in clinical settings, it does not include enough PM tasks to generate many types of performances and does not assess time-based PM performances (Mioni et al., 2014). VR, reaching a higher level of complexity, requires more cognitive resources to perform tasks, therefore could better represent the person in real life.

Based on this assumption, VEs are used to explore central theoretical questions about how the cognitive system successfully codifies and recovers intentional behavior (Gonneaud et al., 2014; Trawley et al., 2014). Gonneaud et al. (2014) assessed the impact of connections between the potential component (PC; remembering that something needs to be done) and the retention component (RC; the content of intention) of PM, in a Semi-Immersive (SI) urban environment where subjects could navigate using a virtual car. More specifically, the link between PC and RC affects the distinction between PM based on the appearance of an external cue (EB) and based on the automatic start of intention after a time interval (TB). Nine tasks were presented to the subjects: with a clear link between PC and RC (Link-EB; e.g., buying stamp book at a post office) or without (noLink-EB; e.g., buying eyeglasses at the fountain). Link-EB produced better performance than noLink-EB and TB, highlighting the importance of the association processes between PC and RC for effective PM. Similarly, Lecouvey et al. (2019) explored in VR the effects of mild AD on PM, showing that both the PM components are significantly compromised in AD patients, but RCs of intentions are altered before PCs. These data supported the hypothesis that early impairments of EM have a great impact on the execution of PM tasks in AD.

This VR approach is a more realistic tool that could help to better highlight planning processes, motivational aspects, time estimation, or eventual difficulties in dual-task processes in PM impairments in everyday life (Gonneaud et al., 2014; Lecouvey et al., 2019). Also, this approach could be useful to provide more efficient therapeutic interventions (Meijer et al., 2009) and a better measure of training effectiveness concerning less naturalistic performance. Indeed, VR technology has also been applied to cognitive training, such as in the paradigm of “Virtual Week” (Yip and Man, 2013; Rose et al., 2015). Mioni et al. (2015) outlined that VR improved PM performance in PD patients for the first time by using emotionally enriched tasks. Participants were asked to remember to carry out actions with positive value (e.g., “tell Roberta that Maria had a baby when you talk to Roberta”); with a negative value (e.g., “Pay a fine for speeding when you go shopping”) or with neutral value (e.g., “Buy your bus ticket after breakfast”). The tasks of PM with positive emotional value showed better results than tasks with negative or neutral value in both normal and PD patients, although the latter showed worse performance than the control group, independently of the emotional valence of the cue. Experimental data also provided improved outcomes in remembering to perform tasks with pleasant content compared to neutrals, since positive stimuli may attract more attention

resources, hence facilitating the recovery and execution of PM actions. Furthermore, results seem to indicate that the use of a fully immersive task is feasible in the elderly: it arouses presence, it is addictive and causes limited symptoms of illness (Ouellet et al., 2018; Corriveau-Lecavalier et al., 2020).

However, it is important to note that due to technological limitations, state of art immersive environments do not correspond exactly to the real world. This contributes to the manifestation, in some users, of symptoms similar to those of the classic motion sickness called cybersickness, resulting from the conflict between the visual, vestibular, and proprioceptive sensory systems. Factors such as the previous familiarity with technology, age, or the presence of diseases can play an important role but, in particular, among older people, factors such as rotational speed and duration of exposure seem to increase cybersickness (Liu, 2014).

In the literature reviewed here, visual stimuli are presented mainly through SI systems, nevertheless, several studies applied fully immersive HMD systems (Nolin et al., 2013; Parsons and Barnett, 2017; Ouellet et al., 2018). A recent study conducted by Dong et al. (2016) compared desktop monitor activity and immersive VR activity with Oculus Rift to investigate whether traditional lab PM activity produced a similar result to PM activity in the VR environment. It was reported that while the performance of standard computer monitor tasks does not significantly correlate to the VR scores, a negative correlation between desktop reaction times and VR scores was observed, as well as a positive correlation between VR and desktop reaction times. VR seems to be more sensitive in accurately assessing PM in life. This is because the slide-based activity requires fewer cognitive resources than VR, where participants' cognitive load has been increased and more correctly identifies worse scores when reaction times are slower. Notwithstanding, as pointed out by Plechatá et al. (2019), research should provide further clarification on the comparison between desktop platform performance and those in HMD, since different results in older people may be because participants have already had previous experiences with desktops but not with HMD platform that would lead to increased fatigue. Consequently, under certain circumstances, the level of immersion and a more complex context could be a problem for memory ability and participants may have limited cognitive resources for the memory task. Structural features such as movement or sensory feedback of the virtual environment can involve participants' attention by draining cognitive energy because older adults unfamiliar with technology may find their management frustrating and this may distract from their virtual reality experience, making the task more demanding than for young adults.

Nonetheless, VR immersiveness is essential for exploiting the procedural involvement that allows us to predict the cognitive behavior of subjects as realistically as possible. In many of the described experimental environments, exploration took place while sitting, but people must navigate in the same way as they do in the real world (Tieri et al., 2018). The need arises from the fact that immersive VR navigation offers the user a much wider range of movement to approach or

physically realistically interact with the virtual world, while the sitting position requires a set visual height, longer movements, and controller-based environment navigation. This freedom of movement is particularly problematic for VR-based PM studies that combine neuroimaging techniques (Dong et al., 2016) to ascertain changes in brain hemodynamic responses, and neuromodulation to improve PM performance in senior subjects (Debarnot et al., 2015).

Neurophysiological changes associated with VR neurorehabilitation can be measured using non-invasive and portable neuroimaging techniques, including fNIRS and/or EEG, equipped with a neuroergonomic and wireless approach, to measure cerebral blood flow in real-time during VR activity. Recently a study by Dong et al. (2017) investigated the function of the prefrontal cortex during a PM activity in an immersive VR environment *via* Oculus Rift and an OEG-16 multi-channel fNIRS system that allowed solving the problems often present for EEG such as the difficulty of movement for the application of electrodes. By using a virtual shopping experience, this study provided early confirmation of Brodmann area activation in a PM activity in VR but further studies are still needed to evaluate if and which other areas could potentially be involved during memory tasks in VE.

## CONCLUSIONS

The present review article provides a snapshot of virtual reality technology applications developed for assessment and improvement of episodic and/or PM, with the idea to suggest the integration of the most recent technological advancements to cognitive and aging neuroscience. It could be considered a flaw in our choice to also include studies employing young adults, though not strictly representative of the aging process. This inclusion, however, is based on a conceptual approach that recognizes that the study of aging must cover both older and younger populations trying to bridge what has been called a gap in geriatric research (Moffitt et al., 2017). The goal is to select additional evidence that currently cannot be obtained with adult subjects.

More importantly, it should be noted that the literature here reviewed included studies that used any form of VR technology, including non, semi, or totally immersive. It is important to note that the keywords of these reviews included the term “virtual reality” although this differs in terms of technical qualities and level of ecological validity. We believe that recognizing the centrality of the nomenclature in this field and the need for greater uniformity of language will create a more coherent and connected research field.

Also, the implementation of clinical VR research outside the laboratory still presents significant challenges that need to be addressed. The type of VR technology and the experimental designs implemented vary greatly between studies with many approaches using completely different hardware, software, or paradigms. However, the positive results provide reasons for more rigorously controlled research, necessary to progress from feasibility studies and pilot tests to standardized protocols that can be shared by the research community.



The literature here reviewed suggests that VR protocols offer an additional tool and excellent opportunity for innovative assessment and training options, particularly important in early identification of the subtle amnesic deficits that usually elude traditional methods. Some of the related advantages are the easy adaptability and the ability to replicate ecologically valid environments present in everyday life, allowing precise measurements of the cognitive processes involved. Furthermore, the possibility of providing a more stimulating context than in traditional laboratories can generate positive motivation in the elderly. However, the possible limitations associated with the perception of VR technology must be taken into account. The results of higher-immersive studies or greater interactivity are inconclusive in terms of the benefits of evaluating or training in the elderly population, particularly in pathological aging. Also, the introduction of the clinical population into VEs raises particular ethical and safety problems: some users experience health problems associated with the use of immersive HMD though these effects are mild and quickly fade. Susceptibility

to cybersickness appears to be limited, but it could be related to short exposure times. Consequently, an in-depth research is needed to investigate how aging can affect motion sickness caused by immersive environments, to avoid the risk of reducing rather than increasing the ecological validity.

Finally, it is interesting to note that VR technology can be easily combined with other forms of technologies such as neuromodulation (tDCS) and neuroimaging (fNIRS/EEG) that can be considered valuable and indispensable tools to increase the benefits of VR. This perspective could provide a more targeted approach to neuro-training and will be the core of future research on the field.

## AUTHOR CONTRIBUTIONS

AR and GG conducted the search and selection of bibliography. VR designed and directed the project. AR, GG, PS, and VR wrote the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Abichou, K., La Corte, V., Hubert, N., Orriols, E., Gaston-Bellegarde, A., Nicolas, S., et al. (2019). Young and older adults benefit from sleep, but not from active wakefulness for memory consolidation of what-where-when naturalistic events. *Front. Aging Neurosci.* 11:58. doi: 10.3389/fnagi.2019.00058
- Bakdash, J. Z., Linkenauger, S. A., and Itt, D. P. (2008). Comparing decision-making and control for learning a virtual environment: Backseat drivers learn where they are going. *Proc. Hum. Factors Ergon. Soc.* 52, 2117–2121. doi: 10.1177/154193120805202707
- Ballesteros, S., and Reales, J. M. (2004). Intact haptic priming in normal aging and Alzheimer's disease: evidence for dissociable memory systems. *Neuropsychologia* 42, 1063–1070. doi: 10.1016/j.neuropsychologia.2003.12.008
- Barak, B., Shvarts-Serebro, I., Modai, S., Gilam, A., Okun, E., Michaelson, D. M., et al. (2013). Opposing actions of environmental enrichment and Alzheimer's disease on the expression of hippocampal microRNAs in mouse models. *Transl. Psychiatry* 3:e304. doi: 10.1038/tp.2013.77
- Bergouignan, L., Nyberg, L., and Ehrsson, H. H. (2014). Out-of-body-induced hippocampal amnesia. *Proc. Natl. Acad. Sci. U S A* 111, 4421–4426. doi: 10.1073/pnas.1318801111
- Bisaz, R., Travaglia, A., and Alberini, C. M. (2014). The neurobiological bases of memory formation: from physiological conditions to psychopathology. *Psychopathology* 47, 347–356. doi: 10.1159/000363702
- Carletti, F., Gambino, G., Rizzo, V., Ferraro, G., and Sardo, P. (2017). Neuronal nitric oxide synthase is involved in CB/TRPV1 signalling: focus on control of hippocampal hyperexcitability. *Epilepsy Res.* 138, 18–25. doi: 10.1016/j.epilepsyres.2017.09.018
- Carletti, F., Sardo, P., Gambino, G., Liu, X. A., Ferraro, G., and Rizzo, V. (2016). Hippocampal hyperexcitability is modulated by microtubule-active agent: evidence from *in vivo* and *in vitro* epilepsy models in the rat. *Front. Cell. Neurosci.* 10:29. doi: 10.3389/fncel.2016.00029
- Clemenson, G. D., and Stark, C. E. L. (2015). Virtual environmental enrichment through video games improves hippocampal-associated memory. *J. Neurosci.* 35, 16116–16125. doi: 10.1523/JNEUROSCI.2580-15.2015
- Coleman, P. D., and Flood, D. G. (1987). Neuron numbers and dendritic extent in normal aging and Alzheimer's disease. *Neurobiol. Aging* 8, 521–545. doi: 10.1016/0197-4580(87)90127-8
- Corriveau-Lecavalier, N., Ouellet, É., Boller, B., and Belleville, S. (2020). Use of immersive virtual reality to assess episodic memory: a validation study in older adults. *Neuropsychol. Rehabil.* 30, 462–480. doi: 10.1080/09602011.2018.1477684
- Craik, F. I. M. (2008). Memory changes in normal and pathological aging. *Can. J. Psychiatry* 53, 343–345. doi: 10.1177/070674370805300601
- Craik, F. I., Govoni, R., Naveh-Benjamin, M., and Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.* 125, 159–180. doi: 10.1037/0096-3445.125.2.159
- Debarnot, U., Crépon, B., Orriols, E., Abram, M., Charron, S., Lion, S., et al. (2015). Intermittent theta burst stimulation over left BA10 enhances virtual reality-based prospective memory in healthy aged subjects. *Neurobiol. Aging* 36, 2360–2369. doi: 10.1016/j.neurobiolaging.2015.05.001
- Dong, D., Wong, L. K. F., and Luo, Z. (2017). Assessment of prospective memory using fNIRS in immersive virtual reality environment. *J. Behav. Brain Sci.* 07, 247–258. doi: 10.4236/jbbs.2017.76018
- Dong, D., Wong, L. K., Luo, Z., and Quan, C. (2016). “A novel approach for assessing prospective memory using immersive virtual reality task,” in *Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, (Qingdao: IEEE), 1888–1893.
- García-Betances, R. I., Arredondo Waldmeyer, M. T., Fico, G., and Cabrera-Umpiérrez, M. F. (2015). A succinct overview of virtual reality technology use in Alzheimer's disease. *Front. Aging Neurosci.* 7:235. doi: 10.3389/fnagi.2015.00235
- Gold, C. A., and Budson, A. E. (2008). Memory loss in Alzheimer's disease: implications for development of therapeutics. *Expert Rev. Neurother.* 8, 1879–1891. doi: 10.1586/14737175.8.12.1879
- Gonneaud, J., Piolino, P., Lecouvey, G., Madeleine, S., Orriols, E., Fleury, P., et al. (2014). “Assessing prospective memory in young healthy adults using virtual reality,” in *Disability Studies. Virtual Reality: People With Special Needs*, eds P. M. Sharkey and J. Merrick (New York, NY: Nova Science Publishers), 107–120.
- Fietta, P., and Fietta, P. (2011). The neurobiology of the human memory. *Theor. Biol. Forum* 104, 69–87.
- Harvey, C. D., Collman, F., Dombeck, D. A., and Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* 461, 941–946. doi: 10.1038/nature08499
- Hedden, T., and Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci.* 5, 87–96. doi: 10.1038/nrn1323
- Huppert, F. A., Johnson, T., and Nickson, J. (2000). High prevalence of prospective memory impairment in the elderly and in early-stage dementia: findings from a population-based study. *Appl. Cogn. Psychol.* 14, S63–S81. doi: 10.1002/acp.771
- Jaul, E., and Barron, J. (2017). Age-related diseases and clinical and public health implications for the 85 years old and over population. *Front. Public Heal.* 5:335. doi: 10.3389/fpubh.2017.00335
- Jebara, N., Orriols, E., Zaoui, M., Berthoz, A., and Piolino, P. (2014). Effects of enactment in episodic memory: a pilot virtual reality study with young and elderly adults. *Front. Aging Neurosci.* 6:338. doi: 10.3389/fnagi.2014.00338

- Kliegel, M., and Martin, M. (2003). Prospective memory research: why is it relevant? *Int. J. Psychol.* 38, 193–194. doi: 10.1080/00207590344000114
- LaValle. (2019). *Virtual Reality*. Available online at: <http://vr.cs.uiuc.edu/>. Accessed December 2019
- Lecouvey, G., Morand, A., Gonneaud, J., Piolino, P., Orriols, E., Pélerin, A., et al. (2019). An impairment of prospective memory in mild Alzheimer's disease: a ride in a virtual town. *Front. Psychol.* 10:241. doi: 10.3389/fpsyg.2019.00241
- Liu, C. L. (2014). A study of detecting and combating cybersickness with fuzzy control for the elderly within 3D virtual stores. *Int. J. Hum. Comput. Stud.* 72, 796–804. doi: 10.1016/j.ijhcs.2014.07.002
- Madan, C. R., and Singhal, A. (2012). Using actions to enhance memory: effects of enactment, gestures, and exercise on human memory. *Front. Psychol.* 3:507. doi: 10.3389/fpsyg.2012.00507
- Makowski, D., Sperduti, M., Nicolas, S., and Piolino, P. (2017). "Being there" and remembering it: presence improves memory encoding. *Conscious. Cogn.* 53, 194–202. doi: 10.1016/j.concog.2017.06.015
- Mantovani, G., and Riva, G. (1999). "Real" presence: how different ontologies generate different criteria for presence, telepresence, and virtual presence. *Presence Teleoperators Virtual Environ.* 8, 540–550. doi: 10.1162/105474699566459
- Maylor, E. A. (1993). Aging and forgetting in prospective and retrospective memory tasks. *Psychol. Aging* 8, 420–428. doi: 10.1037/0882-7974.8.3.420
- Maylor, E. A., Smith, G., Della Sala, S., and Logie, R. H. (2002). Prospective and retrospective memory in normal aging and dementia: an experimental study. *Mem. Cogn.* 30, 871–884. doi: 10.3758/bf03195773
- Meijer, F., Geudeke, B. L., and van den Broek, E. L. (2009). Navigating through virtual environments: visual realism improves spatial cognition. *Cyberpsychol. Behav.* 12, 517–521. doi: 10.1089/cpb.2009.0053
- Mioni, G., Meligrana, L., Rendell, P. G., Bartolomei, L., Perini, F., and Stablum, F. (2015). Event-based prospective memory in patients with Parkinson's disease: the effect of emotional valence. *Front. Hum. Neurosci.* 9:427. doi: 10.3389/fnhum.2015.00427
- Mioni, G., McClintock, S. M., and Stablum, F. (2014). "Understanding, assessing and treating prospective memory dysfunctions in traumatic brain injury patients," in *Traumatic Brain Injury*, eds F. Sadaka and T. Quinn (Rijeka, Croatia: InTeck), 401–436.
- Mitchell, D. B., and Bruss, P. J. (2003). Age differences in implicit memory: conceptual, perceptual, or methodological? *Psychol. Aging* 18, 807–822. doi: 10.1037/0882-7974.18.4.807
- Moffitt, T. E., Belsky, D. W., Danese, A., Poulton, R., and Caspi, A. (2017). The longitudinal study of aging in human young adults: knowledge gaps and research agenda. *J. Gerontol. A. Biol. Sci. Med. Sci.* 72, 210–215. doi: 10.1093/gerona/glw191
- Nilsson, L. G. (2003). Memory function in normal aging. *Acta Neurol. Scand. Suppl.* 179, 7–13. doi: 10.1034/j.1600-0404.107.s179.5.x
- Nolin, P., Banville, F., Cloutier, J., and Allain, P. (2013). Virtual reality as a new approach to assess cognitive decline in the elderly. *Acad. J. Interdiscip. Stud.* 2, 612–616. doi: 10.5901/ajis.2013.v2n8p612
- Optale, G., Urgesi, C., Busato, V., Marin, S., Piron, L., Priftis, K., et al. (2010). Controlling memory impairment in elderly adults using virtual reality memory training: a randomized controlled pilot study. *Neurorehabil. Neural Repair* 24, 348–357. doi: 10.1177/1545968309353328
- Ouellet, É., Boller, B., Corriveau-Lecavalier, N., Cloutier, S., and 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. doi: 10.1016/j.jneumeth.2018.03.010
- Park, D. C., and Festini, S. B. (2017). Theories of memory and aging: a look at the past and a glimpse of the future. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 72, 82–90. doi: 10.1093/geronb/gbw066
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., and Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17, 299–320. doi: 10.1037/0882-7974.17.2.299
- Parsons, T. D., and Barnett, M. (2017). Validity of a newly developed measure of memory: feasibility study of the virtual environment grocery store. *J. Alzheimers Dis.* 59, 1227–1235. doi: 10.3233/jad-170295
- Parsons, T. D., and Rizzo, A. A. (2008). Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. *Cyberpsychol. Behav.* 11, 17–25. doi: 10.1089/cpb.2007.9934
- Plancher, G., Barra, J., Orriols, E., and Piolino, P. (2013). The influence of action on episodic memory: a virtual reality study. *Q. J. Exp. Psychol.* 66, 895–909. doi: 10.1080/17470218.2012.722657
- Plancher, G., Gyselinck, V., and Piolino, P. (2018). The integration of realistic episodic memories relies on different working memory processes: evidence from virtual navigation. *Front. Psychol.* 9:471. doi: 10.3389/fpsyg.2018.00047
- Plancher, G., Tirard, A., Gyselinck, V., Nicolas, S., and Piolino, P. (2012). Using virtual reality to characterize episodic memory profiles in amnesic mild cognitive impairment and Alzheimer's disease: influence of active and passive encoding. *Neuropsychologia* 50, 592–602. doi: 10.1016/j.neuropsychologia.2011.12.013
- Plechátá, A., Sahula, V., Fayette, D., and Fajnerová, I. (2019). Age-related differences with immersive and non-immersive virtual reality in memory assessment. *Front. Psychol.* 10:1330. doi: 10.3389/fpsyg.2019.01330
- Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., et al. (2001). *Neuroscience*. 2nd Edn. Sunderland, MA: Sinauer Associates.
- Plescia, F., Marino, R. A. M., Navarra, M., Gambino, G., Brancato, A., Sardo, P., et al. (2014). Early handling effect on female rat spatial and non-spatial learning and memory. *Behav. Processes* 103, 9–16. doi: 10.1016/j.beproc.2013.10.011
- Redondo, M. T., Beltrán-Brotóns, J. L., Reales, J. M., and Ballesteros, S. (2016). Executive functions in patients with Alzheimer's disease, type 2 diabetes mellitus patients and cognitively healthy older adults. *Exp. Gerontol.* 83, 47–55. doi: 10.1016/j.exger.2016.07.013
- Repetto, C., Serino, S., Macedonia, M., and Riva, G. (2016). Virtual reality as an embodied tool to enhance episodic memory in elderly. *Front. Psychol.* 7:1839. doi: 10.3389/fpsyg.2016.01839
- Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., et al. (2007). Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychol. Behav.* 10, 45–56. doi: 10.1089/cpb.2006.9993
- Rönnlund, M., Nyberg, L., Bäckman, L., and Nilsson, L. G. (2005). Stability, growth and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychol. Aging* 20, 3–18. doi: 10.1037/0882-7974.20.1.3
- Rose, N. S., Rendell, P. G., Hering, A., Kliegel, M., Bidelman, G. M., and Craik, F. I. M. (2015). Cognitive and neural plasticity in older adults' prospective memory following training with the virtual week computer game. *Front. Hum. Neurosci.* 9:592. doi: 10.3389/fnhum.2015.00592
- Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi: 10.1038/nrn1651
- Sauzéon, H., Pala, P. A., Larrue, F., Wallet, G., Déjos, M., Zheng, X., et al. (2012). The use of virtual reality for episodic memory assessment: effects of active navigation. *Exp. Psychol.* 59, 99–108. doi: 10.1027/1618-3169/a000131
- Schultheis, M. T., Himelstein, J., and Rizzo, A. A. (2002). Virtual reality and neuropsychology: upgrading the current tools. *J. Head Trauma Rehabil.* 17, 378–394. doi: 10.1097/00001199-200210000-00002
- Serino, S., Morganti, F., Di Stefano, F., and Riva, G. (2015). Detecting early egocentric and allocentric impairments deficits in Alzheimer's disease: an experimental study with virtual reality. *Front. Aging Neurosci.* 7:1. doi: 10.3389/fnagi.2015.00088
- Serino, S., Pedrolí, E., Tuena, C., De Leo, G., Stramba-Badiale, M., Goulene, K., et al. (2017). A novel virtual reality-based training protocol for the enhancement of the "mental frame syncing" in individuals with Alzheimer's disease: a development-of-concept trial. *Front. Aging Neurosci.* 9:240. doi: 10.3389/fnagi.2017.00240
- Slater, M., Lotto, B., Arnold, M. M., and Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement. *Annu. Psicol.* 40, 193–210.
- Smith, G., Della Sala, S., Logie, R. H., and Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: a questionnaire study. *Memory* 8, 311–321. doi: 10.1080/09658210050117735

- Squire, L. R., and Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proc. Natl. Acad. Sci. U S A* 93, 13515–13522. doi: 10.1073/pnas.93.24.13515
- Sutcliffe, A., Gault, B., and Shin, J. E. (2005). Presence, memory and interaction in virtual environments. *Int. J. Hum. Comput. Stud.* 62, 307–327. doi: 10.1016/j.ijhcs.2004.11.010
- Taillade, M., Sauzéon, H., Arvind Pala, P., Déjos, M., Larrue, F., Gross, C., et al. (2013). Age-related wayfinding differences in real large-scale environments: detrimental motor control effects during spatial learning are mediated by executive decline? *PLoS One* 8:e67193. doi: 10.1371/journal.pone.0067193
- Tieri, G., Morone, G., Paolucci, S., and Iosa, M. (2018). Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. *Expert Rev. Med. Devices* 15, 107–117. doi: 10.1080/17434440.2018.1425613
- Tisserand, D. J., and Jolles, J. (2003). Special issue on the involvement of prefrontal networks in cognitive ageing. *Cortex* 39, 1107–1128. doi: 10.1016/s0010-9452(08)70880-3
- Trawley, S. L., Law, A. S., Brown, L. A., Niven, E. H., and Logie, R. H. (2014). Prospective memory in a virtual environment: beneficial effects of cue saliency. *J. Cogn. Psychol.* 26, 39–47. doi: 10.1080/20445911.2013.852199
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25. doi: 10.1146/annurev.psych.53.100901.135114
- Wilson, B., Cockburn, J., and Baddeley, A. D. (1985). *The Rivermead Behavioural Memory Test*. Reading: Thames Valley Company.
- West, G. L., Zendel, B. R., Konishi, K., Benady-Chorney, J., Bohbot, V. D., Peretz, I., et al. (2017). Playing Super Mario 64 increases hippocampal grey matter in older adults. *PLoS One* 12:e0187779. doi: 10.1371/journal.pone.0187779
- Yip, B. C. B., and Man, D. W. K. (2013). Virtual reality-based prospective memory training program for people with acquired brain injury. *NeuroRehabilitation* 32, 103–115. doi: 10.3233/nre-130827

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

Copyright © 2020 Rizzo, Gambino, Sardo and Rizzo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership