Università di Genova
DIPARTIMENTO DI INFORMATICA, BIOINGEGNERIA, ROBOTICA E INGEGNERIA DEI SISTEMI

The Perception/Action loop: A Study on the Bandwidth of Human Perception and on Natural Human Computer Interaction for Immersive Virtual Reality Applications

Chiara Bassano

Theses Series DIBRIS-TH-2020-42

Università di **Genova**

Dipartimento di Informatica, Bioingegneria,
Robotica ed Ingegneria dei Sistemi

Ph.D. Thesis in
Computer Science and Systems Engineering
Computer Science Curriculum

The Perception/Action loop: A Study on the Bandwidth of Human Perception and on Natural Human Computer Interaction for Immersive Virtual Reality Applications

by

Chiara Bassano

May, 2020

Ph.D. Thesis in Computer Science and Systems Engineering (S.S.D. INF/01)
Dipartimento di Informatica, Bioingegneria,
Robotica ed Ingegneria dei Sistemi
Università di Genova

Candidate
Chiara Bassano
chiara.bassano@dibris.unige.it

Title
The Perception/Action loop: A Study on the Bandwidth of Human Perception and on Natural Human Computer Interaction for Immersive Virtual Reality Applications

Advisors
Fabio Solari
DIBRIS, Università di Genova
fabio.solari@unige.it

Manuela Chessa
DIBRIS, Università di Genova
manuela.chessa@unige.it

External Reviewers
Alexis Paljic
MINES ParisTech
alexis.paljic@mines-paristech.fr

David Rudrauf
University of Geneva
david.rudrauf@unige.ch

Location
DIBRIS, Univ. di Genova
Via Dodecaneso, 35
I-16146 Genova, Italy

Submitted On
May 2020

Phenomenal cosmic powers
Itty bitty living space.

— Genie in Aladdin

Abstract

Virtual Reality (VR) is an innovating technology which, in the last decade, has had a widespread success, mainly thanks to the release of low cost devices, which have contributed to the diversification of its domains of application. In particular, the current work mainly focuses on the general mechanisms underling perception/action loop in VR, in order to improve the design and implementation of applications for training and simulation in immersive VR, especially in the context of Industry 4.0 and the medical field. On the one hand, we want to understand how humans gather and process all the information presented in a virtual environment, through the evaluation of the visual system bandwidth. On the other hand, since interface has to be a sort of transparent layer allowing trainees to accomplish a task without directing any cognitive effort on the interaction itself, we compare two state of the art solutions for selection and manipulation tasks, a touchful one, the HTC Vive controllers, and a touchless vision-based one, the Leap Motion.

To this aim we have developed ad hoc frameworks and methodologies. The software frameworks consist in the creation of VR scenarios, where the experimenter can choose the modality of interaction and the headset to be used and set experimental parameters, guaranteeing experiments repeatability and controlled conditions. The methodology includes the evaluation of performance, user experience and preferences, considering both quantitative and qualitative metrics derived from the collection and the analysis of heterogeneous data, as physiological and inertial sensors measurements, timing and self-assessment questionnaires.

In general, VR has been found to be a powerful tool able to simulate specific situations in a realistic and involving way, eliciting user's *sense of presence*, without causing severe *cybersickness*, at least when interaction is limited to the peripersonal and near-action space. Moreover, when designing a VR application, it is possible to manipulate its features in order to trigger or avoid triggering specific emotions and voluntarily create potentially stressful or relaxing situations. Considering the ability of trainees to perceive and process information presented in an immersive virtual environment, results show that, when people are given enough time to build a gist of the scene, they are able to recognize a change with 0.75 accuracy when up to 8 elements are in the scene. For interaction, instead, when selection and manipulation tasks do not require fine movements, controllers and Leap Motion ensure comparable performance; whereas, when tasks are complex, the first solution turns out to be more stable and efficient, also because visual and audio feedback, provided as a substitute of the haptic one, does not substantially contribute to improve performance in the touchless case.

Publications

Some ideas and figures have appeared previously in the following publications:

Conference Proceedings

Chiara Bassano, Manuela Chessa, Fabio Solari. 'A Study on the Role of Feedback and Interface Modalities for Natural Interaction in Virtual Reality Environments'. Accepted at: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. SciTePress, 2020.

Chiara Bassano, Manuela Chessa, Luca Fengone, Luca Isgró, Fabio Solari, Giovanni Spallarossa, Davide Tozzi and Aldo Zini. 'Evaluation of a Virtual Reality System for Ship Handling Simulations'. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP, INSTICC. SciTePress, 2019, pp. 62–73.

Chiara Bassano, Fabio Solari and Manuela Chessa. 'Studying Natural Human-computer Interaction in Immersive Virtual Reality: A Comparison between Actions in the Peripersonal and in the Near-action Space.' In: VISIGRAPP (2: HUCAPP). 2018, pp. 108–115.

Manuela Chessa, Chiara Bassano, Elisa Gusai, Alice E Martis and Fabio Solari. 'Human-computer interaction approaches for the assessment and the practice of the cognitive capabilities of elderly people.' In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, pp. 66–81.

Elisa Gusai, Chiara Bassano, Fabio Solari and Manuela Chessa. 'Interaction in an immersive collaborative virtual reality environment: a comparison between leap motion and HTC controllers'. In: International Conference on Image Analysis and Processing. Springer. 2017, pp. 290–300.

Alice E Martis, Chiara Bassano, Fabio Solari and Manuela Chessa. 'Going to a virtual supermarket: comparison of different techniques for interacting in a serious game for the assessment of the cognitive status'. In: International Conference on Image Analysis and Processing. Springer. 2017, pp. 281–289.

POSTERS AND ORAL PRESENTATIONS

Chiara Bassano, Giorgio Ballestin, Eleonora Ceccaldi, Fanny Isabelle Larradet, Maurizio Mancini, Erica Volta and Radoslaw Niewiadomski. 'A VR Game-Based System for Multimodal Emotion Data Collection'. In: Motion, Interaction and Games. MIG '19. Newcastle upon Tyne, United Kingdom: Association for Computing Machinery, 2019.

Chiara Bassano, Fabio Solari, Manuela Chessa. 'The Limits of Visual Perception in Immersive Virtual Reality: A Change Blindness Study 42nd European Conference on Visual Perception (ECVP) 2019 Leuven'. In: Perception 48.2 suppl (2019), p. 220.

Acknowledgements

First of all, I would like to thank my supervisors Manuela Chessa and Fabio Solari, for their support, their time and for sharing their experience and expertise with me.
Then, many thanks to everybody who kindly gave me their time, patience and energy to participate in my experiments, receiving no reward!
Finally, to my colleagues I would like to say I don't know half of them half as well as I should like; and I like less than half of them half as well as they deserve.

Contents

List of Figures

List of Tables

Acronyms

AC - Affective Computing

ADL - Activities of Daily Living

CB - Change Blindness

CVE -Collaborative Virtual Environment

COP - Centre of Pressure

DO-IT - Desktop Oriented Interaction Technique

DOF - Degree of Freedom

EEG - Electroencephalogram

FAR - False Alarm Rate

FN - False Negative

FOV - Field of View

FP - False Positive

GPCog - General Practitioner Cognitive Assessment of Cognition

HCI - Human Computer Interface

HIP - Human Information Processing

HMD - Head-mounted Display

HOMER - Hand-centered Object Manipulation Extending Raycasting

HR - Hit Rate

IPQ - IGroup Presence Questionnaire

IR - Infrared

ISI - Inter Stimulus Interval

MCI - Mild Cognitive Impairment

MMSE - Mini-Mental State Examination

MOCAP - Motion Capture

NHCI - Natural Human Computer Interface

PCA - Principal Component Analysis

RGB - Red Green Blue

RGB-D - Red Green Blue-Depth

SMPTE - Society of Motion Picture and Television Engineers

SPMSQ - Short Portable Mental Status Questionnaire

SSQ - Simulator Sickness Questionnaire

TN - True Negative

TP - True Positive

UDP - User Data Protocol

UI - User Interface

VA - Visual Angle

VE - Virtual Environment

VIMS - Visually induced Motion Sickness

VR - Virtual Reality

VWM - Visual Working Memory

WIM - World in Miniature

WM - Working Memory

1

Introduction

Virtual Reality has had a widespread success in the last decade, however, being a constantly evolving sector, research in this area is very rich and diversified, with lots of open questions, issues to be solved and aspects requiring further investigations, from hardware implementation to graphics, from user experience to interaction. Furthermore, with the diffusion of this new technology, the areas of application are multiple and diverse. The following part, thus, aims at giving the reader a complete overview over the specific context and motivation of the research work presented in this thesis and highlights its contributions.

Virtual Reality is an innovative and promising technology which is currently experiencing a widespread success thanks to the recent release of low cost consumer hardware. In the past, in fact, as only expensive systems were available, the use of VR was limited to a few specific sectors, such as military simulation, cinema and multimedia production. In the last decade, however, the release of cheaper devices has led to the expansion of its application domains and, nowadays, Virtual Reality is in common use in various contexts, ranging from video games and entertainment to simulation, training and learning in the industrial (Industry 4.0) and medical field or edutainment.

Virtual Reality is a computer technology that replicates an environment (real or imaginary), creating an artificial sensorial experience that allows isolating the user from the real world surrounding him and giving him the illusion to be physically present in the virtual environment (VE), where he can watch, explore and eventually interact with virtual objects. When talking about VR, however, it is necessary to discern between non-immersive and immersive setups. The first refer to screen-based solutions (desktop or laptop) presenting the VE to a user without occluding his Field of View (FOV), so he can feel involved but does not experience the sense of actually being in the VE. The latter, instead, concern the use of room-filling technologies, head-mounted displays (HMDs) and mobile headsets, occluding the real world surrounding the user, so that he can only see the virtual environment, interact with it and explore it, enhancing the sense of being physically present there.

The sense of immersion these devices provide can be mainly attributed to the involvement of different human sensorial systems (proprioceptive, vestibular and visual), whose information are fused generating a realistic artificial sensorial experience. Although the powerful effect of vision alone over user's self-awareness and *sense of presence*, a VR experience only based on vision would be limited, as the user would have just a passive role. Interaction and navigation, thus, are fundamental and have been the propeller for the birth of new enterprises and researches leading to the implementation of innovative hardware and software. Solutions currently available for interaction, in fact, include not only ad hoc designed joysticks and controllers, sold with the headsets, but also a variety of hand and body gesture recognition approaches, which can be easily integrated with commercial HMDs. Considering navigation, the majority of headsets provide simple tracking systems, able to localize and track the device in a limited area in the real 3D space, without the need of sophisticate motion capture systems.

The potential this technology offers, therefore, has aroused the interest of industries and researchers, who are trying to include Virtual Reality into areas such as data visualization [171], serious games and edutainment [34, 89, 112], support for surgeons in diagnosis, operation planning and minimal invasive surgery simulation, physical and cognitive rehabilitation or for assistive purposes [20, 29, 32, 71, 104], assessment and training of elderly people [44, 55, 98, 101, 123, 159], Industry 4.0 [182], for modelling and products design, complex engines maintenance, assembly and prototyping process, teaching security for work places fatality reduction, training for risky situation managing [84, 85], drive or flight simulator [15, 72, 165, 166, 173]. Moreover, VR technologies are widely applied in the investigation of complex human behaviours and emotions elicitation. For example, VR is a well-established medium for investigating fear perception and treatment [19, 54, 113, 140]. The study of emotional states in VR is important not only in psychology, but also for the advantages it could take to the enhancement of presence perception [53].

1.1 Context

Simulation and training sessions take a lot of advantages from Virtual Reality setups with respect to real-life settings. First of all, trainees can be immersed in dangerous and potentially fatal scenarios and unethical situations while physically being in a totally safe controlled environment, guaranteeing the immunity of the trainee himself and of third parties. Moreover, it is possible to reproduce environmental and social situations that can stimulate the subject in a similar way to the corresponding real-life counterpart [138]. Secondly, VR ensures active learning or learning by doing, meaning the consolidation of theoretical knowledges through practical experience. Indeed, a number of recent studies in cognitive psychology have investigated the potential benefits of active learning [38, 56] through the use of VR technologies, which provide a multisensory, interactive environment that is engaging and that allows learners to construct meaning from experience. Thirdly, being simulations parameters settable and scenarios controllable, training sessions are repeatable, as it is possible to replicate the same exact conditions, and their difficulty can be adapted to users expertise. Meanwhile, people well-being state and improvements can be monitored, by recording multimodal data, i.e. physiological or kinematic measurements, videos, audio, times, used to calculate, for example, the accuracy of an action or the time required to complete a task or a subtask.

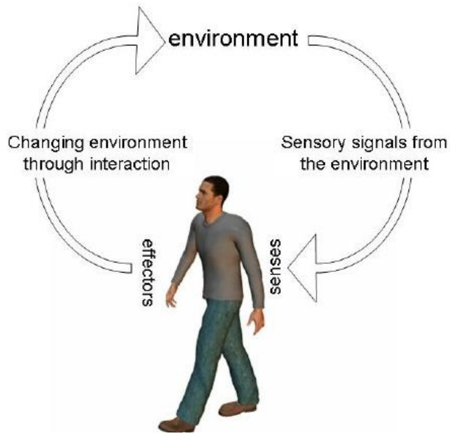
The current work mainly focuses on the general mechanisms underlying simulation and training in VR, considering the industrial and medical context, both surgery and rehabilitation, as the final area of application. Results found, however, can be generalized to any context concerning training and simulation.

1.2 Motivation

When designing a VR application for training and simulation, two main aspects have to be taken into account: firstly, humans must gather and understand all the information presented; secondly, they have to interact with them. So, the whole design and implementation phase revolves around the user, around his perception of the VE and his ability to interact with it. This perception/action loop (see Figure 1) is something we experience in our daily life: we voluntarily and consciously perform actions on the environment, guided by schemas already laid down in memory, so that we can percept and learn from the environment and constantly update our schemas based on our experience. VR, however is not the real world. It is, thus, of crucial importance addressing the perception/action loop in VR, in order to design effective VR systems for the contexts of interest.

On the one hand, my research concentrates on the understanding of the visual system bandwidth, meaning the amount of visual information, coming from immersive 360° VE, humans are able to effectively process, through a systematic methodology inspired to the literature but adapted to VR. Results can constitute a guideline delineating the best way to present information in VR, so that users can take advantage of the immersivity of headsets, without requiring too much cognitive effort.

On the other hand, since user's cognitive effort have to be entirely task directed, I focused on Natural Human Computer Interface (NHCI), trying to define and implement the best interaction modality specifically for selection and manipulation tasks, in terms of naturalness, intuitiveness, stability and efficiency and considering performance and preferences as the main evaluation parameters.



(a)

Figure 1: General perception/action loop schema (a). Images were taken respectively from [39].

1.3 Contribution

The contribution of this thesis consists in the development of a framework and of a methodology for the understanding and analysis of human perception of 3D information displayed in immersive VR and interaction within a 3D immersive VE.

The software framework consists in applications for Virtual Reality, designed taking into account both the needs of the final user and of the experimenter. Each application, in fact, is constituted of two components: a VR scenario, where both the graphics and the interface implementation were considered, and a desktop window, allowing the experimenter to select the preferred modality of interaction (e.g. different devices can be easily exchanged), to monitor what the player is doing and eventually intervene to help him and to set all the experimental parameters, guaranteeing experiments' repeatability and controlled conditions.

Considering the methodology, performance and user experience have been evaluated analysing quantitative and qualitative parameters derived from the collection of heterogeneous data, physiological measurements, head position and rotations, timing, precision and self-assessment validated questionnaires.

In particular, on the one hand we tried to evaluate the bandwidth of human perception in VR by adapting the Change Blindness standard method to be used within a virtual environment and considering the influence of five different conditions on participants ability to detect a change, i.e. the observation time, the number of items to be memorized, the FOV width, the spatial layout and the kind of change [12]. On the other hand, two different state of the art approaches for interaction with virtual objects, namely the controllers and a vision-based solution, the Leap Motion, were tested and compared [10, 11, 68]. In this case, the analysis of performance and user experience, related to specific tasks and VR scenarios, allows to evaluate the limits and the potentiality of both solutions and to design better HCIs for the considered contexts.

Finally, three real world use cases were considered: (i) an industrial use case [9]; (ii) a cognitive research use case [8]; (iii) a medical application for cognitive training [33, 100]. The industrial use case consisted in a collaboration with a local company with a strong background in computer graphics and physical simulation, CETENA, for the evaluation of the usability of a ship handling simulator system. The cognitive research use case concerns the creation of an ecological multimodal dataset for detection and recognition of specific affective states in collaboration with Casa Paganini-InfoMus Research Center. In particular, I was asked to implemented an ad hoc VR game in order to elicit two specific emotions, joy and frustration, according to a well-established psychological theory, i.e., Roseman's "appraisal theory". In the last use case, I participated to the development of a system for the cognitive assessment of a specific population (elderly), in the context of a project my research team is working on.

1.4 Outline

In the next chapters, first, I will briefly introduce state of the art technologies and software frameworks currently available for VR, devices and approaches for interaction, highlighting their pros and cons, potentialities and limitations (Chapter 2).

Then, I will explain how my research on the evaluation of human perception in VR (Section 3.1) and Natural Human Computer Interaction (Section 3.2) has been articulated, with particular attention to the specifically developed methodologies and software frameworks, and I will present obtained results. I will, moreover, describe the three projects to which I participated, representing

three different use cases (Section 3.3).
Finally, I will discuss the obtained findings with respect to the original research questions (Chapter 4) and will identify some open issues and possible future developments (Chapter 5).

2

Background

The following part aims to include all the state of the art necessary to the understanding of the research work described in this thesis. As different topics were involved, i.e. perception and interaction within an immersive virtual environment and User Experience Assessment, this part incorporates multidisciplinary and heterogeneous information.

(a) (b) (c) (d)

Figure 2: Examples of non-immersive (a) and immersive (b, c, d) VR systems. (a) Non-immersive setup based on Nintendo Wii. (b) A room filling technology, the CAVE. (c) Two head-mounted displays, the Oculus (top) and the HTC Vive (bottom). (d) The Gear VR, a mobile headset.

2.1 VR Devices

Although Virtual Reality has had a widespread success only in the last decade, its origins date back to the '60. Since then, lots of researches have been conducted in this field, leading to the creation of new prototypes and release of new devices. Nowadays, however, all solutions for VR have been classified in two main categories, non-immersive and immersive setups (Figure 2).

Non-immersive systems are based on the use of screens (desktop screens, laptop or tablets). The virtual environment is presented to the user without occluding his Field of View, hence, even if he feels involved and engaged in the task, the sense of being in the real world while interacting with the virtual one persists.

Immersive systems, instead, concern room-filling technologies, such as the CAVE [1], the AlloSphere [2] or the YURT[3], head-mounted displays, for example

1 http://www.visbox.com/products/cave/
2 http://www.allosphere.ucsb.edu/
3 https://www.brown.edu/academics/early-cultures/resources-brown/yurt

the Oculus Rift, S, Quest and Go [4], the HTC Vive, Pro, Focus and Cosmos [5] (by HTC), the Project Morpheus [6] (by Sony), and mobile headsets, as the Gear VR by Oculus or the Google Cardboard [7] (by Google).

Room-filling facilities are projection systems composed by multiple full-HD stereo displays covering a 360-degree surface, including, in some cases, overhead and underfoot. Users wear stereoscopic glasses, this way 3D graphics and images appear to be floating, suspended in the air, and their movements are tracked, so that projected images can be adjusted continuously to adapt to the individual perspective [47]. These systems are in general expensive and mainly used by universities or industries in fields such as large and complex data visualization, interaction with a wide range of scenarios, from Mars to plankton to chambers of the heart, reconstruction of ancient artifacts or even entire archaeological sites.

HMDs and mobile headsets, instead, are two different category of mass market devices. HMDs have integrated displays, inertial sensors for head movements detection and external tracking systems. Some of them (Oculus Rift and HTC Vive/HTC Vive Pro) have to be tethered to a high end computer, with a dedicated graphics card (usually last generation NVIDIA or AMD Radeon), a spacious RAM (minimum 4 GB) and a powerful processor; while others (Oculus S, Quest and Go and HTC Vive Focus and Cosmos) represent the newer all-in-one portable solutions with embedded processors and sensors. Nevertheless, the former guarantee high performance, fast refresh rate and a good resolution, whereas the latter ensure portability and good standalone experiences. Finally, mobile headsets require the insertion of a smartphone, used as display and processor. These solutions allow obtaining acceptable VR experiences but for a limited time, with a smaller FOV and with a lower resolution and nowadays, because of their strong limitations, they are exiting the market, substituted by the new generation all-in-one devices.

While immersive environments allow participants to feel as they are inside the environment, as further discussed in Section 2.4.2, non-immersive environments only allow participants to see the contents based on how the device in use, PC, smartphone or tablet, is held and moved [168]. Nonetheless, larger displays, higher resolutions, faster refresh rates and stereoscopic capabilities have increased the display fidelity. A common assumption is that increased immersion is also associated with higher *sense of presence* and better task performance and behaviour change. However, conflicting views exist [78]: by the comparison of different VE platforms (immersive vs. non-immersive) and narrative contexts (emotional vs. non-emotional), some researchers [65] have found that both factors have a significant role in generating an effective VR experience, as they contribute in increasing the feeling of presence, while other [66] stated that only the emotional context impacted on the *sense of presence*. Also, user characteristics (e.g. gender, immersive tendency and previous experience with VR) seem influencing obtained results [79]. Moreover, different VE platforms

4 https://www.oculus.com/
5 https://www.vive.com/
6 https://www.playstation.com/it-it/explore/playstation-vr/
7 https://arvr.google.com/cardboard/

(desktop PC, HMD, and CAVE) have been found to induce different patterns of emotional responses in contexts including high- and low-stress tasks. In particular, a positive relationship exists between the level of immersion and behavioural responses and performance in well-path planning [67], statistical data application [5] and navigation of large-scale virtual environments [143], even if higher simulator sickness is associated to more immersive technologies [82].

The sense of immersion immersive devices provide can be mainly attributed to the involvement of four different human sensorial systems, visual, auditory, vestibular and proprioceptive, responsible for the self awareness of the overall representation of body position, movement, and acceleration. Information coming from these sensory channels is then fused generating a realistic artificial sensorial experience. Vision, in particular, has a very important role in human perception and *sense of presence*. However, considering the potential of immersive VR technology, an experience only based on vision would be limited, as the user would have just a passive role. Interaction with virtual objects and User Interfaces (UI) and navigation of the virtual environment (VE), thus, are fundamental and have been the propeller for the birth of new enterprises and researches leading to the implementation of innovative hardware and software.

Solutions currently available for interaction include not only ad hoc designed joysticks and wireless controllers sold with the headsets (Figure 2 c and d), but also a variety of hand and body gesture recognition approaches, discussed in more detailed in Section 2.3. New controllers integrate the functionalities of two standard videogames interaction tools: joysticks and wand, like the Wiimote or the PlayStation Move. On the one hand, they provide an intuitive interaction through buttons, triggers and touchpads, and feedback through vibration; on the other hand, thanks to inertial sensors and external tracking, positions and rotations in the 3D real space can be used to perform movements like throwing and carrying objects around the scene, raycast, shoot arrows, playing instrument, fighting with sword, punching...

Considering navigation, the majority of headsets include simple external 6DOF tracking systems, able to localize and track the device in a limited area in the real space. HTC Vive and Oculus tracking systems offer even better performance, higher resolution, submillimetric precision and lower weight with respect to expensive specialized hardware [185].
HTC Vive provides a 3 m x 4 m trackable area. The headset and the controllers are covered with photosensors able to recognize the synchronization pulse and the laser IR beams emitted by two Lighthouse base stations sold with the HMD [8]. Whereas Oculus uses a constellation tracking system, based on the detection and triangulation of arrays of infrared micro-LEDs on the headset and controllers surface. The tracking area, however, is limited to cameras' field of view. For this reason, Oculus Rift DK2 was mainly suggested for sitting or standing scenarios [48]. It is worth noting, however, that the new Oculus re-

8 https://www.reddit.com/r/Vive/comments/40877n/vive_lighthouse_explained/

lease provides a better tracking system, a larger tracking area and performance comparable to those of the HTC Vive.

A recent study comparing performance of the HTC Vive and Oculus Rift DK2 in a Pick-and-Place task, pointed out that the HTC Vive tracking system is more reliable and stable than the Oculus Rift one, while there is no difference in terms of precision [157]. So, the choice of the proper system depends on the specific application: if we want the user to be free to move in a virtual room, then the HTC Vive could be a better solution; while if the player has to interact with objects in the peripersonal space, the two devices are equivalent.

The main advantage of having an external tracking system provided with the headset consists in the possibility to develop exergames or applications in which the user is asked to walk in a limited area, without the need of expensive motion capture systems (MOCAP). MOCAP are tracking systems able to localize an object or body in the 3D space. They can be classified in optical and non-optical systems. The former utilize data captured from image sensors to triangulate the 3D position of a subject between two or more calibrated cameras. This data acquisition can be implemented using passive markers (Qualysis [9]), coated with a retroreflective material reflecting the light generated by cameras, active markers (Vicon [10]), emitting their own IR light, or markerless solutions, in which features of the interested surface are dynamically identified. The latter include inertial systems, based on algorithms fusing information recorded by different inertial sensors, exoskeleton MOCAP systems, directly tracking body joint angles, and magnetic systems, calculating real time the relative magnetic flux of three orthogonal coils on both the transmitter and each receiver. Optical systems are the most commonly used as they offer accurate tracking with submillimetric precision, stability to occlusion, as each marker is seen by multiple cameras, and a large expandable trackable area.

2.2 Attention and Working Memory

The moment we open our eyes, we experience a vast, richly detailed world extending well into the periphery and we have the impression to be able to gather and process the majority of the visual information we receive [41]. However, numerous experimental results indicate that the bandwidth of human perception is severely limited. According to [17] we actually have a rich experience of the world, but all this information can not be fully captured by our capacity-limited cognitive mechanisms. Other authors, however, state information is not consciously perceived until it is accessed by higher-order systems, i.e. attention, working memory, and decision-making [87, 91]. While a last group of researchers argue that, even if conscious perception is limited by cognitive mechanisms, the visual information observers have access to is not sparse and postulate the visual ensembles and summary statistics theory [2]: observers

9 https://www.qualisys.com/
10 https://www.vicon.com/

only perceive a subset of the scene at high resolution, but they are aware of the entire scene, which is represented with poorer resolution, averaging across imprecise representations. This is possible because real-world scenes do not comprise random bits of uncorrelated information but have structures, regularities and redundancies. All three theories are based on the existence of a "limited cognitive mechanism", often referred to as Visual Working Memory (VWM).

2.2.1 *Visual Working Memory*

Visual Working Memory is an apparatus dedicated to the active maintenance of visual information to serve the needs of ongoing tasks. This definition encapsules three key points: first, it implies that input information must be visual in nature; second, VWM is based on active maintenance, that is a change in sustained, energy-requiring neural activity, rather than by a change in synaptic strength; third, representations must be used in the service of broader cognitive tasks [97]. This mechanism is limited both in terms of time, it decays in some seconds unless information reinforcement occurs, and of memory capacity, i.e. the number of information which can be stored.

The evaluation of VWM capacity is still controversial. Many researchers claim that a plausible threshold could be around 4 [41, 45] or 7 ± 2 items [60, 109], depending on the task, and that such behaviour might have a physiological base [97]. According to [96], participants can remember a maximum of four objects, independently from the number of features per objects to be remembered (at least up to four features). They state that VWM capacity should be understood in terms of integrated objects rather than individual features. Other authors, in fact, prefer using the term chunk [26], i.e. higher order representations of the items in which individual pieces of information are inter-associated and stored in memory and act as a coherent, integrated group when retrieved. Researchers in [3] found a monotonical relation between the amount of information per item which can be stored and the reciprocal of the number of items which can be memorized, meaning there is a trade-off between complexity (quality) of stimulus, and number of objects stored. Furthermore, [184] have recently proven that individuals with high VWM capacity are able to adjust memory precision flexibly according to task requirements and to voluntary trade-off between precision of internal representations and number of items for a given set size, when they have enough consolidation time. Moreover, they suggested a positive correlation between the ability of voluntary trade-off and VWM capacity. These results might seem in contrast with [96] one, but, hypothesizing the existence of separate stores for individual, basic features such as size, colour, and orientation, reaching the capacity limit for one feature would not prevent the subject from taking on more information about another one. On the same time, objects with multiple values on a single dimension would be more difficult to remember than objects with a single value on that dimension [175]. Besides, the minimal representation of an object includes a minimum set of core features that are always encoded regardless of the task: as the visual information required to discriminate between objects increases, the maximum number of objects that can be stored in memory decreases.

Change Blindness (CB) and Inattentional Blindness experiments represent a common tool for the study of VWM capacity and of the amount of the available information that goes unnoticed.

2.2.2 *Change Blindness paradigms*

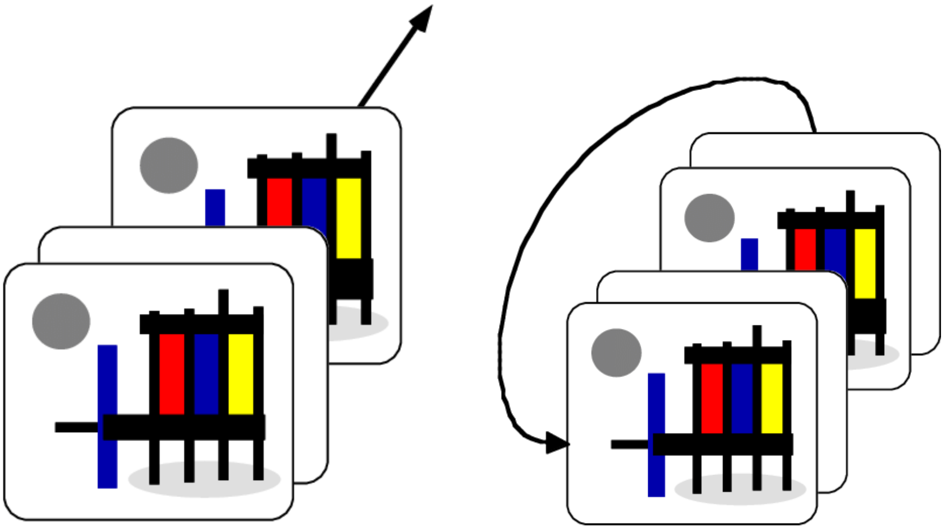
Change Blindness is the inability to detect a change expected or unexpected between two different pictures when a brief interruption occurs between the two images or the change occurs so gradually that it does not automatically draw attention. Inattentional Blindness, instead, is the failure to notice an otherwise visible stimulus when attention is directed elsewhere. My research mainly focuses on Change Blindness.

A first classification for CB experiments consists in the distinction of *intentional* and *incidental* Change Blindness induction[164]: in the first case, observers are intentionally looking for changes; while in the second case, they are not instructed to search for changes. The adopted paradigm may affect participants behaviour and attention, thus influencing the specific kind of failure that underlies Change Blindness, influencing both the results and their interpretation.

Moreover, Change Blindness experiments found in the literature can be of three different types: videos in which the change is gradual enough so that the accompanying motion signal does not draw attention [149]; sequences of pictures in which the change is contingent on an event (such as a brief flash, eye movement, or occlusion) that creates a global motion signal that masks the transient [136]; real life settings, demonstrating how CB can occur in normal daily situations [150].

Considering the sequence of pictures, two paradigms have been implemented, the *flicker* and the *one-shot* (Figure 3). In the first case, the original image is shown for a time T1, then a blank screen (or retention mask) appears for a time T2, after this a modified image (test) is shown for a time T1 and finally the blank screen appears again for a time T2. This is repeated in loop until the observer finds the change. In the second case, instead, the original image is shown for a time T1, then a blank screen appears for a time T2, after this a modified image is shown for a time T3 or until the participant identifies the change. In both cases, T2, namely Inter Stimulus Interval (ISI), can vary from 0.02 s to 9 s, but in general experimenters prefer using a time around 500 ms, in order to mask the transient without impairing VWM. T1, instead, is usually around 275–300 ms, corresponding to the minimum time required to extract the gist of the scene [41].

Another important distinction for *one-shot* experiments is between *single-probe* and *whole-display* paradigms [141] (Figure 4). In the first case, during test, participants must make a recognition judgement, deciding if the proposed target is novel or if it was present in the input stimuli. In the second case, instead, a full set of items is presented at test. This set can be the same as the original,



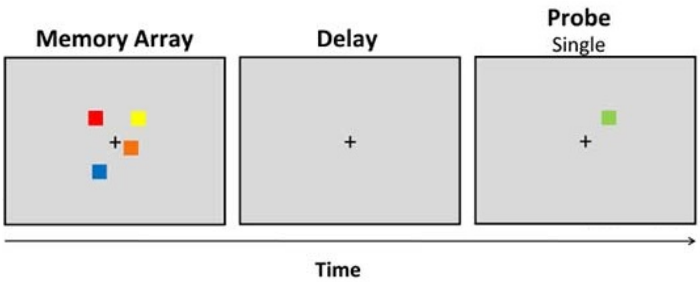
(a) (b)

Figure 3: (a) One-shot paradigm. (b) Flicker paradigm. Images were taken from [135].

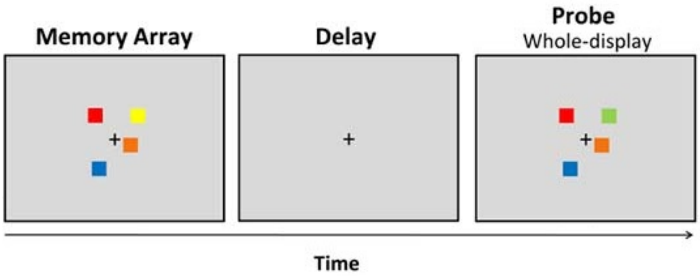
or one item is novel. In other words, in the *single-probe* task, subjects have to evaluate the status of a single element, while in the *whole-display* they have to evaluate the status of all proposed elements. The two tests yield different outcomes and imply the use of different methods for analysis.

Depending on the specific task, stimuli can vary from simple lines, shapes, matrices or letters to complex pictures of real scenarios. It is worth noting that complex inputs, although resembling our daily experience, introduce a difficult to quantify amount of visual information, both in terms of elements to be memorized and semantic of the scene, which could prioritize attention, making it difficult to identify the processes underlying CB. Other major problems related to real-life settings is that they are difficult to control, repeatability is arguable and objective parameters are hard to measure. For this reason, Virtual Reality could represent a valid alternative as it would allow to conduct experiments in a fully controlled 3D environment, manipulate at will the virtual world and different experimental parameters, record measurements, as user's gaze or reaction time, and explore the effect of visual cues, such as stereopsis, motion parallax and depth perception, and pictorial cues, as occlusion, shading and perspective. Also, [57] have recently measured VR headsets, in particular the Oculus Rift, test-retest reliability, assessing its applicability for standardized and reliable assessment of elementary cognitive functions. Another advantage of VR is the possibility to enlarge the space available for stimulus presentation, traditionally limited around 35 degrees of the visual angle, having the potential to raise it up to 360 degrees.

The majority of experiments in VR found in the literature investigate *incidental* Change Blindness induction in dynamic contexts in dual tasks. In general, participants are asked to accomplish a task, sort objects, walk or drive and unexpected changes are applied to the scene [74, 148, 156, 160]. Task irrelevant changes go unnoticed suggesting that human vision has a highly purposive



(a)



(b)

Figure 4: (a) Single-probe paradim: a memory display with a defined set size is shown and, after a delay, user is asked to recognize if an item was in the first image or not. (b) Whole-display paradgm:user is asked to detect if a change have occured between memory display and test probe. Images were taken from [183].

and task specific nature, thus information extracted is used for ongoing computations. Only a few works investigating *intentional* CB have been found. One of the main problem researchers have to face, in fact, is the adaptation of standard paradigms, conceived for 2D, to be used in Virtual Realilty. For example, as flicker paradigm has been found to cause sickness, [154] implemented a solution in which different de-phased images are projected onto the two eyes. Also in [167] authors have tried to design a static passive experiment for Change Blindness, through the introduction of a new paradigm: when people look around in the immersive virtual environment, the object supposed to appear/disappear changes its visible/invisible state each time it falls in and out the visual field of the user. Authors themselves, however, have pointed out that such kind of paradigm allows participants to create their own personal strategy, making the setup less controllable and compromising results.

2.2.3 *Factors influencing Change Blindness*

Results highlight that blindness to changes can be influenced by many different factors [7]: attention; personal preferences; the semantic of the scene; the stimuli's visual saliency; the presence of grouping relevant or irrelevant features; the number of elements to be memorized.

Attention and interest are two different concepts. The term attention defines the subjective process of selecting some bits of information for further processing at the expense of others. Hence, visual perception of change in an object can not occur if that object is given no focused attention, and, in absence of such attention, the contents of visual memory are overwritten and can not be used to make comparisons. Attention is controlled both by top down mechanisms, as a priori knowledges, attentional management or inference over the change, and by bottom up processes, such as the transient generated by an element appearing or moving in the scene. Interest, instead, is an objective parameter: interesting (central interest) and not/less interesting (marginal interest) parts of the scene can be identified by showing a picture to a certain number of people (5 to 6 usually) for some minutes and then asking them to describe it by memory. The items remembered first [136] or reported by at least half of the judges [7, 28] are considered of central interest. In [136], authors proved that a change involving a central interest element is more easily and faster recognised than a marginal interest one, as, in absence of interest, attention is delocalized and detection of change will then require a slow, item-by-item scan.

As opposed to attention and interest, which can be considered high-level cues, saliency and perceptual grouping are low-level cues. Visual saliency is strictly related to the kind of change, considering both low level features (color, shift, appearance/disappearance) and high level characteristics (coherence/incoherence of the modified element), and elements distribution (random or regular). Researchers in [69, 149] demonstrated that the appearance/disappearance of an item is more easy and faster to detect, than other kind of changes; while [26] made a further distinction between appearance and disappearance, assessing that transient onset is more salient than transient offset. Authors formulated different hypotheses explaining this behaviour. Firstly, the appearance of an object introduces a new informative content, thus a new internal representation of the item has to be created, while removal often reveals non-informative regions. In fact, when the deletion of an item disoccludes objects, attention is prioritized. Secondly, no retrieval cue is required in the deletion condition, causing a failure to retrieve information from memory. Thirdly, the distinction between appearance and disappearance could be a consequence of the capacity constraints of Visual Working Memory: this system, in fact, must be constantly updated with new information, but given its limited capacity, relevant one are transferred to long-term memory while older, less relevant information decays.

Perceptual grouping, refers to the Gestalt grouping principles [174] influencing our perception and VWM storage, in terms of accuracy and fidelity, rather than priority, so that objects that appear as being grouped together are processed as a whole and stored together. This influence of grouping is even more evident when the amount of information to be remembered exceeds Visual Working Memory capacity [180]. Grouping occurs outside the focus of attention, at a preattentive stage of visual processing [90], when tested features are group relevant; whereas attentional prioritization may be required for grouping irrelevant features [95]. Example of grouping principles are proximity, i.e. grouping objects according to spatial nearness, similarity, referred to grouping based on similar features (colour, shape..) and connectedness. Proximity is considered to be faster than similarity even if, increasing the processing time, similarity overtook proximity and became to dominate [14]. These two grouping mechanisms can occur separately or combined, but their effect can not be simply accumulated, as they may even interfere with each other [95].

2.3 Natural Human Computer Interaction in VR

The level of naturalism or fidelity of User Interfaces is defined as the objective degree with which actions performed to accomplish the task using the interface, correspond to the actions used for the same task in the real world [107]. For this reason UI designers are, nowadays, particularly interested in 3D interfaces allowing to exploit 6-degree-of-freedom (DOF). According to [25], however, interaction fidelity is a continuum, and not a binary, and the definition itself of naturalism is nuanced and depend on many different factors. For example, given any pair of interaction techniques for a given task, it may be difficult to understand which one has higher interaction fidelity: one, in fact, could better replicate movements to be performed, while the other could more closely approximate the force required by the user. For this reason [107] have developed a framework, the Framework for Interaction Fidelity Analysis (FIFA), for the comparison of interaction techniques to their real-world counterparts. This evaluation scale considers three primary factors: the biomechanical symmetry, involving the comparison of the kinematic, kinetic and anthropometric aspects; the control symmetry, considering the dimensional transfer function and termination characteristics; the system appropriateness, which is concerned with how well a VR system matches the interaction space and objects of the real-world task.

In general, the level of naturalism depends on both the interaction technique and the task context, e.g. a steering wheel is natural when the task is driving but not natural in First Person Shooter games. However, some actions (for example, teleporting) can not be easily translated into natural physical mappings. In these cases, UI designers usually opt for traditional interaction methods, using buttons or hand gestures, although their lower level of interaction fidelity. In other cases, for example when the player must pedal a bicycle, the natural mapping using feet and legs motions would be impractical and could be substituted by hands and arms motions that mimic the real-world feet and legs cyclic movements. This solution is more natural than buttons and joysticks, but is less natural than riding a bike.

Natural interaction techniques bring different benefits. Natural travel methods based on head tracking, for example, result in higher levels of spatial understanding, although they do not allow long-distance navigation due to limited tracking volume. Moreover, natural manipulation techniques with 6-

DOFs significantly outperform the less natural techniques, especially when user is required to control both position and orientation at the same time [25]. It is worth noting, however, that mid-air interactions are often prone to fatigue and lead to a feeling of heaviness in the upper limbs, a condition termed as the gorilla-arm effect. Finally [105] evaluated the effect of the level of interaction fidelity on performance and user experience in a First Person Shooter. The high-interaction condition positively affected performance, was perceived as more familiar, similar to the real world interaction, and influenced presence, engagement and usability rates, which were higher.

Advantages of natural UIs are disparate: first of all they are fun, engaging, intuitive and easy to use also for novice users; secondly, for certain tasks, like pointing, turning and 6-DOF manipulation, humans have finely honed abilities that are hard to beat with other interaction style, as long as tracking quality is high; thirdly, for certain applications, such as training, using a natural UI ensures the training will transfer to the real world. However, natural interface design is not trivial, since simply increasing the level of fidelity does not seem to be sufficient to ensure usability and performance if the UI results unfamiliar. Even techniques that seem to be a good approximation of the real world may have negative effects due to small differences, such as latency or lack of force feedback. It is also important to remember that some tasks have no real-world counterpart that can be used as the basis for a natural UI, even though metaphors can be used to design interactions that are understandable based on real-world experiences. In these cases, traditional interaction techniques could be a good option. Despite their limited naturalism, they are not always inferior as they are familiar, they require minimal hardware and sensing and they are well established and ubiquitous.

Considering the context of this work, the application on simulation and training, interfaces are required to be a sort of transparent layer between user and task, in order to prevent people from diverting part of their cognitive effort managing the interface by taking resources away from understanding and learning the task itself. Ideally, bare hands and human body would be the ultimate form of Human Computer Interaction. Such kind of interfaces are often referred to as natural and ecological, as they are designed in order to reuse existing skills [62], through intuitive gestures requiring a little cognitive effort.

Thus, having many solutions available, one of the main factor to consider in order to make a weighted choice on the proper device to be used is the task, as interfaces are task dependent. In [61], authors classify interactions in four categories: (i) *navigation* and *travel*, which consists in moving the viewport or the avatar through the environment, thus including general movements and target movements; (ii) *selection*, touching or pointing something; (iii) *manipulation*, i.e. the modification of objects position, orientation, scale or shape; (iv) *system control*, corresponding to generating an event or command for functional interaction. My research mainly focuses on selection and manipulation.

Once defined the task, experimenter has to chose the proper device and algorithm. Moreover, according to the perception/action loop, it is important to consider also the kind of feedback users can receive. Haptic feedback, for example, seems playing a fundamental role on interaction, by improving performance and control over actions. In [111] authors point out factors affecting interaction in VR and precise manipulation of virtual objects. First of all, the lack of haptic feedback and physical constraints represents a strong restriction, as humans strongly rely on them in the real world experience. Secondly, most VE systems provide limited input information, i.e. user's head and, in some cases, hands pose, while other input modalities still play a marginal role. Finally, the lack of haptic and acoustic feedback together with inaccurate tracking and whole-hand input restricts users to the coarse manipulation of virtual objects, thus fine-grained manipulation is problematic.

2.3.1 *Devices for interaction*

Existing solutions for interaction in VR include both handleable devices, as those described in 2.1, and 6 DOF hands tracking approaches. Some of them make use of markers or sensorized gloves to improve hand detection [81], while other are based on bare hand interfaces, RGB, RGB-D or IR cameras, as the Leap Motion [50, 80, 131, 142, 188]. In particular, the Leap Motion is composed by two cameras and three infrared LEDs that allow to detect and track hands' joints inside a hemispheric area of approximately 1 meter above the device. Its stated accuracy is 1/100th of a millimeter in a range from 25 to 600 mm above the device. Anyway, for desktop use, it has been demonstrated that in static scenarios, the error in measurement is between 0.5 mm and 0.2 mm, depending on the plane and on the hand-device distance; while in dynamic situations (moving hands), independently from the plane, an average accuracy of 1.2 mm could be obtained [172].

Data acquired are runtime used to generate a virtual representation of the user's hands, representing an intuitive and natural manner to interact with objects and UI. Originally conceived for desktop applications, it has been recently adapted for VR applications: it is possible to attach a support at the centre of the frontal part of the HMDs, so people can move around and their hands are always displayed from their point of view (Figure 5 a). Vision-base solutions often present interaction and tracking challenges, as they are prone to occlusion, noisy reconstruction and artifacts [4] and often have a limited FOV (Figure 5 b), problems that controllers and joysticks can easily overcome. Despite its tracking issues, researchers are very interested in the application of the Leap Motion for different purposes: simulation of experiments [181] or oral and maxillofacial surgery [130], gaming, from puzzle [31] to First Person Shooter [30], model crafting [122] and visualization of complex dataset [43, 133].

2.3.2 *Algorithms for interaction*

The choice of the appropriate algorithms depends on the interaction space. If interactable objects are disposed in the peripersonal space or near action space

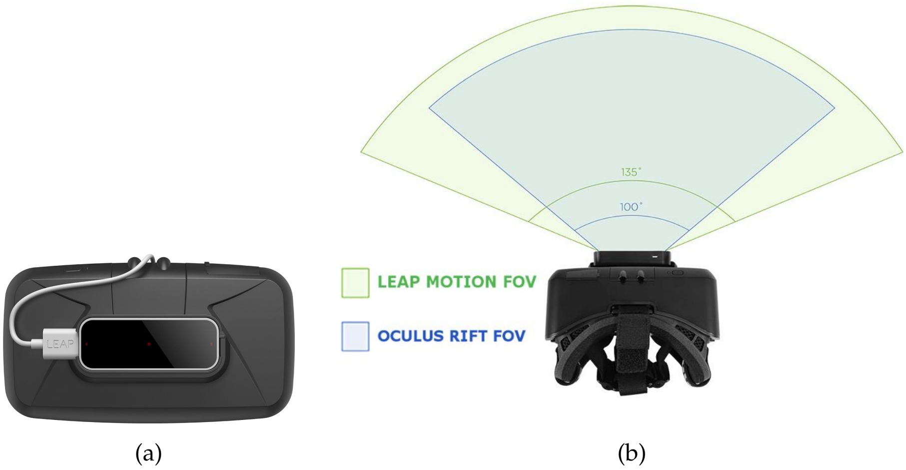


Figure 5: (a) Leap Motion device mounted on the Oculus Rift DK2. (b) FOV of the Oculus Rift and of the Leap Motion.

(i.e. they can be reached performing some steps), then a VR system allowing standing applications (the Oculus Rift) or room scale setups (HTC Vive) and a real-world interaction metaphor would be sufficient [48]. If objects are far away from the users, instead, due to the limited trackable area, a proper travel or reaching algorithm [24] should be implemented.

The real-world metaphor is a paradigm in which the user accomplishes reaching and grabbing task through natural gestures and poses: he reaches out his hands, grabs the object (both pushing a button or closing his own hand), and moves it around in the virtual environment using natural, physical motions [23]. Whereas, the techniques found in the literature for selection and manipulation of remote objects fit in four main categories:

1. *Arm-extention techniques*, where user's virtual arm grows to the desired length, based on hand actual position or velocity, so he can reach objects at a greater distance, while preserving the natural manipulation metaphor;

2. *Raycast techniques*, making use of a virtual laser beam to select and grab an object [110];

3. *World in Miniature* (WIM) [155], where the user holds a small representation of the environment and interacts with objects by manipulating their iconic versions in the miniature;

4. *Scaled-world method*, in which when an object is selected, all the scene is scaled down by a factor depending on hand and object position [111].

In general, during interaction, people look for a precise control over events and desire to put as less effort as possible in the task accomplishment, so a natural interface is not always the most effective solution, also because techniques which resemble to real world actions, may have negative effects due to the lack of haptic feedback or latency [25].

Authors in [23] compare raycast to different arm-extension methods, the so called "go-go" and indirect stretching. In the "go-go" method [126], while user's hand stays within a defined area, whose size depend on the specific implementation, the virtual hand moves in a one-to-one correspondence with the physical hand; while, when it goes beyond the threshold the virtual hand begins following a non-linear increasing function. In the indirect stretching technique, instead, user stretches or retracts the virtual arm by pressing buttons on a 3D mouse. Authors found out that users preferred raycast for selection, as it is simpler to use and more precise, while, arm-extension methods were the favourite solutions for manipulation, as they allow movements with 6 DOF and represent a transparent interaction, based on natural gestures. Consequently, new hybrid solutions have been implemented, the HOMER (Hand-centered Object Manipulation Extending Raycasting) and the scaled HOMER [176], exploiting raycast for selection and hand-centered mechanisms for manipulation. Moreover, [106] investigated the combined effect of interaction techniques and level of immersion on performance. Authors considered three methods: the "go-go", the HOMER and the DO-IT (Desktop Oriented Interaction Technique), which uses a window to mouse input mapping. Two different input devices were used: a 6-DOF wand for "go-go" and HOMER and a standard keyboard and a mouse for DO-IT. The results show that "go-go" and HOMER outperformed DO-IT, while immersivity had no influence on task completion time. These results were confirmed by [127], who compared three interaction modalities in a selection and manipulation task: real-world metaphor, "go-go" and raycast.

Finally, a well-known issue related to natural hand-based interaction is the hand-object interpenetration, due to the absence of real physical constraints between the virtual hand and the virtual object. Different grasping approaches have been investigated. The simpler one consists in the use of 3D models which simply follows the real hand and can interpenetrate objects. Other solutions are based on see-through methods showing the actual position of the fingers inside the objects or hand models constrained to avoid interpenetration or on the use of two hands, a visible one not interpenetrating objects and a ghost one directly following the real one. Finally it is possible providing different kind of feedback during collision, such as visual (object or fingers changing color), haptic (pressure, vibration) or auditory. Actually, many studies have shown that even if people prefer interpenetration to be prevented, described methods do not improve performance, while any kind of feedback seems to be a promising solution [73, 128]

2.3.3 *Feedback provided*

Of course, the choice of input hardware can not be independent with respect to the interaction technique considered, and viceversa. However, other important aspects, influencing this decision are the number of degrees of freedom needed to perform a task [142] and the kind of feedback provided to the user. For input devices and sensors, many different classification methods exist [6].

According to the measured response, they can be classified into *isotonic*, having zero or constant resistance and measuring movement, *isometric*, with infinite resistance (they connect the human limb and the device through force), and *elastic*, with varying resistance, which allows for movements and provides a counterforce [186]. Moreover, [22], define as multisensory input all instruments which combine elements from these groups.

If talking about haptic feedback, instead, systems able to simulate or provide haptic sensation can be divided in active (e.g. gloves), passive (e.g. controllers or mice) and pseudo, when haptic information is generated by the influence of another sensory modality, for example modifying the velocity of an object moved on a plan in order to generate a friction like sensation [22].

Furthermore, considering the level of touch involved, devices can be divided into: *touchful*, e.g. mouse or touchscreens, which require the application of a physical pressure to a surface; *touchless*, e.g. Leap Motion or cameras; *semi-touchless*, combining characteristics of the two previous families. Touchless input systems provide a wide range of input options, can replicate all the necessary DOFs, retrieve information while keeping sterility, be very useful to people with physical impairments and allow the navigation of big complex data volume through intuitive commands and gesture [50]. Nevertheless, they cannot provide haptic or force feedback.

Research on this field has been very active in the last decades: the goal is usually investigating the optimal interaction method to be used for task-specific scenarios, taking into account both quantitative parameters, evaluating performance, and subjective questionnaire, to investigate preferences. Time required to accomplish the task or the subtasks is often considered a reliable way to quantify performance [42].

In general, both for non-immersive and immersive VE, results highlight a contradiction between performance and preferences: while participants performed in general better with the touchful and semitouchless input techniques, they preferred using the touchless one. For examples, these results were obtained by [6], who compared participants performance in a "straight navigation" and a "rotation navigation" task in a 3D non-immersive virtual environment. In particular, they used the Leap Motion as touchless interface and two customized devices designed and developed by authors for the other two categories. Moreover, studies on non-immersive 3D virtual environments comparing the Leap Motion with other devices, e.g. the 3D mouse SpaceNavigator[11] in navigation tasks [161], or the optical mouse in selection and manipulation tasks [40], pointed out that Leap Motion outperforms other devices in simple tasks, while, as the complexity of the exercise improves, it turns out to be less accurate. Considering preferences, however, touchless interfaces are again the favorite ones. Furthermore, similar results were obtained by [153], who com-

11 https://www.3dconnexion.it/spacemouse_compact/it/

pare interaction with an RGB-D camera, the Point Grey Bumblebee2 [12], and with the SpaceNavigator during a navigation task.

Research has been prolific also regarding immersive VE. [80, 187] proposed a comparison between different input devices for navigation: a standard joystick, the Oculus Rift Touches combined with a teleport method and Leap Motion. Of course, the teleport took the least amount of travel time, whereas the joystick and the Leap Motion had comparable results. Moreover, in both studies, people preferred Leap Motion, in terms of immersion, comfort, intuitiveness, low fatigue, ease of learning and use. Similar results, showing contradictions between performance and preferences were obtained by [152], who compared a controller based interface with a realistic interface, composed by a steering wheel and the Leap Motion, in a driving simulation application.

2.4 User Experience Assessment

Two important aspects to be considered when implementing a VR system for simulation and training are *cybersickness* and *sense of presence*

2.4.1 *Cybersickness in VR*

One of the most common problem related to the use of VR is *cybersickness*, defined as a state of malaise and unpleasant side effects associated to the use of immersive simulations and affecting 60-80 % of the users. *Cybersickness* is a type of Visually Induced Motion Sickness (VIMS) that stems from exposure to VR. Although its minor, short term health risk, it is still a potential issue for the broader adoption of this technology, [117]. Typically, *cybersickness* varies between individuals but common symptoms are nausea, eyestrain, dizziness, apathy, sleepiness, disorientation, fatigue and general discomfort, which can cause cognitive impairment and negatively affect user's performance while accomplishing a task [51, 52, 103, 117]. It can occur immediately after training or even up to 5-12 hours later [83, 114].

Cybersickness appears to be subjective, however, works done in this area shaded light on a few of factors influencing it, including individual differences, device feature and task characteristics [88]. Among individual factors, we can find genre, age and health state. In particular, women appear to be more susceptible to *cybersickness* than men; susceptibility is higher between 2 and 12 years of age, decreases rapidly from 12 to 21 years and is almost non-existent around 50 years; people suffering from illness, fatigue, sleep loss, hangover, upset stomach, periods of emotional stress, head colds, flu, ear infection, or upper respiratory illness should avoid using VE simulators. Device factors influencing *cybersickness* are mainly related to displays and technologies issues, for example, position tracking errors causing jitters, lags, due to the time required to send information from a head tracker to the computer, have this

12 https://www.flir.com/iis/machine-vision/stereo-vision

information processed and update the visual display, and flickers, especially in the periphery of our FOV, caused by refresh rates lower than 30Hz.
Finally, also position during the simulation play a role in the individual's susceptibility to *cybersickness*: sitting appears to be the better position in which to reduce malaise since it would decrease the demands on postural control. Moreover, analogously to car sickness, in which passenger gets car sick while driver does not, during multi-player simulations, users who control the simulation are less susceptible to *cybersickness* than those who have a passive role. Thus, ensuring an active control over the exploration of VEs, reduce *cybersickness* [146].

Causes of *cybersickness* are still over debate, but three prominent theories are: poison theory [21], postural instability theory [137] and sensory conflict theory [132]).

The poison theory suggests that physiological effects of the visual, vestibular and other sensory input systems on coordination caused by the ingestion of poison act as an early warning system which enhances survival by removing the contents of the stomach. Thus, the adverse stimulation found in some VE can effect the visual and vestibular system in such a way that the body misreads the information and thinks it has ingested some type of toxic substance thus causing disturbing symptoms which lead to an emetic response.

The postural instability theory is centered on the idea that one of the primary behavioural goals in humans is to maintain postural stability, minimizing uncontrolled movements of the perception and action systems. Whenever the environment changes in an abrupt or significant way, postural control will have to adapt and eventually learn new control strategies. If postural control is degraded or completely lost, people are in a state of postural instability. For example, a subject may use muscular force to resist the tilt of an angular acceleration which is visually perceived during a virtual simulation, but as there is no physical tilt the subject creates an unintended divergence from a stable position causing postural instability. According to postural instability theory, postural instability precedes and is necessary to produce *cybersickness* and the severity of the symptoms scales directly with the duration of the instability.

The sensory conflict theory, finally, is the oldest and the most accepted one. It is based on the premise that discrepancies between visual and vestibular systems cause a perceptual conflict which the body does not know how to handle. For example, in the case of a virtual driving simulator, during the simulation, the optical flow patterns of roads, buildings and other parts of the VE move past the subject's periphery, giving him a sense of vection. The visual system tells the subject he is moving in a certain direction, whereas, the vestibular system provides no sense of linear or angular acceleration or deceleration.

Cybersickness can be quantify using both subjective and objective measures, which were listed by [76, 117]. Subjective scales of evaluation include susceptibility questionnaires (Motion Sickness Susceptibility Questionnaire [63], Reason and Brands Motion Sickness Susceptibility Questionnaire [132]), online reports, usually composed by a symptom or question that participants are asked to rate multiple time during the simulation in order to detect runtime onset, course, severity and trend of VIMS (Fast MS Scale [76], Misery Scale Index [18], Short Symptom Checklist [118]) and standard questionnaires usually filled in before and after the trial, where user is asked to rate the severity level of different symptoms (Simulator Sickness Questionnaire [75], Motion Sickness Assessment Questionnaire [63], Pensacola Motion Sickness Questionnaire [93], Nausea Profile[115]). Simulator Sickness Queationnaire (SSQ), in particular, allows to analyze the participants' status before and after each exposure to VR in order to estimate if the experience inside a virtual environment had some physical effects. It is composed by 16 questions evaluating general discomfort, fatigue, headache, eye strain, blurred vision, dizziness with eyes opened or closed, difficulty focusing or concentrating, fullness of head, vertigo, sweating, salivation increase, nausea, stomach awareness, burping. These 16 aspects are then grouped in three category: Nausea, Oculomotor, i.e. visua-related disorders, and Disorientation. The questions are in a 5-points Likert scale, rating from 0 to 4 how severe a symptom is.

Objective measures, instead, include physiological measurements, some of which have been proven to be correlated with VIMS [52]. [116] studied the effect of *motion sickness* on thermoregulation, using provocative visual stimuli (a VR rollercoaster). During immersion, there is an initial phase of vasoconstriction, due to a defense response associated with arousing effects of the simulated ride, followed by a vasodilation, related to *cybersickness*. Vasodilatation causes heat loss through sweating, an increase of skin conductance, skin warming and tachycardia, which is related to the activity of the sympathetic nervous system, as a defensive reaction against the sensation of nausea [120]. Another study conducted by [83] demonstrated a significant positive correlation of *cybersickness* severity with gastric tachyarrhythmia, increase of eyeblink rate and EEG delta wave and decrease of heart period. Finally, also analysis of sway and center of pressure (COP) could be informative [1, 114]. In fact, body sway differs between participants who report motion sickness and those who do not.

2.4.2 *Sense of presence in VR*

Sense of presence or *spatial presence* is defined as the psychological state where virtual experiences and computer-generated environments feel authentic rather than the actual physical locale, or, in other words, the sense of "being physically there" [70, 102, 147, 179].

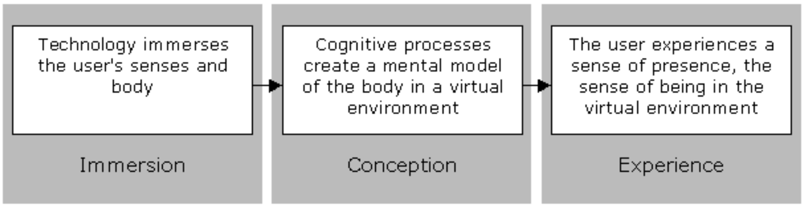
Unlike *immersion*, which depends essentially on the type of technology used and can be objectively described, *sense of presence* is primarily subjective and is linked to the user experience. In particular, some theories asses that *spatial presence* is determined by how efficiently we mentally process the spatial relations within the environment [178] (Figure 6 a). So, persons able to process spatial arrangements effectively will find it easier to create a "mental model" of the spatial environment, thus they will experience a higher *sense of presence*. Authors in [46] proved that self-reports of imagery are positively correlated with reports of *spatial presence*, but *spatial presence* itself is not related to performance.

Although, its subjective nature, there are many factors influencing *spatial presence* [163]: the degree of interaction user has with the virtual environment; the proper implementation of an action/perception loop; the high resolution of information displayed, in a manner that it does not indicate the existence of the display; the consistency of the displayed information across different sensory modalities; the presence of a first person avatar, as a self-representation of the user in the virtual world, which should be similar in appearance or functionality to the individual's body.

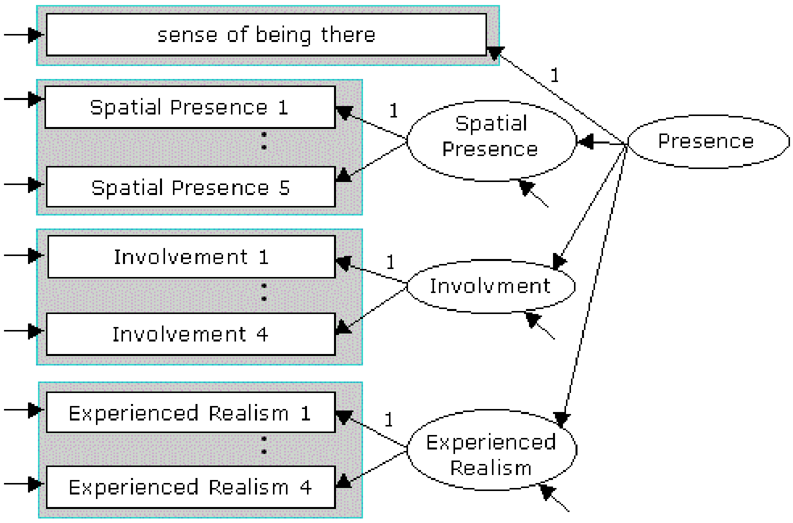
Interaction and *sense of presence* are two concepts strictly related. In fact, the presence of the user in the virtual world implicitly implies his ability to act in it, interact with it and have an active role. So, the more realistic, natural and intuitive an interaction is, the higher the degree of presence is. At the same time, when the *sense of presence* is enhanced, for example reproducing user hands in the VE or allowing him to receive coherent and consistent multimodal feedbacks (e.g. visual, haptic/tactile, proprioceptive, vestibular, audio), also interaction improves.

Standard questionnaire commonly used to measure *sense of presence* are the IGroup Presence Questionnaire [134], the Presence Questionnaire [179] and the Slater, Usoh and Steed (SUS) Questionnaire [151, 162]. Igroup Presence Questionnaire (IPQ), in particular, is composed of 13 7-points Likert scale questions evaluating three different aspects (Figure 6 b): Spatial Presence; Involvement, intended both as attention during the interaction with the virtual world and as perceived involvement; Experienced Realism, i.e. perceived realism of the VR experience. An additional question rates the sense of presence from the original definition, as the sense of being physically "there" in the VE.

Moreover, user head rotation and position can be considered as objective measures of subjects behaviour during the navigation of a VR environment and of their tendency to explore the surrounding environment in a natural way [35].



(a)



(b)

Figure 6: (a) From immersion to sense of presence. (b) IPQ scales of evaluation.

3

Research Work

The following chapters has been divided in five main parts. The two first sections describe the studies specifically designed, implemented and carried out to understand how people perceive and process information coming from the VE and which is the best interface in terms of transparency, able to guarantee a minimum cognitive load on the user so that he can concentrate on the current task. In the other sections, instead, three use cases are considered: the evaluation of a system for ship handling simulation; the development of a game for cognitive research specifically design to elicit particular emotions; the development of an application for the cognitive assessment of elderly.

3.1 Attention and working Memory Capacity in VR

The crucial idea subtending any application for simulation and training is that users must be able to process and memorize information presented in the VR simulated environments and interact with them. As stated before, during active learning trainees take advantage of the action/perception loop in order to improve their knowledge and experience of the world. This is true both for real-life settings and Virtual Reality scenarios. In particular, the mechanism responsible for the processing of information to be used for ongoing tasks, for example change detection, integration of information or planning of reaching movements, is the Visual Working Memory. However, this cognitive system is limited both in capacity, i.e. the number of data which can be remembered, and time, as it decades in few seconds unless recalls occur.

The research described in this chapter mainly focuses on the evaluation of the bandwidth of human perception and VWM capacity in VR, where stimuli can be displayed in a 360 degrees space surrounding the user. The final goal consists in the implementation of more effective VR applications for training and simulation. In particular, we aim at understanding if people ability to gather and process information presented in an immersive virtual environment is severely limited or not. In the first case, in fact, it would be pointless designing and implementing rich and complex scenarios, as users would not be able to perceive and elaborate all the displayed data; whereas, in the second case, it would be interesting investigating conditions and factors which can positively influence and increase subjects VWM capacity. Software framework, methodology and results described below are part of the work presented in [12] and of an article, which is still in preparation and will be submitted soon. This phenomenon has actually drawn researchers' attention from the 1990s, but the majority of studies found in the literature use 2D pictures and videos. Such kind of stimuli can not be compared to the variety and richness of visual input that people continuously receive from the real world surrounding them. Virtual Reality, instead, could represents a valid alternative as it would allow to conduct experiments in a fully controlled 3D environment, diversify levels and combinations of multimodal sensory input, generate a variety of response options, manipulate the virtual world and the different experimental parameters, acquire measurements and explore the effect of visual cues, such as stereopsis, motion parallax and depth perception, and pictorial cues, as occlusion, shading and perspective, on Change Blindness. Moreover, thanks to game engines, like Unity 3D [1], it is possible to generate more or less complex scenarios, at more or less high semantic level, having the possibility to examine sophisticated participants behaviours and reproduce situations which could be impractical, dangerous or ethical questionable [177].
Works conducted in VR, mainly investigate human perception limits in dynamic context or in dual task experiments. We have, thus, designed and im-

1 https://unity.com/

plemented a systematic approach for the study of Working Memory capacity through a *whole-display* change detection *one-shot* paradigm in an immersive Virtual Reality environment, exploiting the *intentional* CB induction. We devised two experiments: in Experiment 1 or *change detection* test, participants have to recognize if a change has occurred in the presented stimulus; while in Experiment 2 or *change localization* test, we asked them to indicate if a change has occurred and where it was located. The second task, requires more precision and should force people to be more concentrated during task execution and remember more details of the scene. It is interesting, in fact, evaluating if and how the kind of task can influence the bandwidth of human perception.

3.1.1 *Material and Methods*

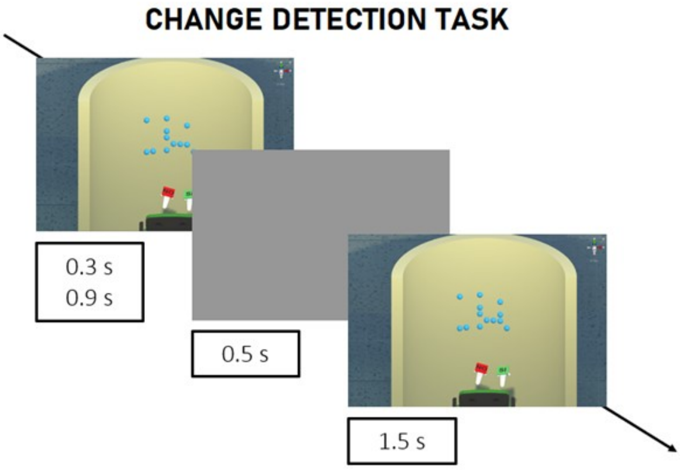
3.1.1.1 *Scenario*

The experimental setup is composed by the HTC Vive headset for visualization, the HTC Vive controllers for interaction and an Alienware Aurora R5 with a 4GHz Intel Core i7-6700K processor, 16GB DDR4 RAM (2,133MHz) and a Nvidia GeForce GTX 1080 graphic card. In this case, we do not use a chin-rest, also because in some cases head rotations are required. However, participants are asked to maintain their head in a specific position, indicated by a white sphere, for the entire duration of the test, otherwise the experiment was restarted.

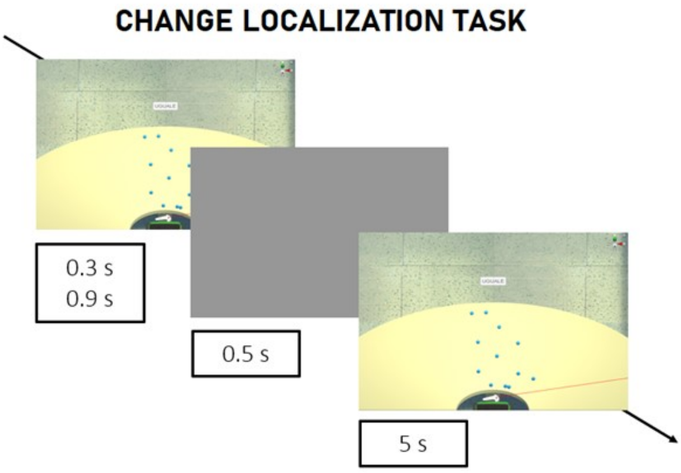
An ad hoc virtual environment, which was used for both the *change detection* and *change localization* tests, has been implemented. The scenario is voluntarily simple and essential: a bare room, with no windows or pieces of furniture except for a green chair and a semi-circular desk or a semi-cylindrical wall, on which items were generated. The virtual chair facilitates the user in locating himself in the correct initial position. Moreover, stimuli consist in light blue spheres with a diameter of 5 cm. The simplicity of the scene is deliberately chosen in order to prevent any interest prioritization based on stimulus features (shape, colour, luminance etc) or on the semantic of the scene.

In order to replicate the *one-shot* paradigm in VR, stimuli are shown on the table or on the wall for an observation time T1; then, a grey canvas covers the entire participants view for a retention time of 0.5 s; after this, spheres are visible again for a time T2; finally, all objects disappear and after 2 s a new trial automatically starts (see Figure 7). The blank colour has been chosen to avoid after-images and not annoy participants, as stimuli and blanks alternate in a rapid sequence; whereas, the ISI duration is long enough to overcome any image persistence on the Dual AMOLED headset displays without impairing memory.

VWM capacity has been measured at the variation of five different conditions: the kind of change (appearance, disappearance and no change), the set size in the original scene (4, 6, 8, 10, 12), the objects distribution, both in terms of spatial layout (vertical with linear distribution or horizontal with perspect-



(a)



(b)

Figure 7: One-shot paradigm in the (a) *change detection* and (b) *change localization* tests.

ive cues) and of occupied horizontal Field of View (40, 80 and 120 degrees), and the observation time (0.3 and 0.9 s).

As stated before, appearance and disappearance, later referred to as Add and Remove, have been found to be easily and fast recognized [69], but they apparently activate diverse mechanisms for detection and prioritization of attention [26], influencing results. In particular, appearance of an object is faster identified, as, for evolutionary reasons, we are more prone to detect threats entering our FOV. The no change condition (Equal), instead, is used as a control condition.

The number of items to be tested has been set after a series of pilot studies, in which the percentage of correct answers tended to decrease around 8 or 10 elements and a strong effect of subitizing over performance was evident up to

4 item. Subitizing is a well-known concept in the literature and is defined as the direct perceptual apprehension of the numerosity of a group and a necessary precursor to counting. This judgement over the number of items in a group have been proven to be fast, in the order of 30 ms for each extra element [158], accurate and confident for small set sizes, up to 4, and decrease for larger set size, where counting or size estimate is required.

Considering the two layouts (Figure 8), in the vertical case (V), we simply replicates the standard 2D experiments conditions but with the addition of a third dimension, as we use spheres as stimuli instead of dots; while in the horizontal case (H), it is possible to better evaluate the influence of visual and pictorial cues, e.g. stereopsis, depth perception, possible occlusions and perspective, over change detection performance. Moreover the two layouts correspond to different spatial distributions of objects: linear for the vertical one and perspective for the horizontal one.

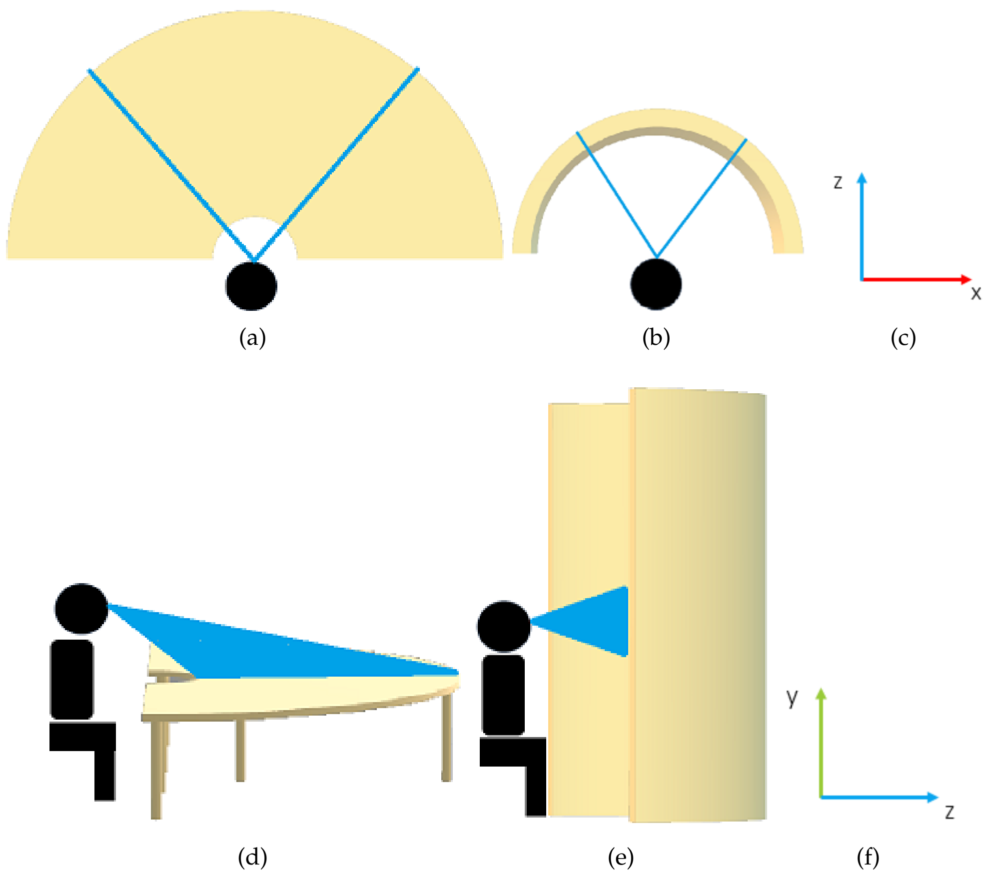


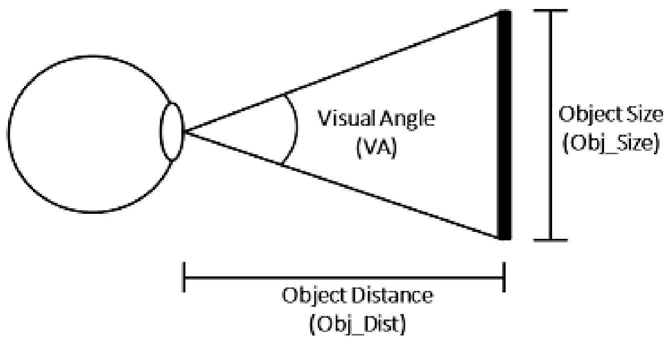
Figure 8: Area of distribution of items as in the VE in the horizontal (a,d) and vertical layouts (b,e). (c, f) Reference axis in Unity 3D.

Equation 1 describes the relation between Visual Angle (VA), i.e. the portion of the Visual Field that an object occupies, size and distance of the item (see Figure 9). Once fixated the head position, thus its distance from the object, and the desired Visual Angle, which is 30° vertically and 40°, 80,°, 120° horizontally, it is easy to define the area where spheres have to be spawned, by using the inverse Equation 2.

$$VA = 2 * \arctan(\frac{Obj\_Size}{2Obj\_Dist}) \tag{1}$$

$$Obj\_Size = \tan(\frac{VA}{2})2Obj\_Dist \tag{2}$$

$$Obj\_Dist = \frac{Obj\_Size}{2\tan(\frac{VA}{2}))} \tag{3}$$



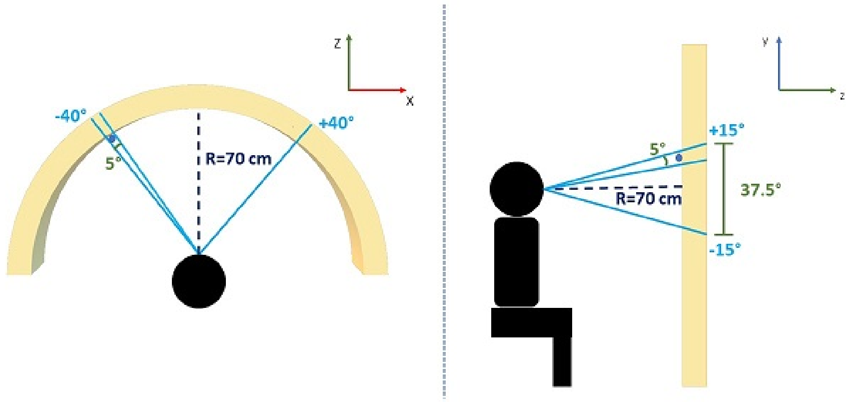
(a)

Figure 9: Schematic representation of the Visual Angle and of the variables involved in Equation 1.

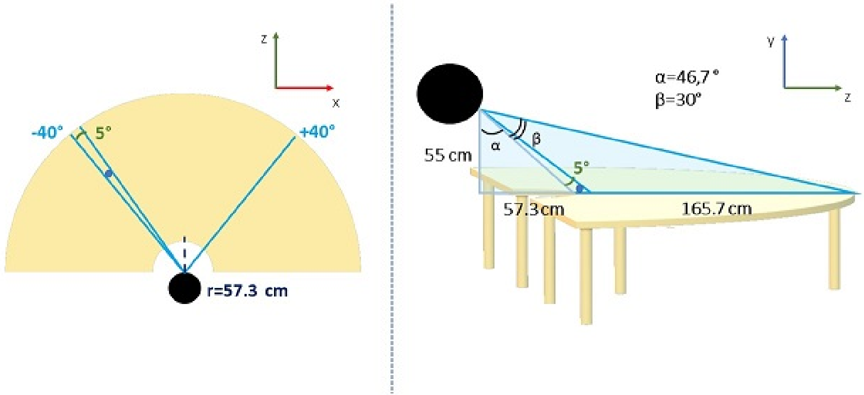
In both cases, an $6 \times n$ grid has been used. The horizontal FOV, in fact, has been divided in 5° large bins, symmetrically arranged with respect to the subject head, from -20° to +20°, or from -40° to +40°, or from -60° +60°. The light blue spheres used as input, were placed at the centre of the bin, from -17.5° to +17.5°, or from -37.5° to +37.5°, or from -57.5° +57.5° (Figure 10 a and b left), respectively. The vertical FOV, instead, has been defined differently for the two layouts.

Starting from the vertical layout, first, we have fixed the head position, indicated by a white sphere, which is 70 cm from the semi-cylindrical wall. Then, 6 heights have been defined, again symmetrically with respect to the head position and occupying a total space of 37.5 cm, which is the size of an object covering 30° of the VA at a distance from the eye of 70 cm (Figure 10 a right). Spheres in both layout cases have a radius of 2.5 cm, in order to occupy around 4° of the (VA) and avoid overlapping between nearby objects. In the horizontal case, the surface where items are spawned is not orthogonal to participant eyes, thus the vertical FOV is conceived in a different manner. We have first calculated the minimum distance at which spheres occupy around 4° of the (VA) without overlapping and found it is 57.3 cm. Then, using basic geometry, we have calculated maximum distance subtending an angle of 30°, starting from the minimum distance and considering that head position was fixed 55 cm above the table (Figure 10 b right).

The choice of gradually enlarging the FOV (see Figure 11), is due to the



(a)



(b)

Figure 10: Spatial distribution of item in the (a) vertical and (b) horizontal layout cases from the top (left) and lateral (right) point of view. FOV 80° is used as example.

fact that experiments found in the literature are in general referred to stimuli presented in a space occupying 30°-40° of the Visual Angle, as described in Section 2.2, while Virtual Reality offers the potential to exploit a 360° world. However, the direct presentation of stimuli distributed in larger FOVs would generate difficult to interpret results, because of the influence of different factors, related both to the binocular vision and to head movements and visual integration. Thus, experiments with the vertical layout and FOV 40° are the exact VR counterpart of those with 2D images, where only tridimensionality is added [69, 97, 125]. Stimuli are presented within the near peripheral vision area, where visual acuity is high and humans are more sensitive to colours and shapes. FOV 80°, instead, refers to the mid-peripheral vision, more sensitive to motion signals, and corresponds to exploiting entirely HTC Vive's Field of View, which namely is 110 degrees, but, due to lenses distortion, is actually limited to 90°. It is worth noting that because HMDs have a 90° FOV, which is far from being "natural", as human horizontal FOV is around 220°, we are well aware that results discussed below can be only referred to VR setups and we do not mean to demonstrate that they are extendible to real-world settings. Finally, with FOV 120° a motorial component is introduced, as participants have to turn their head to get a complete overview over the presented stimuli. We are, in fact, interested in understanding WM performance when the internal representation of the scene has to be fast updated and integrated across different views.

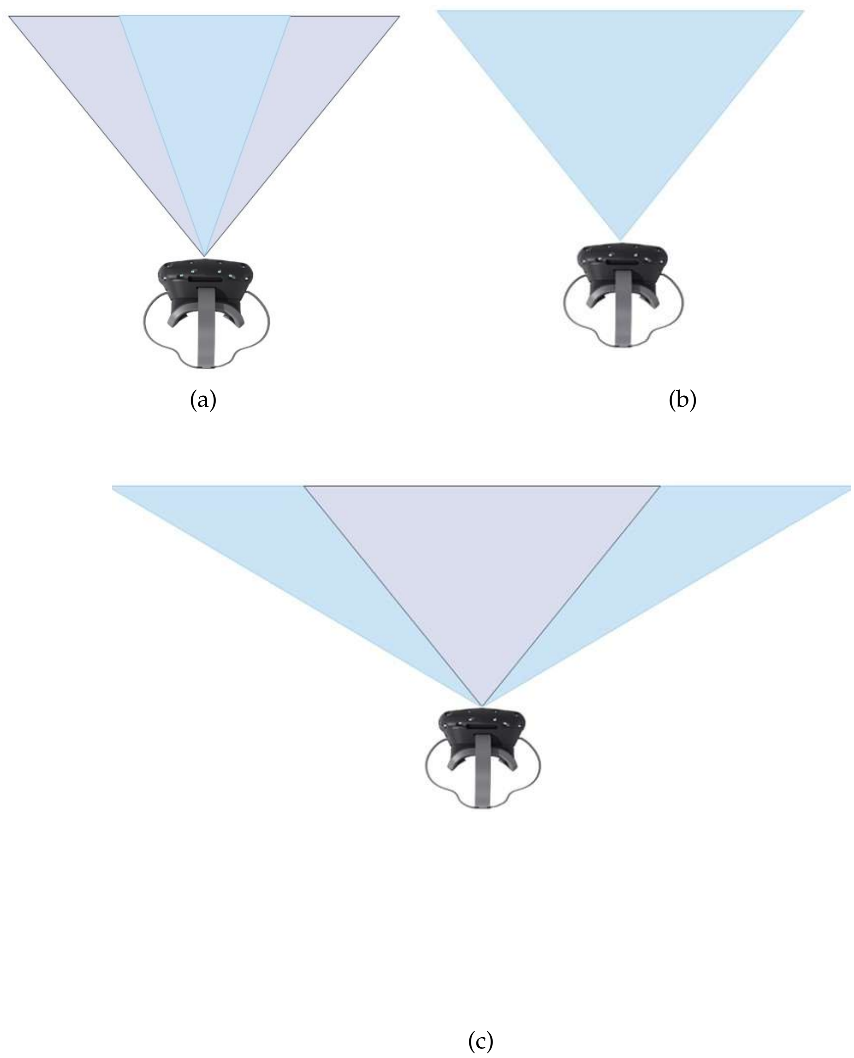
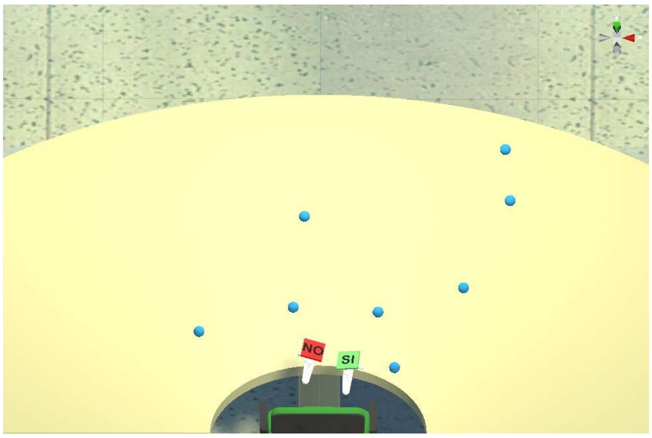


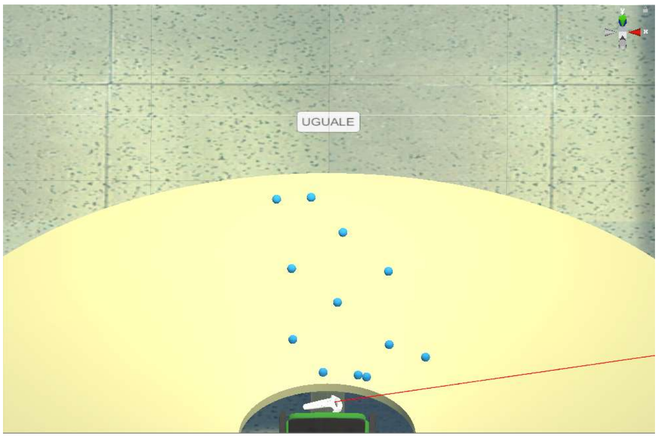
Figure 11: In light blue, proportion of the FOV in which stimuli are displayed (40° (a), 80° (b), 120° (c)) with respect to the actual HTC Vive's Field of View, in pink.

Finally, observation times are chosen considering both the previous literature. A 0.3 s (later indicated by S) observation time, in fact, is enough for participants to generate the gist of the scene, at least in the FOV 40° and 80°

scenes. It does not avoid subitizing but prevents participant from voluntarily counting spheres. In the FOV 120° case, instead, participants have to turn their head to have an entire overview over the presented stimuli. Obviously, a 0.3 s observation time precludes this action, and a higher observation time is necessary. Considering that the Field of View has been tripled, we have therefore decided to increase the observation time proportionally to 0.9 s (later indicated by L).



(a)



(b)

Figure 12: Example of interaction during test phase in the *change detection* (a) and the *change localization* tests (b).

3.1.1.2 *Parameters*

In both tests, for each trial, the given answer, the reaction time, i.e. the time taken to respond, the spheres' distribution and head positions and rotations are recorded. In the *change detection* task, the answer is a value set to 1 for

"Yes", 0 for "No" and 2 when participants could not answer on time; whereas in the *change localization* task the raycasted position in the 3D virtual space is acquired. When subjects report no change or decide not to respond, the cartesian coordinate were respectively set to "100 100 100" or "0 0 0".

We excluded subjects who demonstrated not having understood instructions properly, i.e. they did not answer on time in the majority of trials or showed a random trend in the percentage of correct answers per set size. In the second test, the correct answer was considered as the action of raycasting in trials with change and pressing "Equal" button in case of no change.

Raw data were, then, elaborated and analysed, in order to understand the influence of the number of items, the kind of modification, the spatial layout, the FOV and the observation time over change detection performance, thus over WM capacity.
Common metrics for the evaluation of performance in *one-shot* change detection tasks are the hit and false alarm rate. In the first case, shown in Equation 4, we consider the number of times people correctly reported a change (TP) divided by all trial where a change was applied (TP+FN).

$$HR = \frac{TP}{TP + FN} \tag{4}$$

At first, trials with appearance and disappearance were grouped together, while at a later time, they were considered separately.
False alarm rate, shown in Equation 5, instead, is calculated as the number of time participants perceive a change in the control condition (FP) divided by all the trials without any change (FP+TN).

$$FAR = \frac{FP}{FP + TN} \tag{5}$$

These two parameters quantify how accurate participants are or how random their answers are. In particular, from previous literature [77], we expect high hit rates and near-to-null false alarm rates when the number of items to be remembered does not exceed Working Memory capacity [77]. Once this capacity is overcome, instead, people should start answering in a randomic way, hit rates should decrease while false alarm should raise.

We furthermore analysed the horizontal and vertical hit rate distribution with respect to the distance from the centre of the Field of View. In this case, we want to evaluate if the absolute position of the modified object, independently from its relative position with respect to the other spheres, could influence participants ability to detect a change.

Then, we analysed the reaction time. We hypothesised an influence of the kind of change and number of items over performance. In particular, we expect appearance to be recognized faster than disappearance and no change condition to be the slower one, because of a longer unsuccessful search for change. Moreover, considering the number of spheres, when set size exceeds Working Memory capacity, a slower item by item search of the change is necessary, reducing reaction time.

<table>
<tr><th>Change detection</th><th>Change localization</th></tr>
<tr><th colspan="2">Conditions</th></tr>
<tr><td colspan="2">- Kind of change (Add, Remove, Equal)
- Number of items (4, 6, 8, 10, 12)
- Spatial layout (V and H)
- FOV (40, 80, 120)
- Observation time (0.3 and 0.9 s)</td></tr>
<tr><th colspan="2">Parameters</th></tr>
<tr><td>-HR and FAR
-HR and FAR spatial distribution
-Reaction Time</td><td>-HR and FAR
-HR and FAR spatial distribution
-Reaction Time
-PCA with raycast positions</td></tr>
</table>

Table 1: Summary of conditions considered and parameters calculated in the *change detection* and *change localization* tests.

After having ascertained that the majority of data samples are not normally distributed, through the Kolmogorov-Smirnov test, the Wilcoxon test has been used in order to understand if differences eventually found were statistically significant.

In the *change localization* test, we finally used principal component analysis (PCA) to evaluate if there is a preferential distribution of raycasted position around the target. We expected to see different distributions of pointing errors in the vertical and horizontal case, with more symmetry around the target in the first case and an influence of distances in the second case.

3.1.2 *Experiments*

Combining all the conditions previously described, we obtain 180 combinations per test with 6 repetitions per combination, which were divided in 12 blocks of 90 trials each. In each block, the spatial layout, the FOV and the observation time were fixed, while the number of items presented in the original scene (from 4 to 12) and the kind of change varied. In particular, one third of the trials was with appearance, one third with disappearance and one third without change. In order to avoid a negative effect of fatigue over performance, the blocks were divided in 3 session of 4 blocks each. We designed a counterbalance repeated measures experimental design, i.e. each subject completed the test with all conditions combinations, but block presentation was randomized between subjects. Moreover, the scenes were randomly generated, so that each subject saw different spheres distribution for each trial.

3.1.2.1 *Change detection test*

In the *change detection* test, participants have to answer the question *"Has something changed in the scene?"* by pressing the trigger button of the correct controller: one is marked with a "Yes" label and the other with a "No" label (Figure 12 a). Subjects can decide in which hand they want to handle the two controllers. They have 1.5 s to answer and, once time is over or the answer was given, all spheres disappear for 2 seconds, then a new scene is automatically presented (Figure 7 a). Participants are encouraged to always answer as accurately as possible. Trials without answer are excluded from further analysis. A demo scene precedes the actual trial, so that participants can familiarise with the interface and the task.

Eighteen subjects accomplished the task. They were all students, PhDs and researchers at University of Genoa (age range 20-36, 25.1 $\pm$ 3.8 years). They all reported having normal or corrected-to-normal vision and no deficit in stereo vision. Half of them were naive with respect to the purpose of the study but were familiar with the VR headset used.

3.1.2.2 *Change localization test*

In the *change localization* test, the question is: *"Where was the change, if any?"*. In this case, participants use only one controller and have three options (Figure 12 b): indicating where a sphere have appeared or disappeared through raycasting; pressing the "Equal" button if they have not perceive any change; waiting 5 seconds, if they have perceived a change but can not indicate where (Figure 7 b). In order to ensure a better stability of the answer, position is acquired on release, in fact, the pressure of the controller's trigger button provokes a natural oscillation of the wrist. Despite the limited amplitude of this oscillation, noise could affect measurements, especially when targeting far positions.
Again, a demo scene precedes the actual trial, so that participants can familiarise with the interface and the task.

Twenty-one students, PhDs and researchers at University of Genoa (age range 20-36, 25.3 $\pm$ 3.7 years) took part to the experimental session. They all reported having normal or corrected-to-normal vision and no deficit in stereo vision, they were naive with respect to the purpose of the study but were familiar with the VR headset used.

3.1.3 *Data Analysis*

3.1.3.1 *Change detection test*

Data collected were organised by blocks, obtaining 12 groups based on the layout, the FOV and the observation time.

First we analysed the hit rate, calculated grouping appearance and disappearance together, and false alarm rate, both averaged across subjects. In all groups of trials the first parameter tends to decrease as the number of items increases (see Figure 13 a). Concurrently, standard deviation increases, indicating a higher uncertainty of answers. Repeated two-sided Wilcoxon tests highlight that hit rates referred to 4, 6 and 8 items are statistically different ($p<0.02$) from those referred to 10 and 12 items, exception made for the trials with FOV 120°, vertical layout and short observation time, in which statistically significant differences have been found only between hit rates referred to 4 and 6 items with higher set size. In general, HR referred to 4 or 6 items are very similar to each other and differ from those obtained with 8 items ($p<0.02$). This would confirm a limit of human Working Memory capacity around 7 items. False alarm rates, instead, are usually constant in all trials, even if they slightly increase with the number of items. However, larger standard deviation highlights a higher variability of data and makes differences not statistically significant. Only differences referred to trials with FOV 80°, 0.9 observation time, horizontal layout (4 and 12 items, 8 and 12 items) and vertical layout (4 and 12 items) are statistically significant ($p<0.02$).

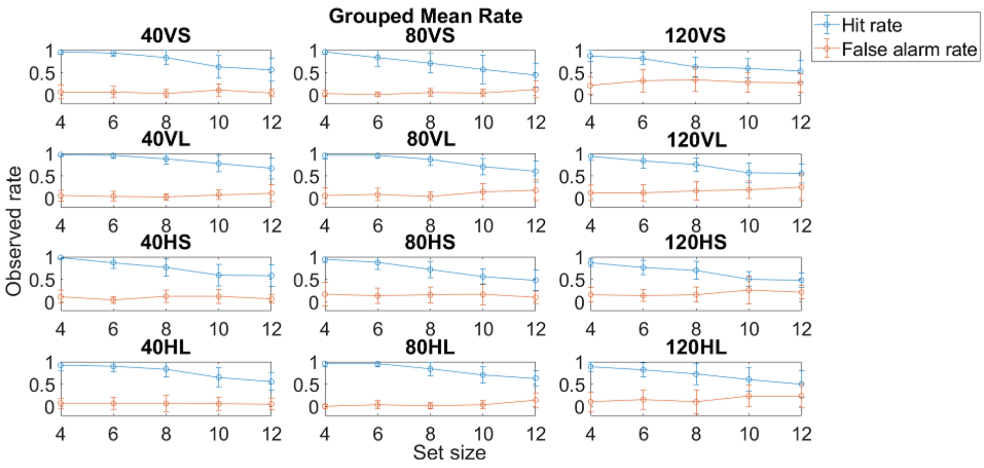
Considering the two metrics, it is possible to notice that better results are associated to the longer observation time, even if these differences are not statistically significant; while layouts seem having no influence. Having more time to observe stimuli, in fact, with FOV 40° and 80° subjects can build the gist of the scene but also try to memorize some structure or spatial distribution of elements; whereas with FOV 120° they can rapidly turn their head and have an impression of the entire scene.

Nonetheless, results in the FOV 40° and 80° trials are, in general, comparable, maybe because stimuli were all in the participants Field of View and no motorial component was involved; whereas, in the FOV 120° trials, hit rates decrease faster and false alarm rates are higher, independently from the observation time length. Also standard deviation is larger, indicating more uncertainty on given answer. Lower performance associated to trials with the largest FOV and the shortest observation time is due to the fact that participants did not have enough time to have a complete overview over the presented stimuli and actually had the possibility to see only two third of the scene, biasing answers, reducing the hit rates and increasing false alarm rate. Considering, instead, the longer observation time, even if participants could see the entire scene, they had to efficiently integrate multiple views in a brief time (0.9 s). Difficulties in image integration could be the reason of the worst results with respect to trials with other FOVs.

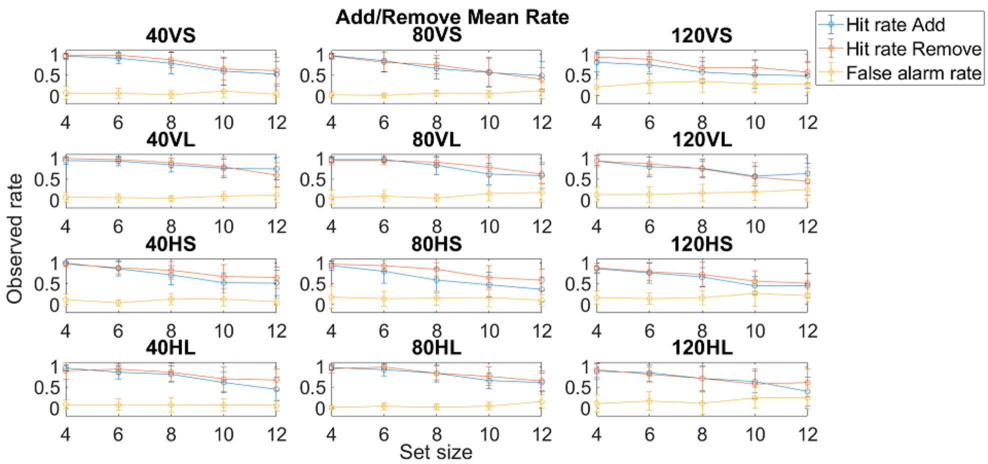
Differences between FOV 40° and 120° false alarm rates, in all conditions, and FOV 80° and 120° false alarm rates, with the horizontal layout and long observation time and with the vertical layout and short observation time, are statistically significant ($p<0.02$).

These findings are confirmed by the analysis of hit rates referred to the two separate kind of change. Although HR, is generally better when an object is removed, differences between appearance and disappearance are not statistically significant (see Figure 13 b).

In order to understand if the probability to correctly detect changes was associate to the absolute position of the modified item in the Field of View, we defined a threshold for hit rate of 0.75 and we excluded from each block trials



(a)

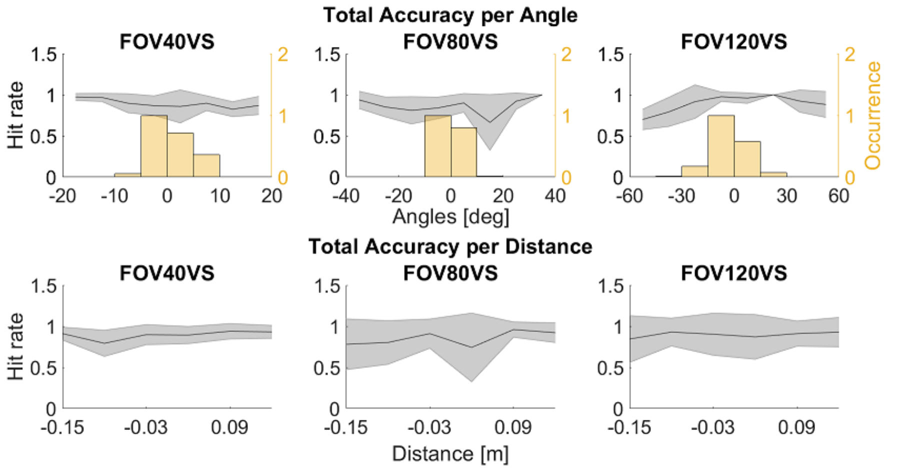


(b)

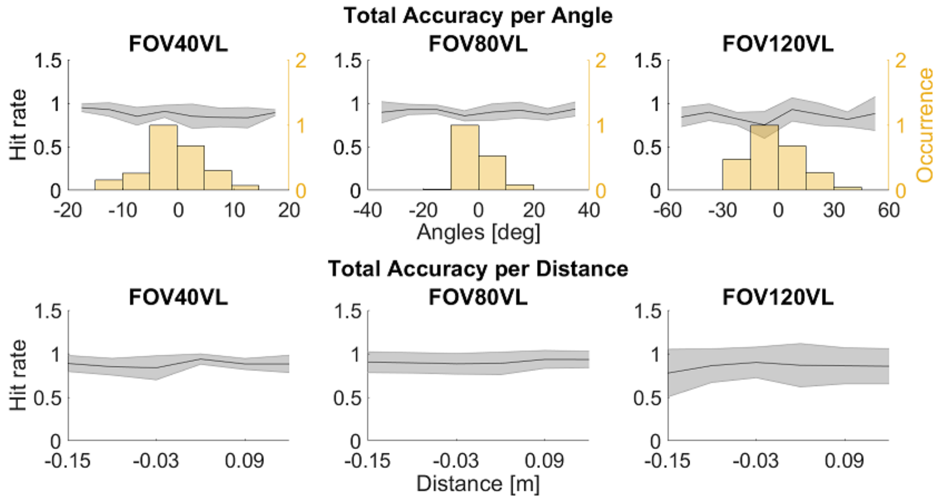
Figure 13: *Change detection* test: mean and standard deviation of hit rate and false alarm rate in the different blocks considering appearance and disappearance together (a) or separately (b). V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

referred to the number of item, for which the hit rate falls below the established threshold. We then calculated the hit rate referred to each possible position of the $6 \times n$ grid, and mediate first on the rows and then on the columns. This way, we consider separately the mean hit rate distribution at the variation of the horizontal angular distance from the FOV's centre and of the height or ray distance, respectively in the vertical and horizontal layout case. As each position has an equal probability to be selected to spawn the modified item, in the horizontal FOV analysis we have considered bins of different size, 5° for FOV 40°, 10° for FOV 80°, 15°for FOV 120°.

Results in Figure 14 and 15 show that, in general, performance is better and with a lower variability when a longer observation time is provided; while, fixating the observation time, it decreases with FOV enlargement. No particular trends related to absolute spatial position of the modified object has been



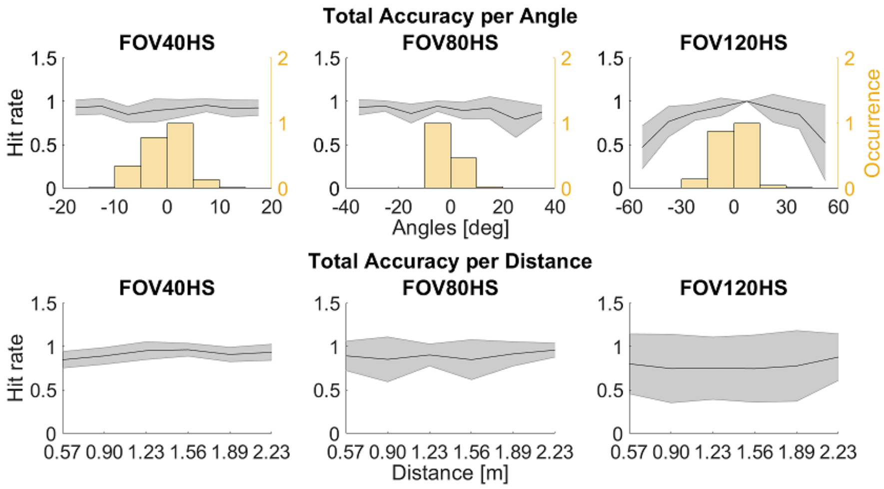
(a)



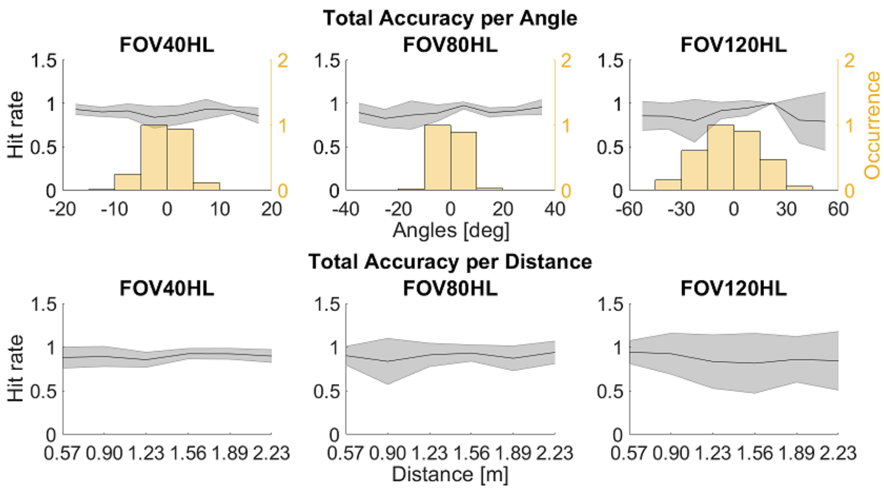
(b)

Figure 14: *Change detection* task: mean and standard deviation of the hit rate in the different blocks at the variation of the horizontal angular distance from the center (top) and of the height (bottom). In the first case, together with the angular variation, a normalized histogram of yaw head rotations is shown. V = vertical; S = 0.3 s observation time; L = 0.9 s observation time.

found, both for the lateral and vertical Field of View, except for the FOV 120° with short observation time case. In Figure 14 and 15 in the top raw, together with the hit rate distribution inside the horizontal FOV, a normalized histogram of head rotations around the vertical axis during the observation of the original scene is shown. Histograms are indicator of participants tendency to look around and explore the scene. As expected, with FOV 40° and 80°, rotations are clustered around a central value, because all stimuli are presented inside the visual field. Also in the FOV 120° trial with short observation time (Figure 14 a and 15 a) head rotations are limited, in fact hit rates decrease in the periphery; whereas, having enough time to explore the entire scene,



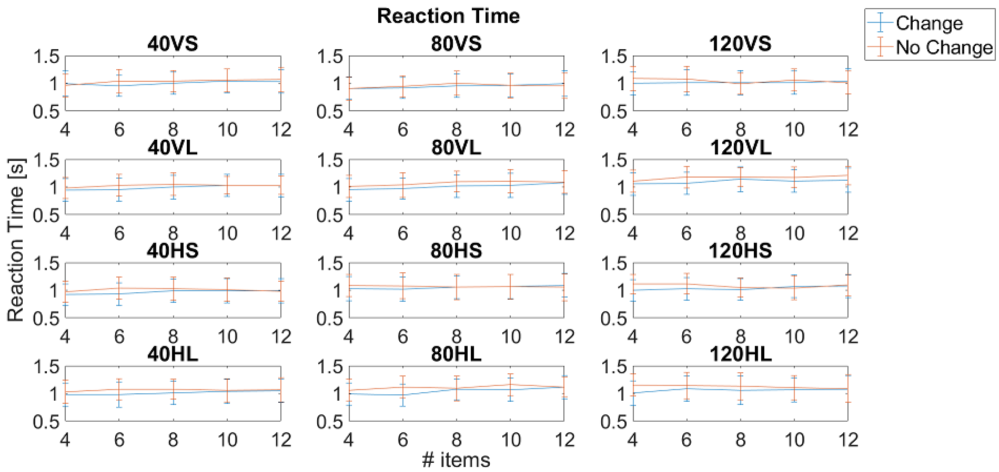
(a)



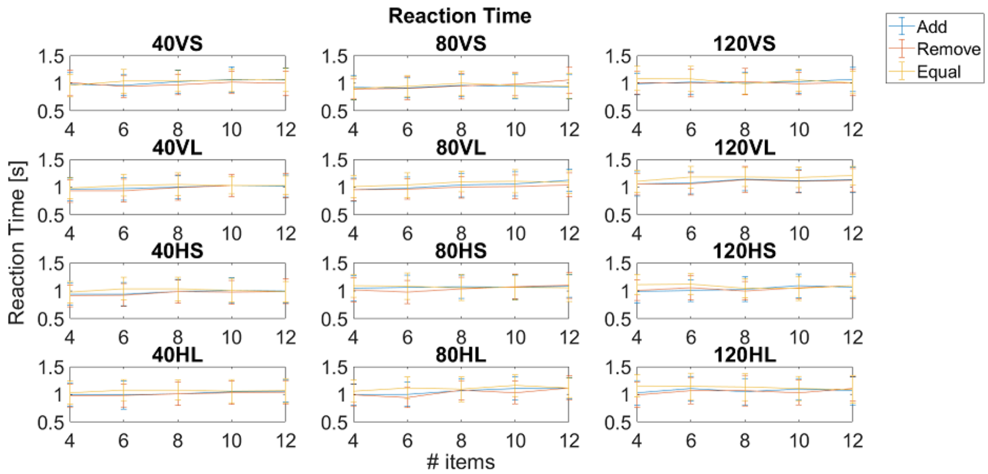
(b)

Figure 15: *Change detection* task: mean and standard deviation of the hit rate in the different blocks at the variation of the horizontal angular distance from the center (top) and of the distance ray (bottom). In the first case, together with the angular variation, a normalized histogram of yaw head rotations is shown. H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

participants performance increases again (Figure 14 b and 15 b). Apparently, considering the head rotations histograms, subjects were more prone to look at different part of the scene in the vertical layout scenario than in the horizontal layout one, when they had a 0.3 s of observation time. Reasons are not clear as in any case participants were not prevented from observing a different subset of the scene. This could explains the slightly better results obtained in the vertical layout case. Results only exclude an effect of the absolute position of objects inside the FOV, while the effect of their relative position with respect to the other elements in the scene is a still a hot topic and could lead to further and in-depth researches.



(a)



(b)

Figure 16: *Change detection* task: mean and standard deviation of the reaction times in the different blocks considering appearance and disappearance together (a) or separately (b). V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

As participants had only 1.5 s to answer, reaction time differs only in a small range around 1 s. Results in the appearance and disappearance case are comparable, and differ from those in the Equal condition (Figure 16 b). A change is usually recognized slightly faster than the absence of change, maybe because when people perceive that an object has been added or removed they immediately answer, whereas in the no change condition they keep looking for the change for a while before answering. Statistically significant differences have been found between trials with and without change only in the trials with FOV 120°, both with the horizontal and the vertical layout, and in the

trials with FOV 40° with horizontal layout and long observation time (Figure 16 a).
Moreover, reaction times depends on the number of items, in fact responses when 4 or 6 items are presented are faster then those with 8, 10 and 12 elements ($p<0.02$). The only exception is in the FOV 120° case with vertical layout and short observation time, where participants give answers based on what they can actually see in their Field of View, inferring on what they can not see.

Trials with FOV 40° and 80° differ only when observation time is short ($p<0.02$); trials with FOV 80° and 120° differ in the vertical layout case ($p<0.02$), while those with 40° and 120° are different in all cases except with vertical layout and short observation time. In trials with FOV 120° and longer observation time, in fact, before answering volunteers first fast explore the entire scene, thus they take a little more time to answer.

If we consider the influence of observation time, statistically significant differences ($p<0.02$) have been found in the FOV 40° wih horizontal layout and in the FOV 80° and 120° with vertical layout case. Finally, the layout seems having a mild-to-null influence over results, in fact, we found statistically significantly different results only in the FOV 80° with short observation time case and in the FOV 120° with long observation time case.

3.1.3.2 *Change localization test*

Again, we first organised data collected in 12 groups based on the layout, the FOV and the observation time. Then, we calculated for each group the mean hit rate and the false alarm rate per number of items across participants. We excluded from the first part of the analysis the pointing accuracy, which was later evaluated through the PCA.

Mean hit rate tends to decrease as the number of items increases, while false alarm rate increases, unlike previously seen (see Figure 17 and 19). Differences from previous experiment could be due to the fact that volunteers had more time to answer, so errors like involuntarily pressing the wrong button or not answering in time occurred less frequently. Through repeated two-sided Wilcoxon test statistical analysis these trends have been found to be statistically significant: in particular, hit rates with 4, 6 and 8 items are statistically different ($p<0.02$) from those referred to 10 and 12 items; whereas, only false alarm rates with 4 and 6 items are statistically different ($p<0.02$) from those referred to 10 and 12 items. These results would confirm the hypothesis that when number of elements to be remembered do not exceed WM capacity, performance accuracy is high, while, once Working Memory capacity is overcome, participants start guessing the answer. The only exceptions are trials with FOV 120° and short observation time, with both layouts, where false alarm rates arrive to 0.5 and hit rate fast decrease below 0.75. This behaviour is due to the fact that people could see just two third of the scene, so they tended to guess answers.

Moreover, it is possible to notice that better results are associated to the higher observation time, especially in the case of FOV 80° and FOV 120°, where false alarm rates differences have been found to be statistically significant ($p<0.02$). Having more time to observe the scene, in fact, subjects can rotate their head and integrate multiple views or, when no rotation is required, they can find efficient strategies to memorize more elements or their spatial distribution.

Nonetheless, while results in the FOV 40° and 80° trials are comparable, differences between false alarm rates referred to FOV 40° and 120° and to FOV 80° and 120° are, in general, statistically significant ($p<0.02$), except in the vertical layout long observation time case. In other words, giving people enough time to have a complete overview over the scene, performance with the largest FOV is comparable to those referred to trials where stimuli are all visible inside the visual field. Layouts, instead, seem having no influence on hit or false alarm rate.

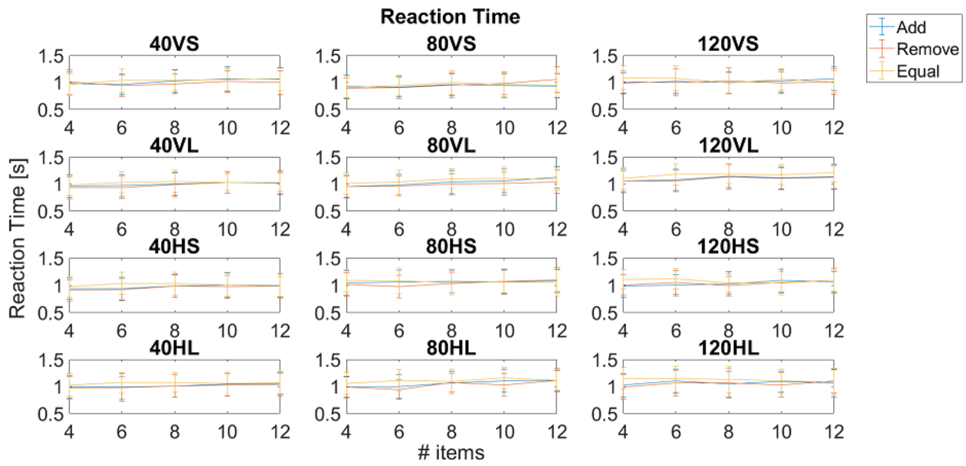
Results obtained considering the addition or removal of a sphere separately, confirm these findings (see Figure 19). Contrary to expectation, they are in general slightly better in the disappearance case. This difference is statistically significant for FOV 40° ($p<0.02$) and 80° ($p<0.05$) with horizontal layout and short observation time, for FOV 80° vertical layout and short observation time ($p<0.05$), FOV 80° horizontal layout and long observation time ($p<0.02$) and FOV 120° and long observation time ($p<0.02$).

Also in the *change localization* test we decided to define a threshold of 0.75 for hit rate and we excluded from each block trials referred to the number of item, for which the hit rate falls below the established threshold subjects having hit rates higher than 0.75. On the remained data we calculated the per bin hit rate, in order to understand if the probability to correctly detect changes was associate to the absolute position of the changing item, shown in Figure 18 and 19. Results confirm what we previously highlighted for the *change detection* test. As expected, performance is better and with less variability when a longer observation time is provided and decrease with FOV enlargement.

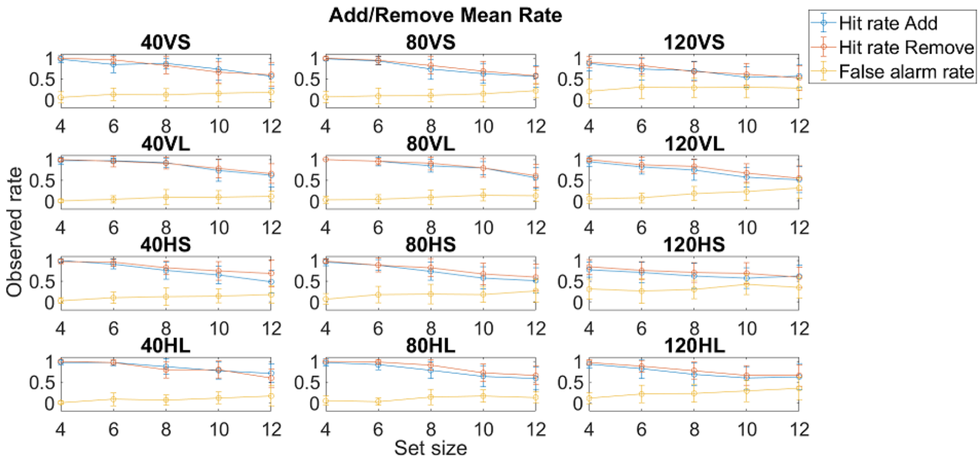
Considering both the horizontal angular distance from the FOV's centre and the height or ray distance, respectively in the vertical and horizontal layout, no particular trends related to spatial position has been found, except for the FOV 120° case with short observation time. In this trial, participants could not actually have a complete overview of the scene, before test, thus hit rate decreases in the periphery of the Field of View, reaching a value of 0.5, which indicates casualty. Whereas, with larger observation time, performance becomes stable in the entire FOV.

Even if people had 5 s to give an answer, average of reaction times are in general smaller than 1.5 s (see Figure 20). They are similar in case of appearance and disappearance of an object; while, when no change is applied, participants tended to answer in a slightly faster way (see Figure 20 a). Maybe because in the first case a precise pointing action is required, while in the second case volunteers just press the "Equal" button. This consideration is valid only in the trials with vertical layout, where differences has been found to be statistically significant ($p<0.02$).

Moreover, reaction times do not seem to depend on the number of items except for trials with change with FOV 40° and vertical layout and FOV 80° with



(a)



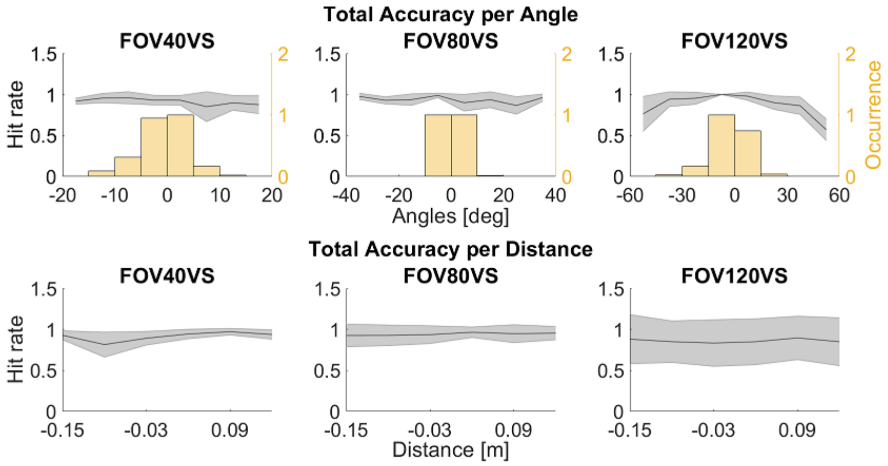
(b)

Figure 17: *Change localization* task: mean and standard deviation of hit rate and false alarm rate in the different blocks considering appearance and disappearance together (a) or separately (b). V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

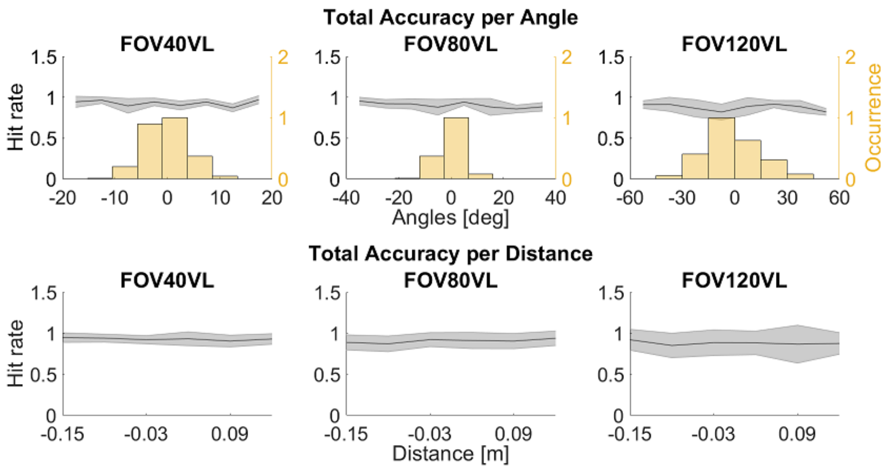
short observation time, where statistically significant differences have been found between 4, 6 and 8 items and 10 and 12 items ($p<0.02$).

Comparing the variation of reaction time among FOVs, FOV 40° or 80° show comparable performance, while they are statistically significantly different from trials with FOV 120° ($p<0.02$). This may be due to the fact that FOV 120° trials are conceptually different. When observation time is 0.3 s and people do not have time to see the entire scene, they guess the answer, whereas when observation time is 0.9 s they have to rotate their head and explore the scene during the test phase, thus they spend more time on search.

Layouts seem influencing reaction times only in trials with 120° and short observation time and with FOV 40° with both short and long observation time. The effect of observation time over reaction time, instead, is more evident in trials with FOV 40° and 120°, in both the horizontal and vertical layout case:



(a)

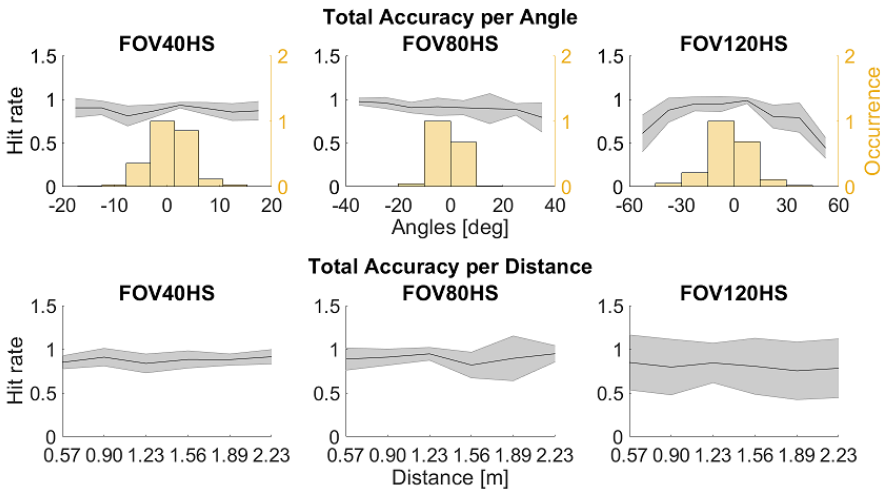


(b)

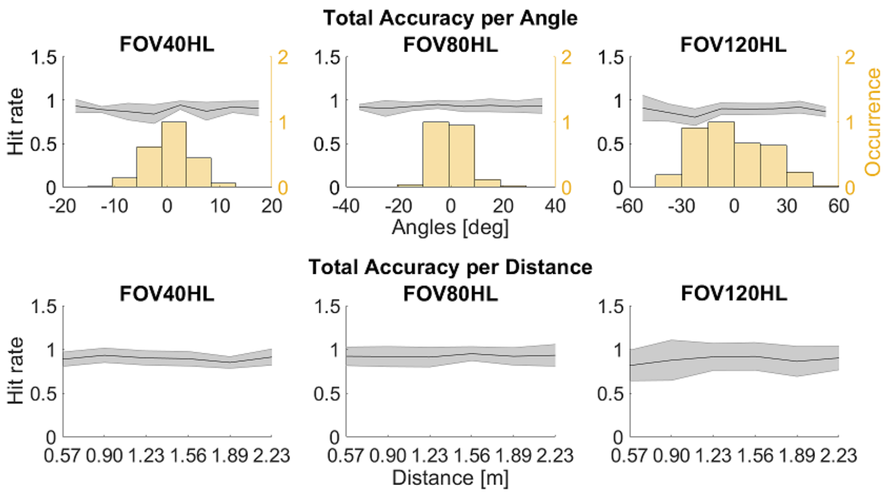
Figure 18: *Change localization* task: mean and standard deviation of the hit rate in the different blocks at the variation of the horizontal angular distance from the center (top) and of the height (bottom). In the first case, together with the angular variation, a normalized histogram of yaw head rotations is shown. V = vertical; S = 0.3 s observation time; L = 0.9 s observation time.

with longer observation time, also reaction time is slower. In the FOV 40° case, this can be explained by the fact that participants memorized more details, thus more time is required to compare the old and new information before answering. In the FOV 120 ° trial, instead, research of the error is longer as the space where items were distributed is larger, so people need time to explore the test scene.

Unlike *change detection*, in the *change localization* test, subjects had to indicate where they thought an object had been added or removed by using a raycast. It is, thus, interesting investigating the pointing precision at the variation of layout, FOV and observation time. For this reason, we calculated the PCA of



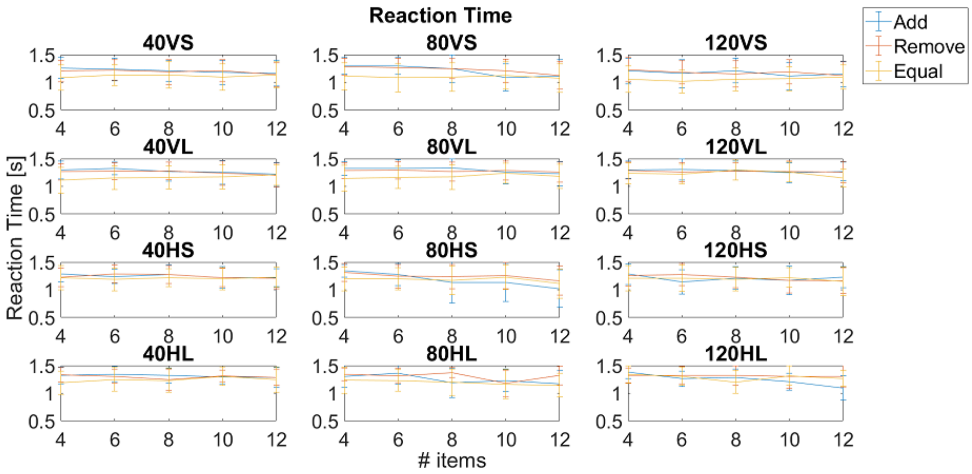
(a)



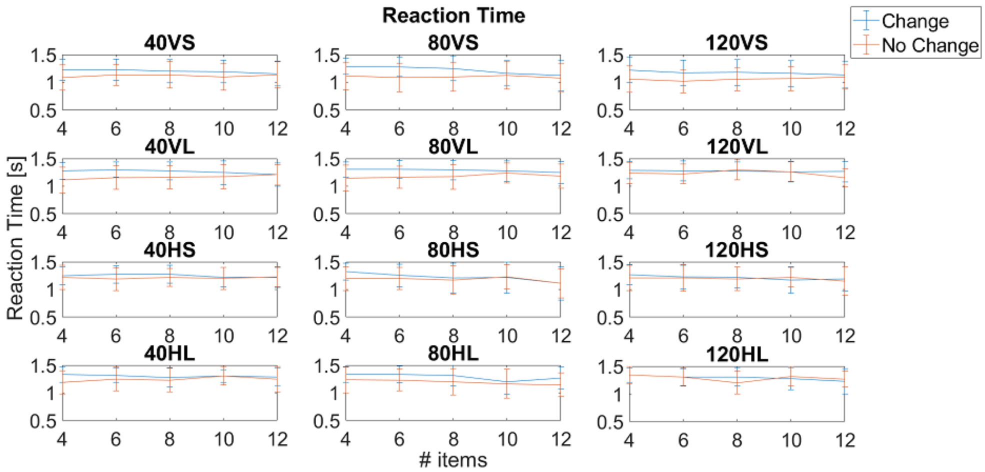
(b)

Figure 19: *Change localization* task: mean and standard deviation of the hit rate in the different blocks at the variation of the horizontal angular distance from the center (top) and at the variation of the distance ray. In the first case, together with the angular variation, a normalized histogram of yaw head rotations is shown. H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

the raycasted positions referred to each target point. We considered only trials referred to number of items for which hit rate is above the threshold, 0.75. This means that we analyse pointing precision only in those trials, in which people have correctly perceived the change with a probability of 75%, which is well above the random response threshold, 50%. Again, as each target position had the same probability to be selected to spawn the modified object, enlarging the FOV, less data per target position are available, making results unstable. Thus, in order to ensure the same data "density", depending on the trial, we divide the horizontal FOV in sectors of different size, 5° for FOV 40°, 10° for FOV 80°,



(a)



(b)

Figure 20: *Change localization* task: mean and standard deviation of the reaction times in the different blocks disappearance together (a) or separately (b). V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time.

15°for FOV 120°. So, in Figure 21 and 22, the points indicate the actual target positions in the FOV 40°, and the centre between two or three adjacent target positions respectively in the FOV 80° and 120° case.

In general, raycasted positions respect the shape of the $6 \times n$ original matrix, indicating that participants are more or less aware of the total space where stimuli are distributed even if they have never seen the matrix. With FOV enlargement pointing error increases and the greatest inaccuracy is associated to FOV 120°. However, participants easily recognize the zone of their Field of View in which the change was applied, in fact the mistake in generally referred to adjacent positions. The observation time, instead, does not seem playing a fundamental role in raycast precision.

Considering the layout, due to the nature of the raycast action used and to the spatial distribution of element, pointing errors have different characteristics. In the vertical case, when people indicate the correct position, error seems to be symmetric with respect to the real position but mostly distributed along the vertical axis, thus we have ellipses centred in the real target position (see Figure 21). A possible explanation for this result could be that participants tended to raycast the target position from below and not frontally, thus a higher error on the vertical axis is plausible. In the horizontal case, instead we have an overestimation of depth, as shown in Figure 22. Even if ellipses tend to get bigger as the distance increases, the actual angular error remains almost constant. Small oscillations of the wrist around the lateral axis, in the order of 1-2 degrees, generate bigger error for distant spheres than for nearer ones.

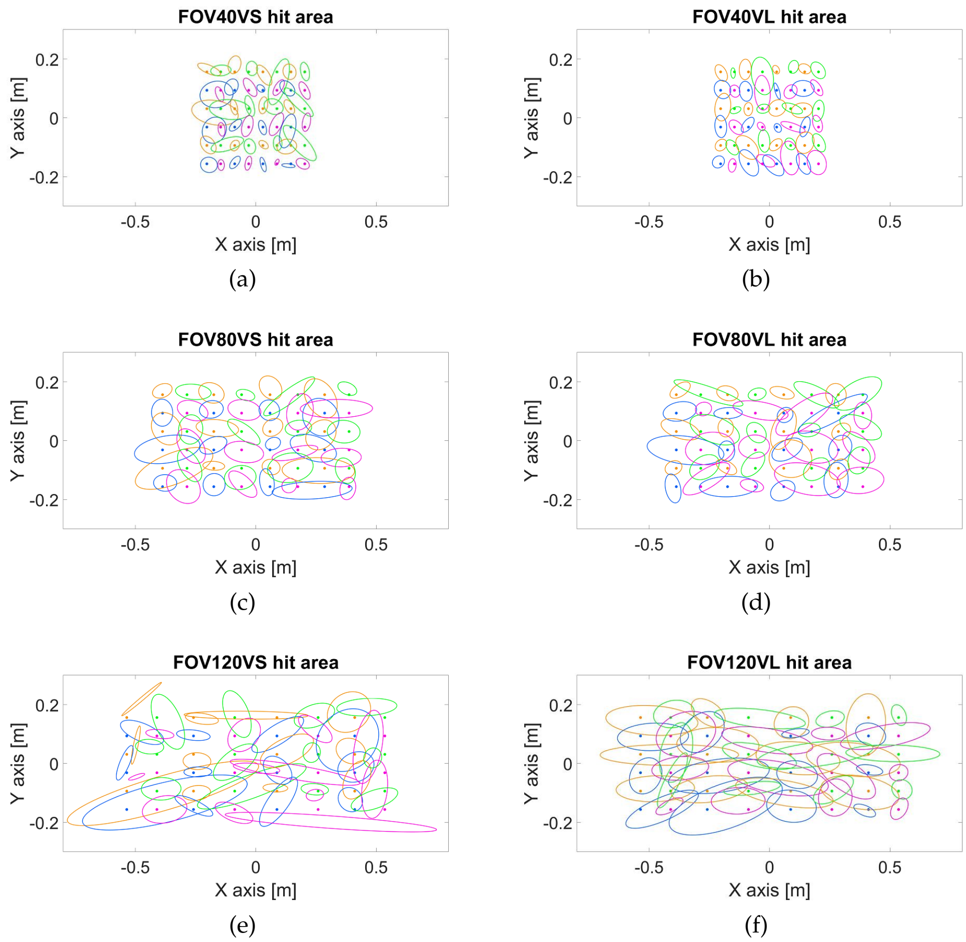


Figure 21: *Change localization* task: PCA showing the distribution of pointed positions around the target for the vertical layout trials with FOV 40° (top), 80° (middle) and 120° (bottom). V = vertical; S = 0.3 s observation time; L = 0.9 s observation time

Same results can be obtained analysing separately PCA referred to appearance and disappearance (Figure 23 and 24). Nonetheless, when participant have to raycast an added object, in general precision is higher, except when

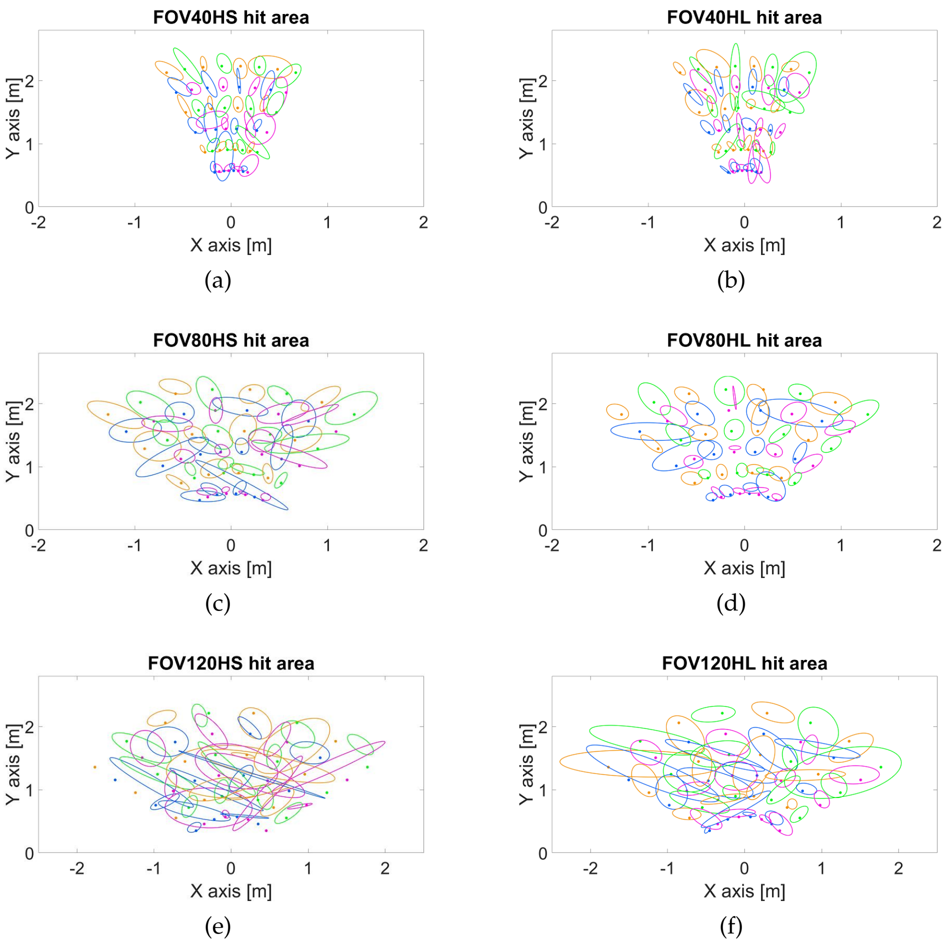


Figure 22: *Change localization* task: PCA showing the distribution of pointed positions around the target for the horizontal layout trials with FOV 40° (top), 80° (middle) and 120° (bottom). H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time

they have not identified correctly the sphere they have to indicate. Whereas, when they have to point the location where an item has disappeared, they are more inaccurate as they have to completely rely on their memory or on allocentric cues, i.e. the identification of the position of an object based on the position of other elements in the scene.

3.1.4 *Outcome*

The goal of this research is the evaluation of how people perceive and process information in VR in order to define some guidelines to improve interaction and gathering of information in simulated virtual environment, by better conveying input to the user. In particular, we decided to focus on the evaluation of the bandwidth of human perception and VWM capacity. Literature in this field is rich, but the majority of studies use 2D images and videos as input, so we first had to define a systematic approach based on the design and im-

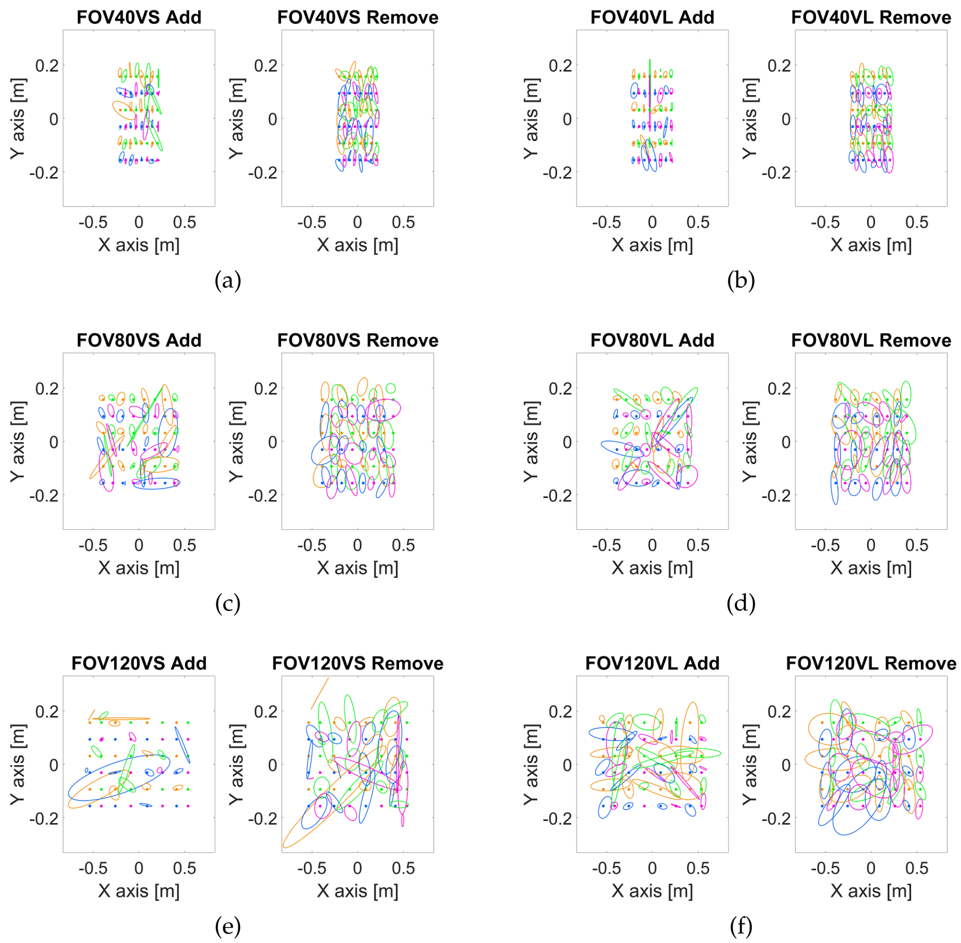


Figure 23: *Change localization* task: PCA showing the distribution of pointed positions around the target for the vertical layout trials with FOV 40° (top), 80° (middle) and 120° (bottom), considering appearance and disappearance separately. V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time

plementation of an ad hoc software framework and a rigorous methodology. We thus adapted the *whole-display* change detection *one-shot* paradigm to be used in an immersive Virtual Reality and devised two experiments: a *change detection* test, where participants have to recognize if a change has occurred and a *change localization* test, where we ask them to indicate if a change has occurred and where it was located. We moreover analysed the influence of five different conditions on change detection: the kind of change (appearance, disappearance and no change), the set size in the original scene (4, 6, 8, 10, 12), the objects distribution, both in terms of spatial layout (vertical with linear distribution or horizontal with perspective cues) and of occupied horizontal Field of View (40, 80 and 120 degrees), and the observation time (0.3 and 0.9 s). Data acquired on FOV 40° allow us to have a baseline for the analysis of results obtain with larger FOVs and fill the gap between non immersive 2D and immersive 3D experiments.

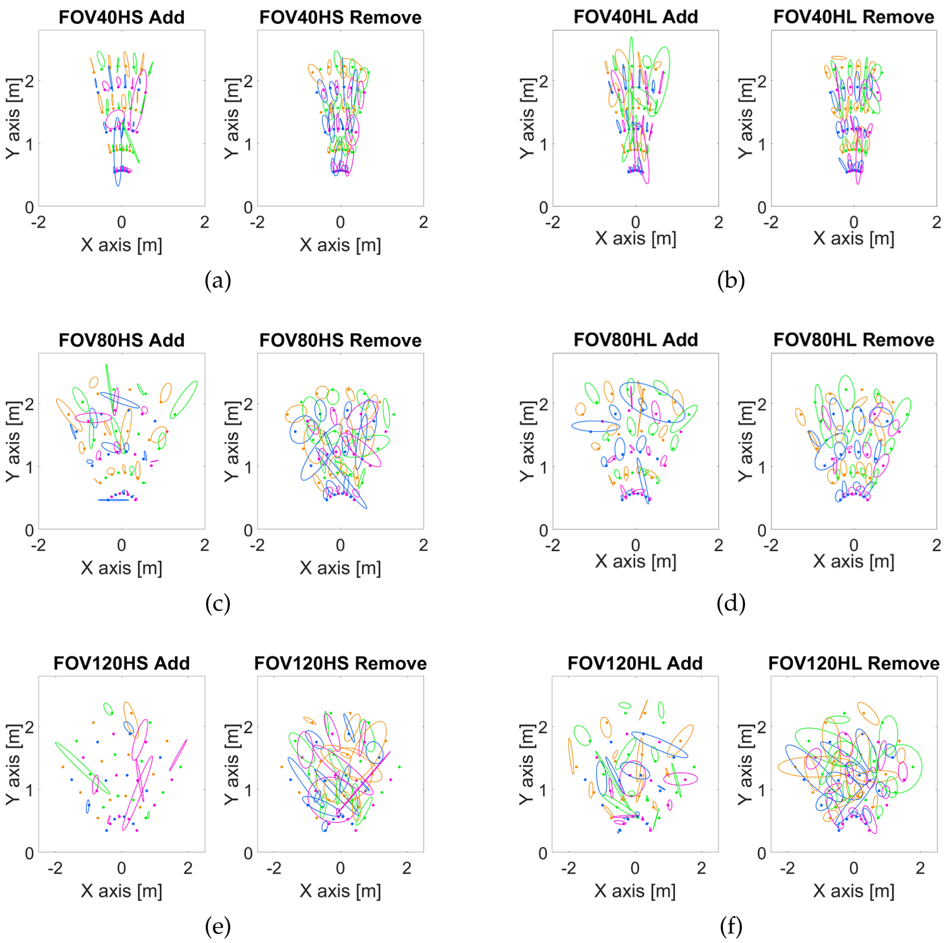


Figure 24: *Change localization* task: PCA showing the distribution of pointed positions around the target for the horizontal layout trials with FOV 40° (top), 80° (middle) and 120° (bottom), considering appearance and disappearance separately. V = vertical; H = horizontal; S = 0.3 s observation time; L = 0.9 s observation time

Concerning FOV 40° and 80°, where stimuli are inside the participant's visual field, results confirm findings of the previous literature. These results are obtained also in the FOV 120° only when subjects are given enough time, 0.9 s, to turn their head and build a gist of the entire scene. In both tests, people are able to recognize a change with a 0.75 accuracy when up to 8 elements are present in the scene, except for the trial with FOV 120° and short observation time, where the upper limit is 6 items. So we can state that cognitive mechanisms for information perception and elaboration used in real-life settings and Virtual Reality scenarios are the same or at least comparable. Layout seems not influencing performance, even if people often complained about the vertical wall, stating they found it uncomfortable and more fatiguing and made them feel trapped. So, the 70 cm ray of the semi-cylindric wall contributed to this feeling. Depending on the task, information can be thus disposed both vertically or horizontally, but, in the first case, more attention should be paid to the distance at which information have to be placed in order not to annoy the user. Finally, also the kind of change did not affect results, maybe because of an absence of pictorial cues, such as occlusions or shades.

Future developments of this research include considering a further increment of the FOV and of the observation time, in order to be able to exploit the entire 360 degree virtual world surrounding the user, or analysing the influence of the relative position of items in the scene on change detection. On the one hand, these findings and future ones will certainly enrich literature in the perception, psychological and medical field, studying the limits of human Visual Working Memory; on the other hand, they will be useful in the design and development of application for simulation and training in VR, but also in other areas, for example data visualization.

3.2 Natural Human Computer Interaction in VR

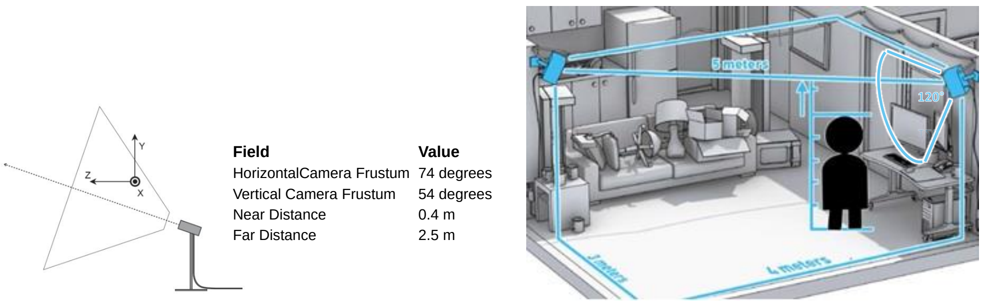
The goal of immersive Virtual Reality simulation systems for training and simulation is to improve trainees' knowledges and skills through active learning. Thus, interaction is required to be a sort of transparent layer allowing the user to interact with the virtual environment in an intuitive and efficient way, without distracting him from task execution or forcing him to redirect any amount of cognitive effort from the task to the interaction. Although the multiple solutions available, some open problems still persist.

Our research particularly focuses on two state of the art interaction modalities, touchful (i.e. HTC Vive controllers) and touchless, particularly vision-based (i.e Leap Motion). Controllers ensure stability, reliability and effectiveness, thanks to their inertial sensors and external tracking, and are the more commercially diffused tools for gaming, so they are familiar. However, interaction is achieved by pressing buttons and triggers or sliding fingers on a touchpad, which decrease the degree of the interaction fidelity. Moreover, the fact of holding an external device while being immerse in the VE could potentially reduce *sense of presence*. Non-wearable solutions, instead, allow interacting with virtual objects using natural and intuitive hand poses and gestures (grasping, grabbing, pinching, opening the hand for release...), thus improving *sense of presence*, agency and ownership, and, at the same time, handling real tools, which could be used to accomplish specific actions in industrial or medical contexts, going towards Mixed Reality solutions. However, hands detection and recognition problems are still common issues in vision-based approaches. Nonetheless, hand poses and gestures are not as natural and intuitive as they appear: in order to overcome tracking issues, in fact, people have to keep their hand inside the limited Leap Motion Field of View and avoid poses which could cause joints occlusion.

In other words, both devices allow having a 3D interaction exploiting 6DOFs. Considering them in a naturalness scale, Leap Motion is more natural than HTC Vive controllers, because of the bare hand interaction and the higher interaction fidelity degree it provides. On the other hand, also controllers offer some advantages: they are familiar, as joysticks and controllers in general are a common tool for gaming, and intuitive to use. The higher naturalness degree of an interface, however, does not imply it is the best solution, in fact, stability problems or a lower control over events, as in the Leap Motion case, could impair performance and frustrate users.

The aim of our research consists in the evaluation of the best interaction method able to guarantee a natural, stable and efficient interaction with the VE. For this reason, controllers and Leap Motion-based interfaces were used and compared in terms of performance and preferences. Considering the advantages of bare-hands interfaces, we moreover tried to improve Leap Motion interaction, by solving some of the well-known open issues.

Firstly, an ad hoc algorithm for the Leap Motion was implemented and tested in a simple selection and manipulation task, a Shape Sorter game. This activity requires a limited interaction space, so the Oculus Rift DK2, suitable for sitting and standing application (Figure 25 a), was chosen for the VR setup. Two versions of the application were, however, implemented, considering as the game area the peripersonal and near action space. A larger interaction space implies wider movements, thus more tracking issues. In particular, the idea was understanding if problems could be due to the Oculus Rift DK2 tracking system, the Leap Motion or a combination of both and defining the boundaries for a stable interaction. Results suggested that, in order to ensure a stable interaction in the near-action space, a new tracking system is required. For this reason, since then, experiments have been conducted with the HTC Vive, allowing for standing applications and room-scale setups (Figure 25 b). Furthermore, from the comparison between controllers and Leap Motion, it emerged that, even if people appreciated the idea of interacting with the VE using their own hands, they expressed doubts on Leap Motion usability and preferred the touchful device.



(a) (b)

Figure 25: (a) Oculus Rift DK2 and (b) HTC Vive tracking system limits.

As these outcomes seemed controversial, we wondered if providing some kind of feedback during interaction would further help improving performance with the Leap Motion, overcoming the lack of haptic feedback, which is an essential component of the action/perception in our daily experience. For this reason two different feedback modalities, visual and audio, have been implemented, as a substitute of the haptic one, which would require specialized hardware. New improved interaction algorithm was tested in two selection and manipulation tasks, the Shape Sorter and the assembling of the Ironman suit. The two tests differ for the level of complexity: in the first case (the simple task) the space for interaction is a desk and lower precision is required in reaching the target position, i.e. it is sufficient carrying the shape nearby the correct hole; in the second case, (the complex task), instead, interaction space is an area of around 3.5 m by 3.5 m and a finer manipulation is requires, as objects have to arrive at the exact target point with a specific rotation. In both cases, the complexity of the task depends more on the interaction, the fineness grade of the manipulation or the amplitude of movements to be performed, than on the cognitive loads required to accomplish the task itself, so that performance should be mainly affected by the chosen interaction device. Again the tasks were accomplished with the Leap Motion and the controllers and performance was compared.

In the end, we have obtained a software framework, constituted by the implemented interaction algorithm and the designed applications, which can be used with different HMDs (Oculus and HTC Vive) and devices for interaction (controllers and Leap Motion).

In this chapter, the designed scenarios, the implemented interaction algorithm and the objective and subjective parameters used for the evaluation of performance and preferences will be described in Section 3.2.1. Then, the devised experiment and the obtained results (Sections 3.2.2 and 3.2.3) will be presented. Finally, outcomes will be discussed considering the main goal of our research (Section 3.2.4): the implementation of a Leap Motion-based interface, able to overcome the stability and efficiency issues and to compete in terms of users' performance and preferences with controllers and in the evaluation of the best modality for the interaction in VR. Experiment are divided as follows:

- Experiment 1: evaluation of the usability of our implemented algorithm for the interaction in the peripersonal and near-action space in the accomplishment of the Shape Sorter task, using a setup composed by Oculus Rift and Leap Motion [11];
- Experiment 2: comparison between HTC Vive controllers and Leap Motion-based interface in the accomplishment of the Shape Sorter task [68];
- Experiment 3: evaluation of the improvement of performance in the Shape Sorter task accomplishment when adding visual and audio feedback to the Leap Motion based interface [10];
- Experiment 4: analysis of the influence of different feedback modalities on the Leap Motion-based interface in the accomplishment of the Ironman task [10];
- Experiment 5: evaluation of performance in the Ironman task with the controllers, through the analysis of data acquired during an in-the-wild experimental session [10];

3.2.1 *Material and Methods*

Both applications, the Shape Sorter and the Ironman, have been designed to serve, in future, as collaborative virtual environment (CVE) and require the presence of two users: a main user, wearing the HMD and physically interacting with the VR scenario, and a supervisor, who uses a desktop application to have a complete overview over the scene and can eventually intervene to help the player. The supervisor, in this case, is the experimenter, conceived as a sort of "deus ex machina". Of course, his interventions must be limited to situations of extreme necessity, which the user in VR can not manage on his own. Moreover, he must pay close attention to the participant's health state and be able to stop the application in case of severe simulator sickness. The two views are part of the same application running on a computer which is connected both to the HMD worn by the main user and to the desktop screen, mouse and keyboard used by the supervisor. The computer has the minimum requirement to support VR, it is an Alienware Aurora R5 with a 4GHz Intel Core i7-6700K processor, 16GB DDR4 RAM (2,133MHz) and a Nvidia GeForce GTX 1080 graphic card.

3.2.1.1 *Scenario 1: The Shape Sorter*

The Shape Sorter game, used in Experiment 1, 2 and 3, is set in a school class. It has been developed using Unity 3D and 3D models created in Blender [2] or downloaded from SketchUp [3].

As shown in Figure 26 a, in front of the player there is a desk with 12 objects, different in shape and color, and a black structure with 12 holes corresponding to the items. All around the room, there are other non interactable decorative elements and pieces of furniture. The task consists in grabbing the items, one per time and using one hand, and insert them in the correct hole. The game ends when all the shapes are correctly positioned.

Before the main scene, there is an optional demo scene, where participants can explore and act in the VE, to get familiar with the scenario and the selected interaction modality and explore the boundaries of the tracked game area.

The desktop application (Figure 26 b), showing the virtual room from a fixed point of view, allows the supervisor to see what the player is doing and help him in three ways: deleting some objects, in order to make the scene clearer; repositioning objects accidentally thrown out of the game area; selecting a hole, i.e. illuminating it with a white light. The operator will also be able to reset the scene, start and pause the timer for time acquisitions or stop the application at any time.

In Experiment 1, the virtual scene has been rescaled in order to have a correspondence with the real environment where experimental sessions took place. In particular, the virtual desk in front of the player and the bookcase behind him correspond in position to the real ones (Figure 27). So, for example, by touching the surface of the virtual desk, the user receives an haptic feedback from the real desk. The position of these elements is defined in the reference system of the Oculus tracking camera, which is considered the origin of our VE. The camera, in fact, is the only fixed point in the setup, existing in both realities, the real and the virtual one. Also the position and rotation of the headset in the 3D space is referred to the system of reference of the camera. In the peripersonal space case, the game area corresponds almost to the real desk, 110 cm long and 60 cm wide, and the volunteer sits on the chair; while in the near-action space case, the game area occupies the real desk and part of the space around it, for a total surface 156 cm long and 140 cm wide, and the participant has to stand up and play on foot.

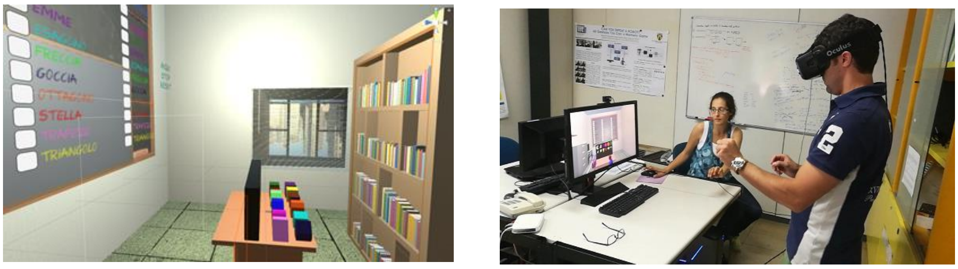
As stated before, Oculus tracked area is limited and when the headset is out of the boundaries of the camera FOV, lags occur in the rendering. This can disorient the user and cause *cybersickness*, for this reason, we decided to display red lines, representing the camera frustum: everytime the user crosses

2 https://www.blender.org/
3 https://www.sketchup.com/it

the boundaries of the tracked area, he can understand the reason of rendering problems and re-enter to the safe zone.

(a) (b)

Figure 26: Two images of the Shape Sorter game. (a) The main user view in Virtual Reality. (b) The supervisor desktop application.



(a) (b)

Figure 27: Experiment 1: virtual and real view of the Experiment 1 setup with modified Shape Sorter game, for Oculus Rift version. (a) The virtual scenario. (b) A picture of a user playing the game inside the real room.

3.2.1.2 *Scenario 2: The Ironman*

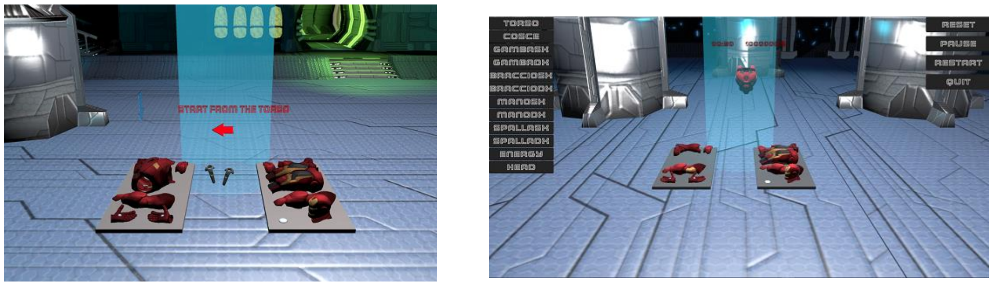
The Ironman game of Experiment 4 and 5 is set in a futuristic sci-fi lab. The game was developed using Unity 3D, the environment was taken from a free asset available on Unity Store, while the Ironman model was downloaded from Sketchfab[4] and modified with Blender to split the suit into pieces.

The VR game is composed by two scenes. In the first scene, the player reads the instructions: he has to assembly the Ironman suit in the shortest possible time, grabbing a piece per time with one hand; suit parts moved but not correctly inserted in the suit, are respawned in their original position after 5 seconds. In the main scene (Figure 28 a), the player finds himself between two tables. On the tables there are the suit pieces and in front of him there is the cylinder, where gravity is set to zero and where he is instructed to assembly the suit starting from the torso. Timer starts when torso is correctly positioned inside the cylinder. The suit is dimensioned proportionally to participant height, computed as the distance of the headset from the floor at the beginning of the game, and the playing area is around 3.5 by 3.5 meters. In

4 **https://sketchfab.com/**.

order to improve engagement and isolate user from the surrounding virtual environment, a soundtrack is played in the background. Furthermor, a metallic click sound is played when a suit piece is correctly assembled.

Again, the desktop application (Figure 28 b) allows the supervisor to see what the player is doing and help him repositioning objects accidentally thrown out of the game area and not automatically re-spawned, restarting, pausing or quitting the game, or resetting the initial position of the torso in the cylinder.



(a) (b)

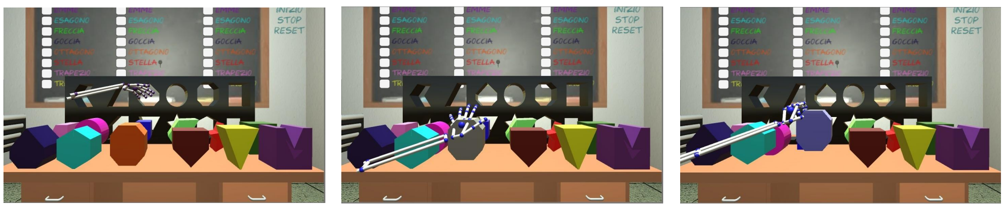
Figure 28: Two images of the Ironman game. (a) The main user view in Virtual Reality. (b) The supervisor desktop application.

3.2.1.3 *Interaction algorithms*

In the two scenarios and with both interaction devices, we have decided to use the real-world metaphor for the selection and manipulation of virtual objects. Thus, all objects inside the playable area can be reached, grabbed and moved around using natural gestures and movements, without the need of implementing ad hoc solutions, like raycast or stretching the virtual limbs. In the controller case, the solution provided in the Steam VR asset for Unity was used: players approach an element, press the trigger button to grab it and release the trigger to leave it. In the Leap Motion case, instead, an ad hoc solution was implemented. When using the Leap Motion to interact with the virtual environment, sometimes users have the impression to lose their control over events, as they perceive objects behaviour as strange and difficult to explain. For example, when they are grabbing an object and their virtual hand glitches or disappears, the item suddenly falls or is thrown away. However, it is difficult for them to predict or avoid actions leading to unexpected behaviours. In other words, tracking problems or joints occlusions cause problems in ensuring a stable and efficient interaction with virtual objects, which cause frustration in users. An important issue is that new users have to learn that, when the hand is out of the Leap Motion camera FOV, it ceases to exist in the virtual world. Even if this concept is easy to learn, it is not intuitive: humans, in fact, are not used to constantly watch their hands during interaction in the real environment.

For this reason, we have decided to impose a more restrictive control over the physics of objects: when an item is not colliding with the hand, its position and rotation is locked. People do not notice the presence of elements unnaturally floating in the air, because they tend to intuitively take and lay objects on the table and, when they turn their head around while grabbing an item, they rarely move the hand outside of their FOV.

Moreover, in order to further improve the interaction with the Leap Motion, a *grab algorithm* was implemented. Everytime a hand collides with an object, the object itself turns grey (Figure 29 b) and the distance between the palm and the baricenter of the item is calculated. Then, the hand's GrabAngle, which is a Leap Motion Unity Asset Core precalculated value, is checked. The GrabAngle is the angle between the fingers and the hand in a grab hand pose. It is computed by considering the angle between the direction of the fingers, excluding the thumb, and the direction of the hand. Its value ranges from 0 radians, for an open hand, to pi radians, when the pose is a tight fist. The GrabAngle is compared to a threshold experimentally defined, by grabbing the different items multiple times and considering the average value. When this parameter is higher than the threshold, the object becomes purple (Figure 29 c) and its new position is runtime calculated so that the distance between hand and object remains constant. Thus, if a user wants to grab an item, he has to approach it keeping his hand open and close the hand after the collision. While, to release the object, he just has to open his hand and move it away from the item. This is the way humans usually interact with real objects. The only difference is the absence of haptic feedback, substituted by the visual one.



(a) (b) (c)

Figure 29: Grab algorithm functioning in the Shape Sorter scenario: (a) no collision between hand and object; (b) a hand-object collision is detected; (c) the user closes his hand in a grasping pose.

In order to understand the importance of feedback in the Leap Motion interaction, we further implemented *four different feedback conditions*, used in Experiment 3 and 4:

- None, where no feedback is provided;
- Visual feedback Visual1, where interactable objects change their appearance twice, when a collision is detected (Figure 29 b and 30 b) and during grasping (Figure 29 c 30 c);
- Visual feedback Visual2, where shapes change appearance once, during grasping;
- Audio feedback, when an object is grabbed a music is played, when the grabbing action finishes music is paused.

The last modality represents the auditory counterpart for Visual2.

Considering Visual1 and Visual2, in the Shape Sorter game we modified the colour of the objects, gray on collision and violet during grabbing; in the Ironman game, for aesthetics reasons, we preferred modifying the "smoothness" parameter of suit pieces, i.e. the level of brightness of the metallic material, which was 0.2 normally, 0.6 on collision and 0.8 during grabbing.

We decided to implement and test both visual and audio feedback because in VR audio is often a minor input: users receive lots of visual information, causing delay and loss of information in their processing, while audio channel is almost unused. Nonetheless, similarly to the real world, haptics is a non-visual sensorial channel.

(a) (b) (c)

Figure 30: Grab algorithm functioning in the Ironman scenario: (a) no collision between hand and object; (b) a hand-object collision is detected; (c) the user closes his hand in a grasping pose.

3.2.1.4 *Parameters*

In all experiments, interaction modalities were evaluated in terms of performance and preferences considering both objective and subjective parameters.

Performance was estimated through the analysis of the total execution and partial times and the number of times the experimenter had to intervene. The total execution time is defined as the time required to accomplish the task from the moment in which the supervisor starts the timer, in the simple task, or from the moment in which the player release the torso in the cylinder, in the compex task. The partial time, finally, is the time required to correctly position each shape or suit piece. Obviously, every time the player changes the grabbed object, the timer is restarted. For every task, we counted the number of times the supervisor repositioned or deleted objects, selected holes or reset the scene. As the experimenter rarely intervened, however, this metrics was used only in Experiment 1 and 2.

Considering preferences, instead, subjects were directly asked to express a preference between controllers and Leap Motion, in Experiment 2, and between the different feedback modalities, in Experiment 3 and 4, or had to fill in the User Experience Questionnaire (UEQ), in Experiment 3 and 4, evaluating, through a series of questions, different aspects of the feedback modalities they had tried. UEQ [92, 145] is a validated questionnaire firstly released in German in 2005 and then translated in different languages, among which Italian and English. It is commonly used to compare usability of different devices and is composed of 26 questions in the form of a semantic differential, i.e. each item is represented by two adjectives with opposite meaning, and adopts a seven-stage scale for rating. The 26 items are clustered in 6 groups:

- *Attractiveness*: describing if users liked the product proposed or not;
- *Perspicuity*: valuating how easy is learning to use the product;
- *Efficiency*: defining on what measure people can accomplish a task without unnecessary effort;
- *Dependability*: assessing the level of control user felt he had over events;
- *Stimulation*: estimating people excitement and motivation;
- *Novelty*: a measure of how innovative and creative the device is.

In Experiment 1, instead, participants were asked to evaluate their experience in the two different interaction spaces and give a preference, through another custom questionnaire, composed by an open question and eight 5-points Likert scale questions, where rate 1 corresponds to "Strongly disagree" and 5 to "Strongly agree":
Q1) Did you prefer the peripersonal space or the near-action space trial? Why?
Q2) Did you feel safe to freely move your hands in the peripersonal space trial?
Q3) Did you feel safe to freely move your hands in the near-action space trial?
Q4) Was the interaction natural in the peripersonal space trial?
Q5) Was the interaction natural in the near-action space trial?
Q6) Level of frustration in the peripersonal space trial
Q7) Level of frustration in the near-action space trial
Q8) In the task, you could use just one hand, did this fact limit you?
Q9) Did the change in color of reached/grabbed objects help you?

In Experiments 1 and 2 we also administered the Simulator Sickness Questionnaire. However, once ascertained that selection and manipulation tasks with real world metaphors and using both headsets, HTC Vive and Oculus Rift DK2, do not cause *cybersickness*, it was no longer used. After the Ironman task (Experiments 4 and 5), however, participants were asked to fill an overview questionnaire, investigating both their well-being state and the test perceived complexity. The questionnaire is composed of 5 questions rated in a 5-points Likert scale, where rate 1 corresponds to "Strongly disagree" and 5 to "Strongly agree":

- How intuitive were the game commands?

- How difficult was understanding to what the single suit pieces corresponded?
- How difficult was positioning the suit pieces?
- How did you feel immersed in the virtual environment?
- Do you feel any simulator sickness symptom, like vertigo, nausea, dizziness, difficulty focusing?

3.2.2 *Experiments*

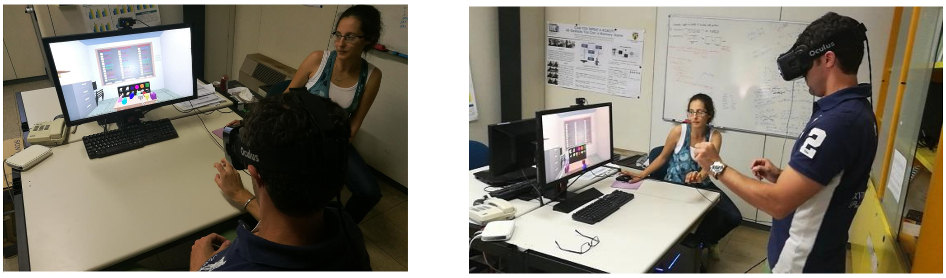
3.2.2.1 *Experiment 1*

The aim of this experiment is the assessment of the stability, efficiency and naturalness of our Leap Motion-based interface during the interaction with virtual objects in the peripersonal and near-action space. We used the Oculus Rift DK2 for visualization and mounted the Leap Motion on the HMD, as shown in Figure 31. The experiment was composed of two trials in the two interaction space setups. A repeated measures counterbalance experimental design was used: all participants tried both setups but half of the them carried out the first task in the peripersonal space setup and half in the near-action space setup. First, subjects were asked to submit the Simulator Sickness Questionnaire (Pre), in order to have their baseline physical status before the exposure to VR. Then, they had to perform the task in one of the two setups. After this, they had to submit a second SSQ (Post), which was used to evaluate users' well-being after the first exposure to VR and before the second one. Next, the second trial was executed in the complementary modality. In the two VR scenes the position of the objects was changed, to prevent people from learning the task. Subsequently, participants had to submit two questionnaires: another SSQ (Post2) and the custom questionnaire.

In total, 21 volunteer healthy subjects, 12 males and 9 females, aged between 24 and 52 (mean 30.8 ± 6.7 years) took part to the experimental session, receiving no rewards. They all had normal or corrected to normal vision, the majority of them had already tried the Leap Motion (71%) and VR devices and applications (90%) in the past, but only 4 of them had already tried the Oculus Rift.

3.2.2.2 *Experiment 2*

In Experiment 2 Oculus Rift DK2 is substituted by the HTC Vive, which provides better tracking system, allowing to design room-scale applications.
This experiment mainly focuses on the evaluation of the interaction in Shape Sorter scenario comparing two different devices, HTC Vive controllers and Leap Motion (Figure 32). In particular two aspects were taken into account: the performance and the quality of the experience. Two separate experimental sessions were conducted: in the preliminary one we used the Leap Motion interaction algorithm provided in the Unity Asset Core; whereas, in the second



(a) (b)

Figure 31: Experiment 1 setup, composed by the Oculus Rift DK2 for visualization and the Leap Motion for interaction. A computer is connected both to the HMD and to the desktop screen used by the experimenter to supervise the main user. (a) Peripersonal space setup. (b) Near space setup.

one, we used the algorithm described in Section 3.2.1.3. In the first case, in fact we compare commercially available solutions for interaction in an immersive VE, whereas in the second case, we want to evaluate if our algorithm can actually improve performance with the Leap Motion.

Similarly to the previous experiment, a repeated measures counterbalance experimental design was used: all participants tried both setups but order of execution varied between subjects. The experimental procedure in both cases included two trials for each participant. Firstly, players were asked to submit a pre-exposure SSQ and performed the task with the selected interaction device. After completing the first trial, they had to submit the post-exposure SSQ. Then, the second trial was executed with the complementary device. Again, to avoid learning the position of all the objects on the desk has been modified in the two trials. Subsequently, users had to submit final-exposure SSQ in order to evaluate their state of malaise at the end of the experiment. Finally, all the volunteers were asked which interaction mode they preferred.

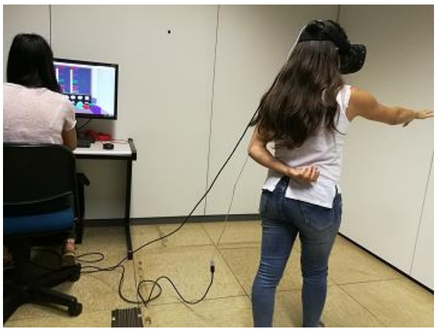
30 volunteer healthy subjects, 15 males and 15 females, aged between 20 and 49 years (mean 27 $\pm$ 6.6 years) took part to the preliminary acquisition phase, while 35 volunteers, including 30 Adults aged between 20 and 45 years (mean 26 $\pm$ 7.3 years) and 5 Children/Teenagers between 7 and 17 years (mean 10.6 $\pm$ 4.6 years) took part to the second session. The aim is understanding if participants of different ages can accomplish the task and use the proposed interfaces. Adults and Children, in fact, can react in different way, have different preferences and needs while interacting with the virtual environment. Only few of the participants had already had past experiences in VR, while the majority of them had never got in touch with it, even if they had heard about it.

3.2.2.3 *Experiment 3*

The goal of this experiment is understanding the influence of different feedback modalities over performance in a Leap Motion interaction. For this reason,

(a)

(b)



(c)

(d)

Figure 32: Experiment 2 setup, composed by the HTC Vive for visualization and the HTC Vive controllers or the Leap Motion for interaction. A computer is connected both to the HMD and to the desktop screen used by the experimenter to supervise the main user. (a) Experimental setup with the HTC Vive controller. (b) Interaction with the HTC Vive controller. (c) Experimental setup with the Leap Motion mounted on the HMD. (d) Interaction with the with the Leap Motion.

the four different feedback modalities described in Section 3.2.1.3 were compared in terms of performance and preferences.

Participants were asked to complete the task with two of the four feedback modalities, according to a counterbalanced repeated measure experimental design. Each combination was assigned to 3 subjects and 18 samples for each feedback modality were acquired. A different initial position of shapes is associated to the 4 feedback modality scenes. During each experimental session, first, users performed the task with one of the two feedback modalities and filled in the first UEQ. After a short break, subjects accomplished the task with the second feedback modality and submitted a second User Experience Questionnaire.

36 volunteer healthy subjects, aged between 15 and 45 years (average 25.4 $\pm$ 5.5 years), took part to the experimental session, receiving no reward. They all had normal or corrected to normal vision and had to sign an informed consent. The majority of them had previous experience with the VR headset but not with the Leap Motion.

3.2.2.4 *Experiment 4*

Experiment 4 has been designed in order to understand if results obtained in the previous experiment with the simple task, can be extended to the complex task of assembling the Ironman suit, where larger movements and a finer manipulation are required. Thus, this experiment replicates the previously described experimental procedure: a counterbalanced repeated measure experimental design, where each participant was asked to complete the task with two of the four feedback modalities. Order of execution varied among subjects. After accomplishing each task, participants were asked to submit the UEQ considering as device to evaluate the combination of Leap Motion and the feedback modality.

36 volunteer healthy subjects aged between 15 and 50 years (average 25.2 $\pm$ 8.1 years) took part to the experimental session, receiving no reward. They all had normal or corrected to normal vision and had to sign an informed consent. The majority of them had previous experience with the VR headset but not with the Leap Motion.

(a) (b)

Figure 33: (a) Experiment 4 setup, composed by the HTC Vive for visualization and the the Leap Motion for interaction. A computer is connected both to the HMD and to the desktop screen used by the experimenter to supervise the main user. (b) Interaction with the Leap Motion.

3.2.2.5 *Experiment 5*

This experiment was devised in-the-wild, during the Smack and Play, a game festival which took place in Genoa. The idea is to understand how people naive to VR and HTC Vive controllers can accomplish the complex task with the touchful interface. Moreover, acquiring data with state of the art technology for interaction allows obtaining a baseline for the analysis of performance in the complex task.

Procedure in this case was simplified, due to the chaotic and crowded environment. First participants were introduced to the hardware used and the game instructions, then they freely played the game.

90 volunteer healthy subjects accomplished the task. They were between 7 and 53 years old and were divided in three groups based on their age: Children

(up to 10 years), Teenagers (from 11 to 21 years) and Adults (from 22 years on). In total 30 Children (average 9 ± 1.4 years), 41 Teenagers (average 14.4 ± 2.8 years) and 19 Adults (average 36.2 ± 12.3 years) took part to the experiment. All of them had normal or corrected to normal vision and signed an informed consent. Permissions were given by parents, when required.

(a) (b)

Figure 34: (a) Experiment 5 setup, composed by the HTC Vive for visualization and the HTC Vive controllers for interaction. A computer is connected both to the HMD and to the desktop screen used by the experimenter to supervise the main user. (b) Interaction with the HTC Vive controllers.

3.2.3 *Data Analysis*

3.2.3.1 *Experiment 1*

Objective parameters (total execution, partial times and number of helping actions performed) recorded during the experimental sessions were evaluated considering different factors as the independent variables of the analysis: (i) the order of execution, (ii) the considered setup (peripersonal or near-action space), (iii) the order of execution and the setup, (iv) the personal preference and the setup. In each analysis the statistical significance of the differences between dataset was calculated by making a paired t-test analysis, after ascertaining that data were normally distributed.
In the first case, the goal is understanding the influence of learning on performance. In order to discern between familiarizing with the interaction modality and learning the task, the disposition of shapes on the table was different in the two trials. Comparing the mean total execution times, a consistent improvement of performance between first (mean 89.7 ± 26.4 s) and second trials (mean 73.3 ± 16.8 s) emerges (Figure 35 a). The difference is statistically significant ($p<0.05$) and proves the existence of a learning curve in the use of the interface. This result is confirmed also by the total execution time found in the analysis which considered both the execution order and interaction space as the independent variables (see Table 2). Also partial times slightly improve, passing from 3.4 ± 1 s to 3.1 ± 0.96 s, but this difference is not statistically significant (Figure 35 b).

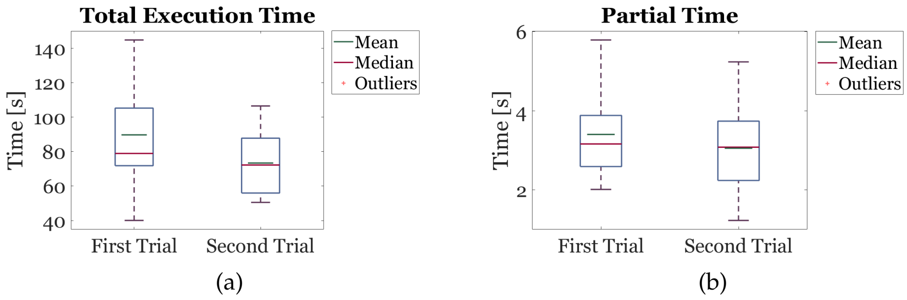


Figure 35: Experiment 1: mean and median of the total execution time (a) and partial time (b), considering the order of execution.

Considering external helps, experimenter had to intervene a few time, usually only to reposition items, thus the number of repositioned elements is mildly affected (Figure 36). This is an indicator of the stability and reliability of the developed interface.

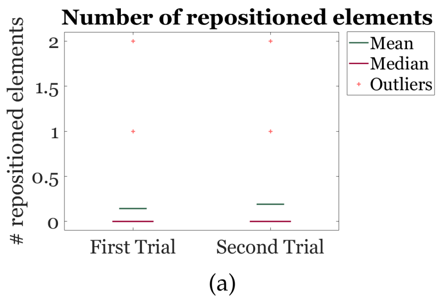


Figure 36: Experiment 1: mean and median of the number of repositioned objects (a) and final score (b), considering the order of execution.

The space of interaction, instead, seems having no influence on performance, in fact times and number of interventions are comparable and no statistically significant differences emerged (Figure 37 and 38).

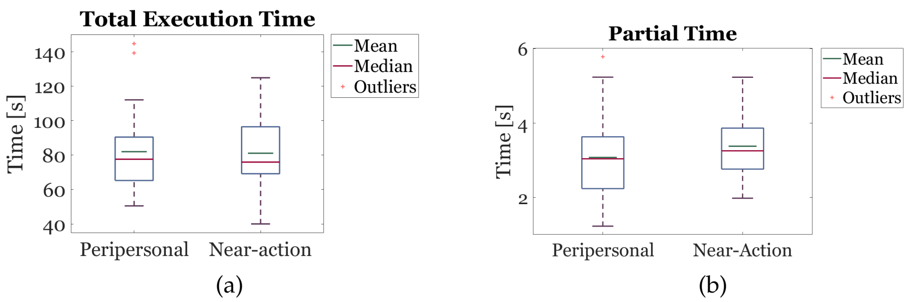
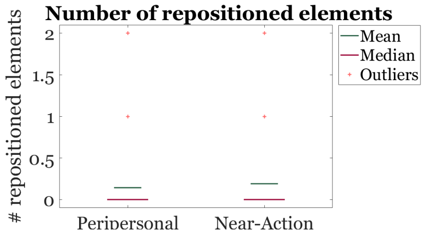


Figure 37: Experiment 1: mean and median of the total execution time (a) and partial time (b), considering the space of interaction.

Finally, there is no correlation between preferences and performance, in fact, observing total execution time in Table 3, they are even worst in the preferred



(a)

Figure 38: Experiment 1: mean and median of the number of repositioned objects (a) and final score (b), considering the order of execution.

Parameters	Trials	Peripersonal space	Near-action space
Total Execution Time [s]	**1st**	94.8 ± 27.5	84.1 ± 25.5
	2nd	67.9 ± 14.6	78.3 ± 17.8
Partial Time [s]	**1st**	3.3 ± 1	3.6 ± 1.5
	2nd	2.9 ± 1.2	3.2 ± 0.7
Repositioned elements	**1st**	0.2 ± 0.6	0.1 ± 0.3
	2nd	0.1 ± 0.3	0.3 ± 0.6

Table 2: Experiment 1: mean and standard deviation of total execution and partial time and number of repositioned elements over all participants considering the order of execution and the space of interaction.

interaction space. For the other objective parameters evaluated no significant difference has been highlighted, maybe because of the high variability of data.

From the analysis of collected Simulator Sickness Questionnaires an increase of the final SSQ grade emerges, however, it is not statistically significant (see Figure 39). Thus, we can state that the VR setup do not cause *cybersickness*. Moreover, three people reported a strange behaviour of the system: in the peripersonal space, when they rotated their head looking at the table, they perceived an oscillatory movement accompanied by a sort of zooming effect. This did not happen when they were staring at their hands while rotating the

Parameters	Preference	Peripersonal space	Near-action space
Total Execution Time [s]	**Peripersonal**	97.6 ± 36.1	93.5 ± 24.6
	Near-action	74 ± 16.1	77.8 ± 20.6
Partial Time [s]	**Peripersonal**	3.1 ± 1.4	3.6 ± 1.8
	Near-action	3.1 ± 0.9	3.4 ± 0.9
Repositioned elements	**Peripersonal**	0.3 ± 0.8	0.5 ± 0.8
	Near-action	0.1 ± 0.3	0.1 ± 0.3

Table 3: Experiment 1: mean and standard deviation of total execution and partial time and number of repositioned elements over all participants considering the space of interaction and the expressed preference.

head. This phenomenon did not interfere with their performance nor caused sickness.

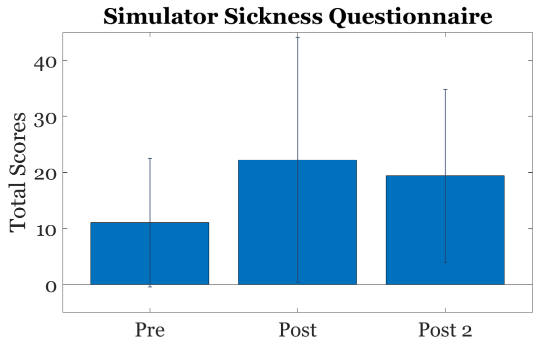


Figure 39: Experiment 1: mean and standard deviation of the Simulator Sickness Questionnaires submitted before the first trial (Pre), between the first and the second trial (Post) and after the second trial (Post 2).

Finally, the subjective questionnaire allows to have an overview over volunteers preferences and opinions (see Figure 40). Considering, the open question about the preferred setup, the majority of them prefer the near-action space, 6 the peripersonal space and 4 have not a preference. The near-action space is in general considered more realistic regarding distances and more comfortable, as it allows more freedom of movements and more interaction with the scene and the virtual objects. Consequently, the task is perceived as funnier and more engaging. Participants who judged the peripersonal space setup as the the best one, instead, say that they felt safer, more at ease and free to move and complain about the wires and the limited game area in the other setup. On the boundaries of the tracked area, in fact, the Oculus Rift camera frequently lose the signal from the HMD, causing lags on the rendering of the scene, which are perceived as annoying and confusing. The same problem occurs when users rotate their head too much or look backward, as there are no infrared sensors on the back part of the headset. Q2, Q3, Q4, Q5 sum up all these considerations: on the one hand, in the peripersonal space setup, due to the characteristics of the setup itself, people feel their movements are limited, but the tracking is stable; on the other hand, in the near-action space setup, they feel free to move in a larger space but are limited by the tracking system and the wires. Focusing on the naturalness of the interaction, the interdiction of using both hands to grab objects has a minimal influence (Q8), also because in general people tend to instinctively use just their dominant hand and virtual objects have no weight, and subjects can easily complete the task (Q6, Q7), so we can say that our grab function is well implemented. Finally, the visual feedback during collisions and grasp actions (Q9) is appreciated, especially by those who had stability problem with the Leap Motion tracking. These issues are mainly related to too fast movements and participants necessity to learn to use the interaction device properly.

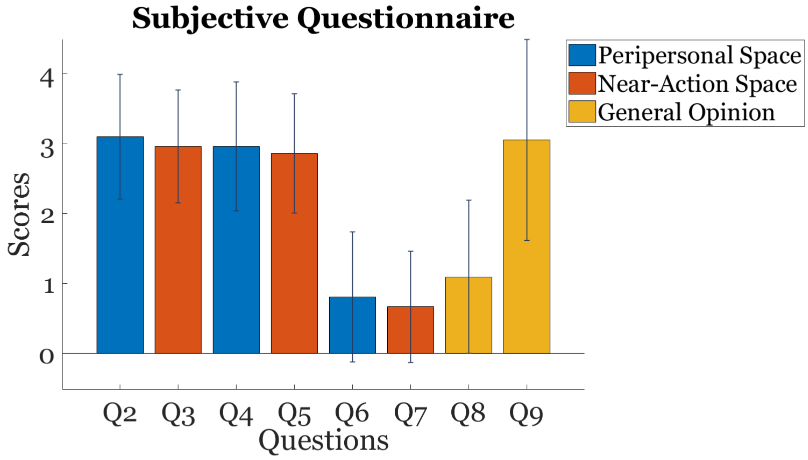


Figure 40: Experiment 1: mean and standard deviation of the answers given to the 8 5-points questions in the subjective opinion questionnaire. P = *peripersonal space* setup, NA = *near-action space* setup, General = questions not referred to a specific setup.

3.2.3.2 *Experiment 2*

The metrics defined to evaluate the task performance, i.e. total execution and partial times, have been evaluated considering different parameters as the independent variable of the analysis: firstly, the device used (HTC Vive controller or Leap Motion); secondly, the order of execution (first and second trial); thirdly, both the kind of device and the execution order. In each analysis the statistical significance of the differences between samples was estimate by making a paired t-test analysis, , after ascertaining that data were normally distributed.

From the analysis of results obtained in the preliminary test, it emerges that performance with the controller is significantly better ($p<0.02$) than the one achieved by using the Leap Motion, in fact all the considered parameters are significantly lower for the HTC controller, because of Leap Motion instability. The total execution time is in average lower with the controller (68.5 s against 161.1 s for the Leap Motion), as shown in Figure 41 a. Similarly the partial time is around 2.3 s for the touchful interface and 5.0 s for the touchless one (Figure 41 b). While partial time is a direct evaluation of the stability and reliability of the interface, total time is influenced both by the interaction device and the cognitive load of the task. For this reason, total time is higher than the partial time multiplied by the number of items (12). In this case, however, being the Shape Sorter a simple and intuitive game, the effect of cognitive effort over total time is assumed to be limited.

The median value related to the number of repositioned elements is equal to 0 in the HTC controller case and to 2 the Leap Motion case, while the average value is 0.3 versus 3.4 (Figure 42 a), indicating a higher instability of interaction using the Leap Motion.

The analysis conducted considering the order of execution shows no statistically significant differences for any of the considered parameters and slightly better results in the second trial only in the time analysis. Because of the coun-

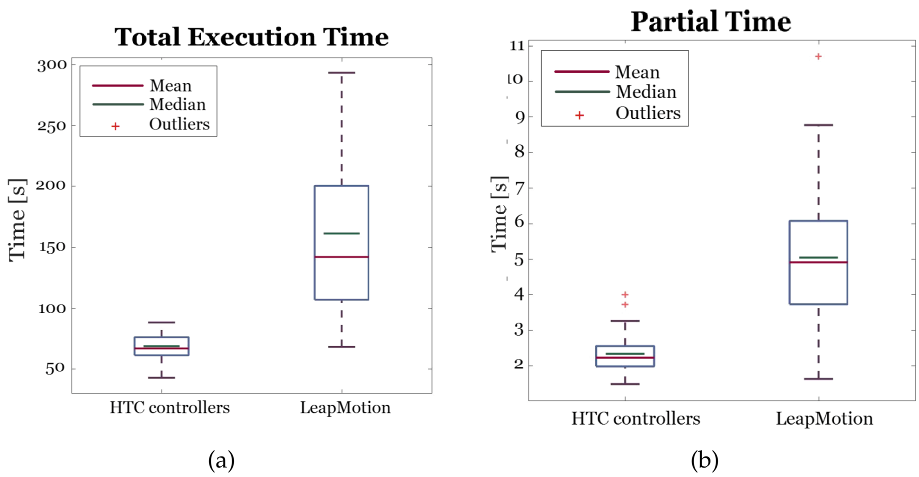


Figure 41: Experiment 2: mean and median of the total execution time (a) and partial time (b), considering the device used.

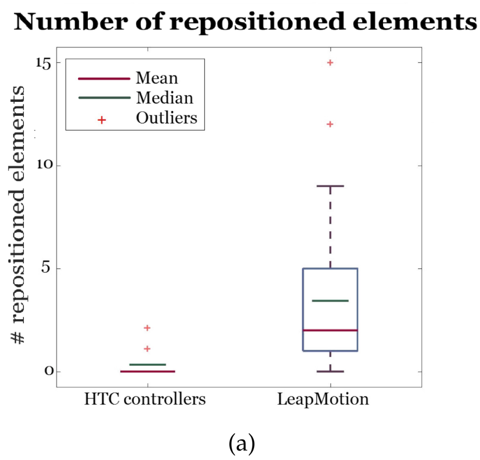


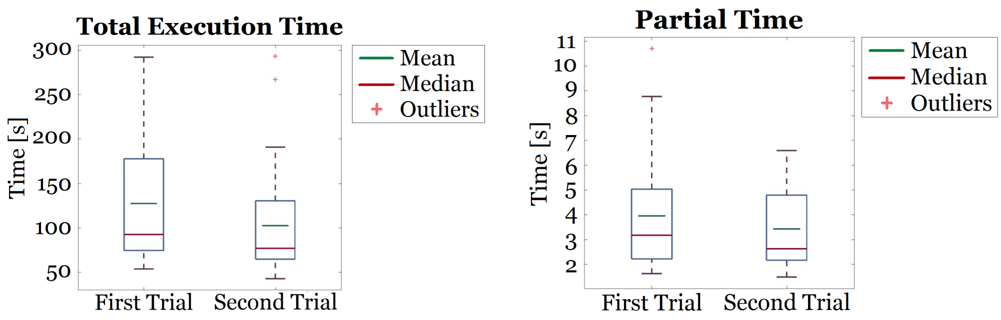
Figure 42: Experiment 2: mean and median of the number of repositioned objects, considering the device used.

terbalanced nature of the experimental design and of the use of two different interaction modalities, those results can be explained as an absence of any learning process, due to changing the interaction device.

The crossed comparison, which considers both the execution order and the device, confirms that performance is better with the controllers and highlights a mild effect of learning (Table 4). Keeping the device fixed and varying the order of execution, all the considered metrics have a little lower value in the second trial, though the differences are not statistically significant.

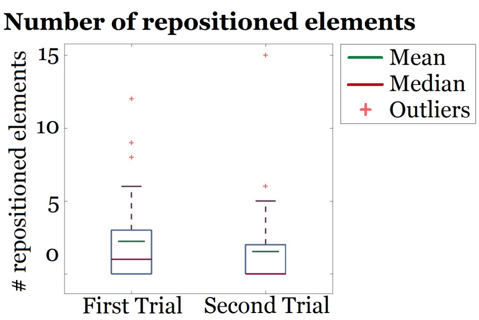
The Simulation Sickness Questionnaire analysis does not show any statistically significant difference between the data distributions relative to the Pre and Post and Post 2 exposures. So it can be stated that there are no adverse effects of our immersive VR setup on users in terms of *cybersickness*.

Finally, 73% of the subjects feel more comfortable in using the HTC controller,



(a) (b)

Figure 43: Experiment 2: mean and median of the total execution time (a) and partial time (b), considering the order of execution.



(a)

Figure 44: Experiment 2: mean and median of the number of repositioned objects, considering the execution order.

while only the 13% of them the Leap Motion for the greater naturalness of interaction that it allows.

Parameters	Trial	HTC controllers	Leap Motion
Total Execution Time [s]	**1st**	73.9 ± 10.4	180.7 ± 67.9
	2nd	63.7 ± 10.9	141.1 ± 65.8
Partial Time [s]	**1st**	1.2 ± 0.7	2.7 ± 3.3
	2nd	1.3 ± 1.2	2.3 ± 5.8
Repositioned elements	**1st**	0.4 ± 0.8	4.1 ± 3.7
	2nd	0.3 ± 0.6	2.8 ± 3.9

Table 4: Experiment 2: mean and standard deviation of the total execution and partial time and of the number of repositioned objects, over all participants considering the device used and the order of execution.

After having substituting the interaction algorithm with our improved version, results associated to the use of the two interaction devices are definitively much more comparable. For what concerns the total execution time, it is still statistically lower in the HTC Vive controller case, 73.8 s, than with the Leap

Motion, 111.9 s (Figure 45 a), but results are now more similar. As shown in Figure 45 b , partial times confirm this result even if the difference is not relevant (3.2 for the HTC controller and 3.4 for the Leap Motion). Regarding the external helps received by the main player, as shown in Figure 46, the supervisor had to rarely intervene and the number of repositioned objects is around 0.5 for the HTC Vive controller and 0.3 for the Leap Motion. As before, this was the only help action actually used. This factor indicates an improvement in the stability of the free-hand interface, as, thanks to the custom interaction algorithm, the player is almost autonomous in the accomplishment of the task. The p-value indicates that no statistically significant difference exists between the two values.

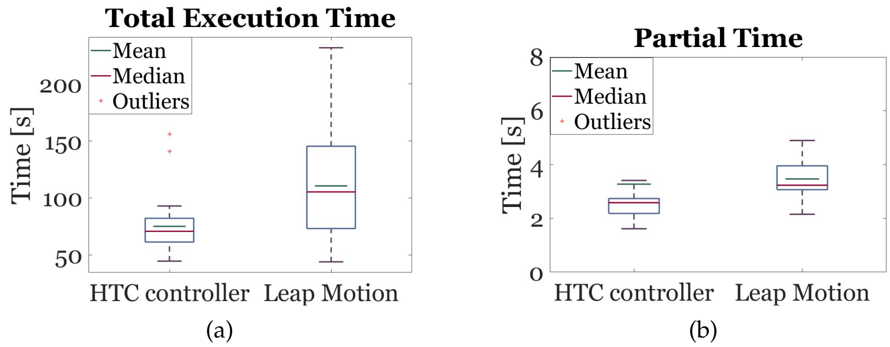


Figure 45: Experiment 2: mean and median of the total execution time (a) and partial time (b), considering the device used.

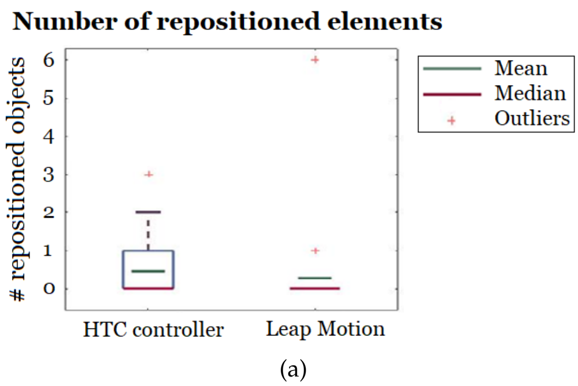


Figure 46: Experiment 2: mean and median of the number of repositioned objects (a) and final score (b), considering the device used.

Considering, instead, the execution order, no learning processes is highlighted for all the examined parameters, although the performance is slightly better in the second case: 99.7 s vs 86 s for the total execution time and 3.7 vs 2.9 for the partial time, 0.4 vs 0.3 for the number of repositioned elements. The p-value confirms that there are no statistically significant differences.

From the cross comparison between interaction modality and execution order, a mild learning process appears. Furthermore, these data confirm that the performance associated to HTC controller and Leap Motion is almost compar-

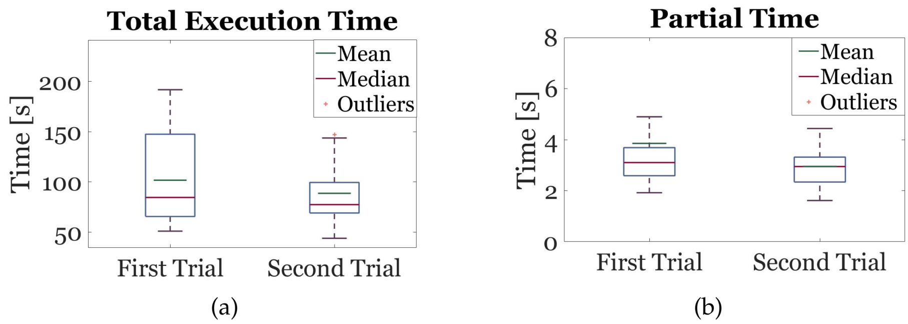


Figure 47: Experiment 2: mean and median of the total execution time (a) and partial time (b), considering the order of execution.

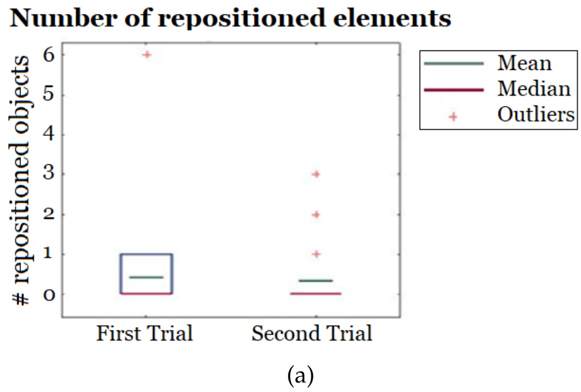


Figure 48: Experiment 2: mean and median of the number of repositioned objects, considering the order of execution.

able except for the total execution time, where higher times are associated to the Leap Motion.

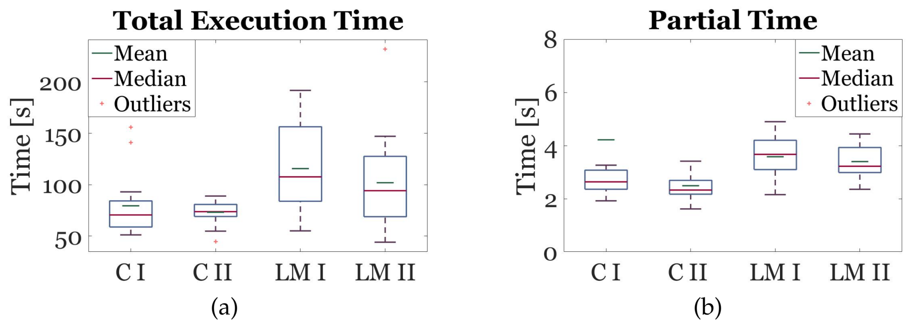


Figure 49: Experiment 2: mean and median of the total execution time (a) and partial time (b), considering the device used and the order of execution.

In the second experimental session, volunteers were divided into two age group, Adult and Children. Adults have, in general, shorter total execution time, while other parameters are comparable (Table 5) and not statistically different in terms of number of repositioned objects and partial time. Moreover, considering the used device, Children and Adults performed remarkably bet-

Parameters	Age Group	HTC controllers	Leap Motion
Total Execution Time [s]	**Adults**	68.1 ± 13.5	105.5 ± 42.5
	Children	112.4 ± 33.4	143.3 ± 53.4
Partial Time [s]	**Adults**	3.2 ± 0.8	4.2 ± 4.5
	Children	2.8 ± 0.2	3.7 ± 4.5
Repositioned elements	**Adults**	0.5 ± 0.7	0.3 ± 1.1
	Children	0.2 ± 0.4	0.2 ± 0.4

Table 5: Experiment 2: mean and standard deviation of total execution time, partial time, number of repositioned elements and final score over all participants considering the device used and the age group.

ter by using the HTC controller, even if this difference is statistically significant only for Adults ($p < 0.05$).

Finally, the open discussion underlines that 66% of participants still prefer to interact by means of the HTC controller, 31% appreciated the Leap Motion, even if they pointed out the greater accuracy and easiness of use of the controller, and 3% feel equally comfortable with both devices.

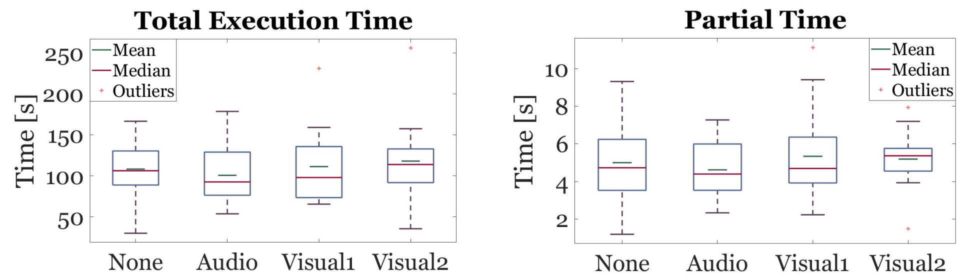
3.2.3.3 *Experiment 3*

Previous experiments proved that our grab algorithm increases stability and reliability of the device, thus positively affecting performance, which are comparable to those obtained with a state of the art technology for interaction, the controllers. The algorithm, however, considered two different aspects: (i) a better control over the physics of the objects and the hand-object interaction; (ii) a visual feedback labelling the hand-object state.

The aim of the current experiment is investigating the impact of these two facets over results, in particular whether the kind of feedback modality (visual, audio or none) can affect performance, by overcoming the absence of haptic feedback during the interaction with VR objects.

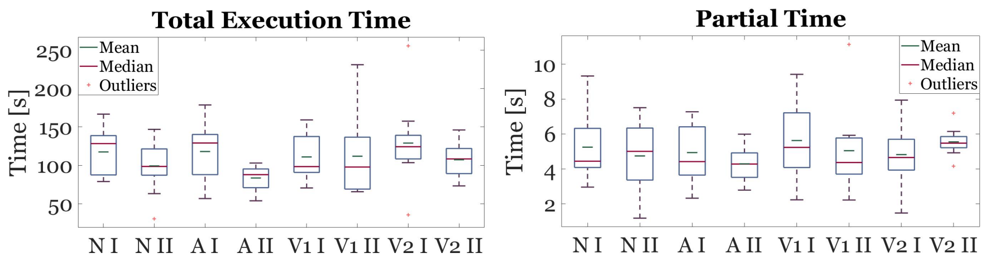
Quantitative data collected during the experimental session have been analysed considering as the independent variables first, the different feedback modalities, then, the order of execution and the feedback modality, finally, the preference expressed by the users. In this case, the number of repositioned objects has not been considered, as the experimenter had to help the main player only in two cases, which can be considered outliers. In each analysis the statistical significance of the differences between samples was estimate by making repeated paired sample t-tests.

If we compare times referred to the different feedback modalities, shown in Figure 50, we can notice that total execution times are comparable (ranging from 100.7 ± 33.5 s to 118.1 ± 44.6 s). However, Audio is slightly better than the other, while Visual2 is the worst. Also partial time are very similar, with the one referred to Audio being in average the lower one (4.6 ± 1.4 s) and the one referred to Visual2 being the higher one (5.3 ± 2.3 s). As stated before,



(a) (b)

Figure 50: Experiment 3: mean and median of (a) the total execution time (averaged across subjects) and (b) partial time (averaged across subjects and objects), considering the performance with different feedbacks.



(a) (b)

Figure 51: Experiment 3: mean and median of (a) the total execution time and (b) partial time, considering the order of execution and the different feedbacks. N = None, A = Audio, V1 = Visual1, V2 = Visual2, while I = first trial and II = second trial.

the difference between partial and total execution time could be due to the fact that while the first is mainly related to the core interaction between user and object, the second takes into account also the strategy used to accomplish the task and time necessary to understand in which hole an object should be inserted. However, we do not directly address cognitive aspects in this study, for this reason the task is simple. Statistical analysis found no between group statistically significant difference, both for total execution time and partial time. So, we can state that feedbacks, even if appreciated in general, do not provide any substantial improvement to performance.

The cross comparison analysis, considering both the feedback modality and the execution order, highlights an effect of learning on total execution time, confirming that this parameter is strongly influenced by the cognitive load of the task. Also partial time slightly decreases in the second trial, except for Visual2 (see Figure 51). Statistical analysis, ran first fixating the feedback modality and considering the order of execution as the dependent variable, then fixating the order of execution and considering the feedback as the dependent variable, found no statistically significant differences, except for total time of the first and second trial with Audio feedback ($p<0.05$) and both total and partial time of the second trial with Audio and Visual2 modality ($p<0.05$).

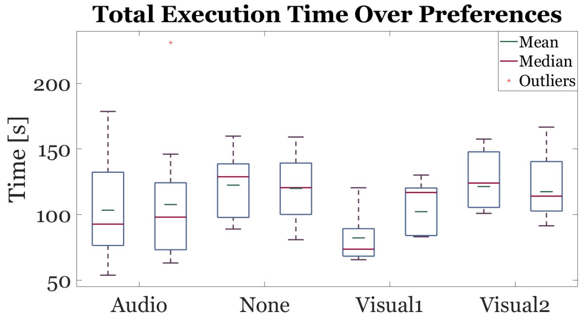


Figure 52: Experiment 3: mean and median of (a) the total execution time and (b) partial time, considering preferences. N = None, A = Audio, V1 = Visual1, V2 = Visual2, while I = first trial and II = second trial.Data are organized in bins, i.e. *Audio*, *None*, *Visual1* and *Visual2*. In each bin two boxes are drawn, the first is referred to total execution time of people who preferred a certain feedback in the trial with that feedback; the second box is the total time of people who preferred that feedback in the non preferred feedback modality trial.

Moreover, Figure 52 shows a lack of correlation between performance and preferences: it is not true that people performed better in the trial with the performance modality they preferred. However, when designing an interface, it is strongly recommended not to ignore testers preferences. T-test performed considering both the feedback modality and the preference as the independent variables highlights no statistically significant difference.

Finally, we examined the User Experience Questionnaire answers. As described in Section 3.2.1.4, this questionnaire provides 6 values, corresponding to the 6 scales of evaluation in which items are divided. Interpretation is quite simple: values between -0.8 and +0.8 represent a neutral evaluation; rates above +0.8 correspond to a positive evaluation, while rates inferior to -0.8 a negative evaluation. It is well known, however, that in general it is hard finding values above +2 and below -2 [92]. In our case, as shown in Figure 53, results show experience is good with all feedback modalities; all the six scales of evaluation for each feedback modality have positive values, widely above 0.8, exception made for the *efficiency* in Visual2 (0.8 ± 0.5). This factor received the lowest ratings, probably because of the problems certain subjects had with hand and gesture recognition. On the contrary, *perspicuity*, indicating how easy is getting familiar with the interaction modality, is well rated. Questionnaire results also show people had a moderate to good control over the interaction modalities and felt stimulated. There is not, however, a clear difference among feedback modalities. Indeed, in general, people liked playing the Shape Sorter game with all the different feedback and could accomplish the task easily. 38% of them expressed a preference for the Audio feedback, asserting that it required less attention in order to be caught, 22% preferred the None modality, 19% the Visual1, 13% the Visual2 and finally 8% stated they have no preference.

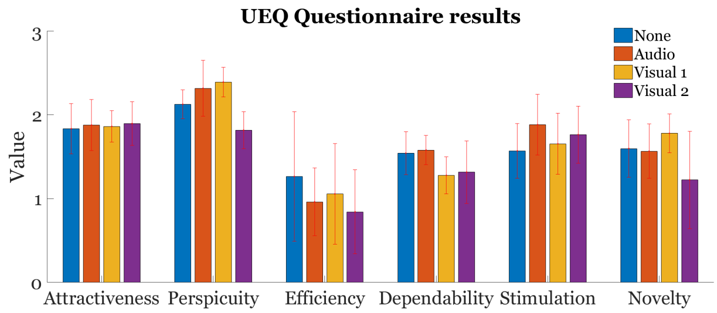


Figure 53: Experiment 3: UEQ results for the 6 scales of evaluation referred to the 4 feedback modalities.

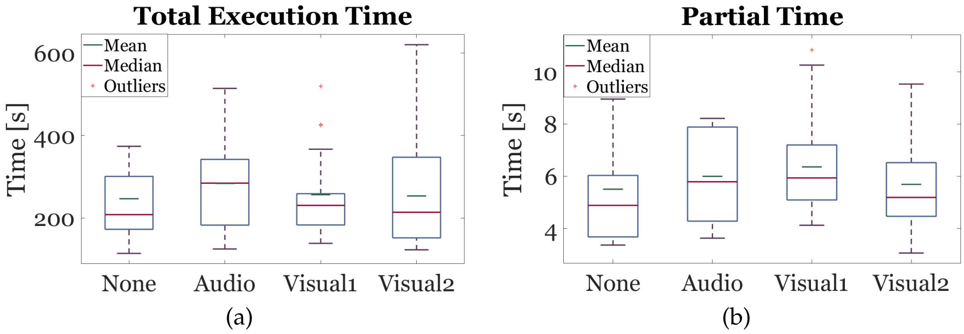


Figure 54: Experiment 4: mean and median of (a) the total execution time and (b) partial time, considering the performance with different feedbacks.

3.2.3.4 *Experiment 4*

Even if feedback seem having no effect over performance in the selection and manipulation simple task, this does not exclude that they could be useful during the accomplishment of a more complex task. The second scenario and the assembly task were thus designed and implemented and participants performance and preferences were, again, evaluated.

As in the previous experiment, quantitative data have been analysed on three levels: considering first the feedback modalities, then, both the execution order and the feedback, finally, the feedback modalities and the preferences. In each analysis the statistical significance of the differences between samples was estimate by making repeated paired sample t-tests.

Total execution times referred to different feedback modalities are comparable (see Figure 54 a), and vary in a small range from 246.5 s, in the case of None, to 283.7 s, in the case of Audio feedback. As shown in Figure 54 b, partial times confirm this trend. Taking into account that the suit pieces were 12, the difference between partial and total time, indicates that people spent the majority of their time organizing a strategy and understanding how to position different objects than on the interaction itself. The t-test we performed to calculate the between group statistical significance never rejected the null hypothesis, i.e. differences among different feedbacks are not statistically significant.

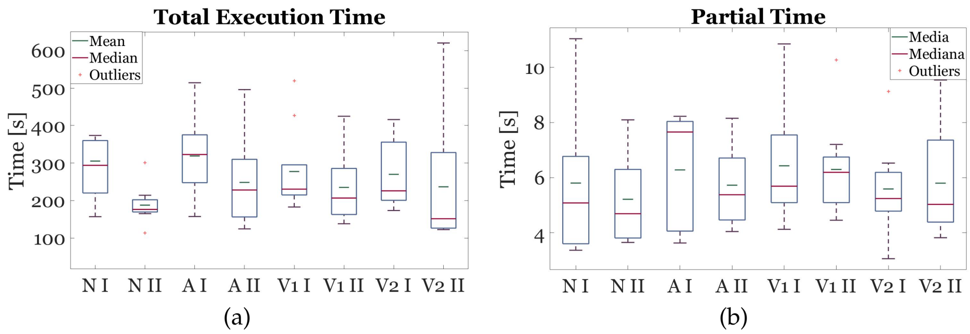


Figure 55: Experiment 4: mean and median of (a) the total execution time and (b) partial time, considering the order of execution and the different feedbacks. N = None, A = Audio, V1 = Visual1, V2 = Visual2, while I = first trial and II = second trial.

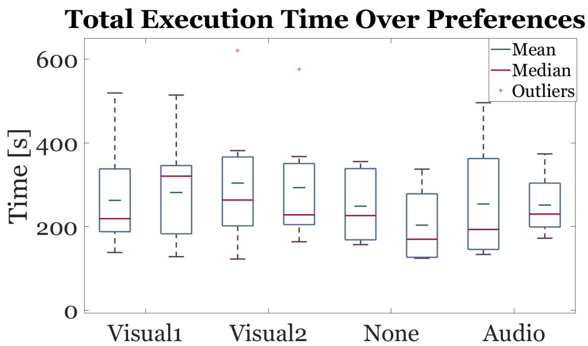


Figure 56: Experiment 4: mean and median of (a) the total execution time and (b) partial time, considering preferences. N = None, A = Audio, V1 = Visual1, V2 = Visual2, while I = first trial and II = second trial.Data are organized in bins, i.e. *Audio*, *None*, *Visual1* and *Visual2*. In each bin two boxes are drawn, the first is referred to total execution time of people who preferred a certain feedback in the trial with that feedback; the second box is the total time of people who preferred that feedback in the non preferred feedback modality trial.

Considering both the order of execution and the interaction modality as the independent variable, we can notice a learning process more evident in the case of the total execution time in Figure 55 a, as participants have learnt how to build the suit. Partial times slightly improve in the second trial (see Figure 55 b), except for Visual1. Repeated paired t-tests performed in order to find some between or within group statistical significance, found that all these differences are not statistically significant.

Furthermore, performance are not dependent from preferences, i.e. volunteers total execution times are not better in the task accomplished with the feedback modality they preferred (Figure 56). Again within and between group differences are not statistically significant.

Finally, UEQ questionnaire results are quite interesting: for each feedback all the values referred to the different scales are above +0.8, meaning positive opinions. Rates on *efficiency* and *dependability*, indicating the level of control user feels having on events, are the lowest one, while *attractiveness* and *perspicuity*, i.e. how easy is learning how to use the interface, are the highest. Visual1 received the worst rate, Visual2 the best, Audio and None are in general comparable. While people appreciate the naturalness and intuitiveness of the interaction through the virtual hand, in fact, they often complain about the perceived poor control.

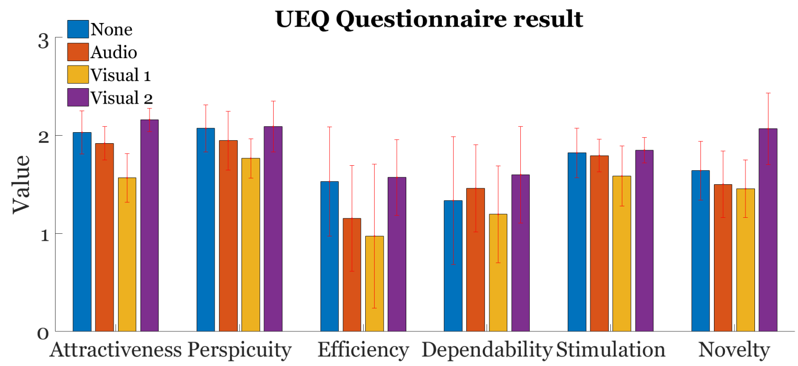


Figure 57: Experiment 4: UEQ results for the 6 scales of evaluation referred to the 4 feedback modalities.

Finally, in general, everybody completed the task successfully and liked the experiment, only 9% of participants said they did not appreciate it. Half of the volunteers preferred the visual feedback (31% Visual 1 and 19% Visual 2), 14% said the None was the best one, 11% chose the Audio and 25% had no preference. In this case, participants preferred the visual feedback to the Audio one, maybe because of the presence of multiple audio sources, i.e. the soundtrack and the metallic sound played when a piece was correctly positioned.

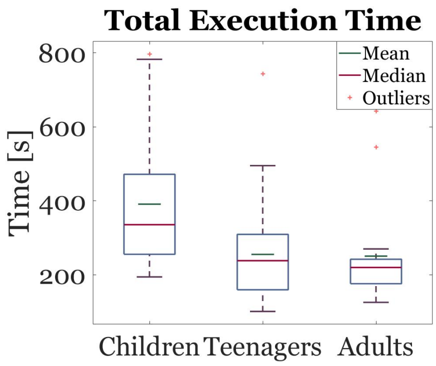
Answers to the overview questionnaire (on a Likert scale from 1 to 5) suggest that people found the interaction intuitive (4.2 $\pm$ 0.7 points) and the experience immersive (4.3 $\pm$ 0.6 points). Nobody felt symptoms of simulator sickness (1.1 $\pm$ 0.4 points) and in general participants found more difficult positioning suit pieces(2.9 $\pm$ 0.9 points), so using the interaction device, than understanding what they corresponded to (1.9 $\pm$ 1.1 points).

3.2.3.5 *Experiment 5*

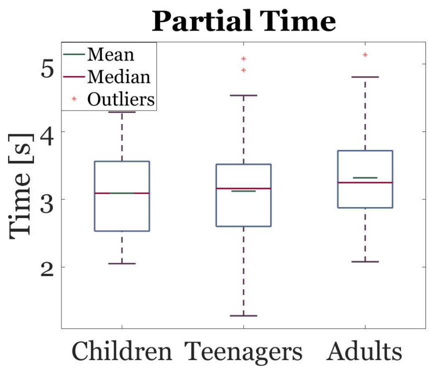
In general, every participants could accomplish the task, except for two children (5 and 6 years old respectively), who were excluded from the final analysis.

Quantitative data collected during the experimental acquisition have been divided in three groups based on player's age (Children, Teenagers and Adults). Performance, in fact, differs between groups. In each analysis the statistical significance of the differences between groups has been estimated by performing a paired sample t-test analysis.

Considering the total execution time shown in Figure 58 a, we can notice that Children have more difficulty in completing the task (390.9 $\pm$ 180.9 s) with respect to Teenagers (225.1 $\pm$ 129.8 s) and Adults (250.4 $\pm$ 135.3 s), whose



(a)



(b)

Figure 58: Experiment 5: mean and median of (a) the total execution times and (b) the partial times divided based on participant age.

performance is almost comparable. Variability is also higher. T-test analysis, in fact, highlights a statistically significant difference between Children and Teenagers ($p<0.001$) and between Children and Adults ($p<0.01$), while Teenagers and Adults are not statistically significant different. Instead, partial times are very similar, ranging from 3.1 s for Children and 3.3 s for Adults, and differences are not statistically significant (see Figure 58 b). Furthermore, partial times, obtained in this experiment with HTC Vive controllers, are shorter if compared to those obtained in Experiment 4 with all feedback (Figure 50 b), indicating the touchful interface allow a more natural and reliable interaction for the complex task accomplishment. In both experiments, instead, total execution time is more influenced by cognitive skills of the player and strategy adopted than by the intuitiveness or efficiency of the interaction: in fact Children tended to have more difficulties in recognizing what suit pieces corresponded to or in understanding the instruction, i.e. they positioned objects in the cylinder but did not connect them to other suit pieces losing time.

Answers given to the overview questionnaire confirm this hypothesis: for participants it is more difficult understanding what pieces correspond to (3.9 ± 1.1 points) than actually positioning them (2.9 ± 1.2 points). Everyone liked the game and had no simulator sickness (1.3 ± 0.9 points), commands have been considered as intuitive (4.2 ± 1.1 points) and environment immersive (4.7 ± 0.8 points).

3.2.4 *Outcome*

The aim of our research is the evaluation of the best method for the interaction within an immersive virtual environment, considering on the one hand its naturalness and intuitiveness and on the other hand its stability and reliability. To this purpose, performance and preferences have been quantified and analysed. The used metrics for performance evaluation includes the number of times the experimenter had to intervene to help the participant, by repositioning objects, the total execution time and the partial time, necessary to position each item. In general, in all experiments, partial time reflects the usability, stability and ease of use of the interaction modality; while total execution time resumes different characteristics, like interaction related aspects but also task complexity and user cognitive skills, which are out of the scope of this paper. For preferences, instead, we used both direct interviews and a validated questionnaire, the User Experience Questionnaire. As a supplementary analysis, we evaluated participants physical condition after the exposure to VR, through the Simulation Sickness Questionnaire.

Two of the most diffuse solutions for interaction in VR have been taken into consideration, tested and compared: a touchful device, the HTC Vive controller, and a touchless one, the Leap Motion. The former is the better in terms of performance, because the Lighthouse basestations combined with inertial sensors ensures a stable and accurate tracking; whereas the latter is prone to tracking issues and occlusions, responsible for an unstable interaction. However, users in general appreciate the possibility to see a virtual representation of their hands and use it to interact with the surrounding environment. We implemented a grab function to be used with the Leap Motion, in order to overcome its stability issues and improve performance, and we tested it using two selection and manipulation tasks in two different ad hoc designed scenarios, the Shape Sorter and the Ironman. In the Shape Sorter game, participants are asked to put 12 items, different in colour and shape in the corresponding hole; while in the Ironman game they have to assembly the movie character suit starting from building blocks. The two tasks have a different complexity level, both in terms of fineness of the items manipulation and required cognitive skills. In Table 6, total execution and partial times referred to the two scenarios and divided by age group show that, while interaction with items is comparable, the cognitive load of the task is different and strongly affect the time required to accomplish the task, which is always higher in the Ironman case.

Parameters	Age Group	Shape Sorter	Ironman
Total Execution Time [s]	Adults	68.1 ± 13.5	250.4 ± 135.3
	Teenager		255.1 ± 129.8
	Children	112.4 ± 33.4	390.9 ± 180.9
Partial Time [s]	Adults	3.2 ± 4.5	3.3 ± 0.8
	Teenager		3.1 ± 0.8
	Children	2.8 ± 0.2	3.1 ± 0.7

Table 6: Mean and standard deviation of total execution and partial times obtained by different age group in the two selection and manipulation tasks with the HTC Vive controller.

For visualization, we used the Oculus Rift DK2, in Experiment 1, and the HTC Vive in the other experiments. For standing applications, where interaction is in the peripersonal space, the two systems are equivalent, while only

HTC Vive can ensure a stable interaction in the near-action space, thanks to the possibility it offers to design room-scale setups.

As shown in Table 7, interaction through HTC Vive controllers has in general better results than interaction with the Leap Motion. However, our grab algorithm ensures comparable performance. Total and partial times referred to the different feedback modalities are similar, indicating that the improvements brought by our algorithm depend on the more restrictive control on the behaviour of the objects and on the distinction of hand-object states (no collision, collision detected, grab action recognized). So, controllers and Leap Motion allow to obtain comparable performance in simple and complex task, but Leap Motion require more control over the physical behaviour of interactable items, i.e. objects are blocked when no collision is detected.

Considering preferences, subjects in general prefer HTC Vive controllers because of the greater simplicity of the interaction with virtual objects and the better stability. Our grab algorithm, however, improve performance thus increasing preferences, from 13% to 31%. Moreover, even if feedback modalities provided with the touchless interaction do not affect performance, they are appreciated. In the Shape Sorter scenario there is not an evident preference between Audio (38%) and visual (32%), including Visual1 (19%) and Visual2 (13%), while in the Ironman scenario visual feedback receives 50% of preferences (31% for Visual1 and 19% for Visual2) and Audio only 11%. This difference is due to the fact that, in the first case, the Audio feedback during grabbing was the only music played whereas, in the second case, there were multiple sound sources, i.e. a soundtrack and the metallic click when a suit piece was correctly positioned. Nonetheless, it is worth noting that performance is not dependent on preferences.

Finally, even if *sense of presence* was not directly addressed in the presented experiments, engagement, involvement, spatial presence and experienced realism are in general high. For instance, when objects where occluded by other items in the VR scene, people tended to step over and avoid the occlusive object, instead of penetrating it, which was certainly the more logical solution, being VR elements unreal. Moreover, multiple time, at the end of the trial participants tried to place the controllers on the virtual table.

Parameters	Scenario	HTC Vive	LM - None	LM - Visual1	LM - Visual2	LM - Audio
Total Time [s]	**Shape Sorter**	75 ± 22.9	108.3 ± 34.6	111.4 ± 42.8	119.1 ± 43.6	100.7 ± 32.6
	Ironaman	253.6 ± 130	246.5 ± 110.6	256.3 ± 107.1	253.3 ± 131.7	283.7 ± 115.6
Partial Time [s]	**Shape Sorter**	3.3 ± 4.3	5 ± 1.9	5.3 ± 2.3	5.1 ± 1.4	4.6 ± 1.4
	Ironman	3.2 ± 0.8	5.5 ± 2.2	6.4 ± 1.8	5.7 ± 1.8	6 ± 1.8

Table 7: Mean and standard deviation of total execution and partial times obtained in the two scenarios with the HTC Vive controllers and the Leap Motion combined with the grab algorithm.

3.3 Industrial Use Case

This work aims at assessing if immersive Virtual Reality ship handling simulation systems can be valid substitute of non-immersive systems for training of nautical personnel. In the literature, in fact, simulation has been proven to be important in all those domains where complex and demanding tasks are required and ship handling simulators, in particular, have always taken advantages from non-immersive VR technologies and computer-based environments representing a replica of the real world, in which the ship is operating [15, 165, 166]. However, the diffusion of immersive virtual reality HMDs gives the users the possibility of being involved and totally immersed in synthetic environments for more realistic experiences. Even if researchers still debate if higher *sense of presence* depends on a combined effect of VE platforms type (immersive or non-immersive) and narrative contexts (emotional or non-emotional) [65] or emotional context only [66] or on user characteristics [79], many studies have confirmed a positive relationship between the level of immersion and behavioural responses and performance in different application domains [5, 67, 143]

The current work has been accomplished in collaboration with a local company with a strong background in computer graphics and physical simulation, CETENA S.p.A., in the context of the project MIT - Leadership Tecnologica[5]. The goals of the collaboration were the creation and the evaluation of an immersive VR ship handling simulation system. CETENA S.p.A. provided the software framework and the Virtual Reality setup; whereas, the role of DIBRIS Department consisted in the assessment of the immersive VR system usability for training purposes as a substitute of traditional non-immersive solutions.

First of all, a simulation system is required to provide a realistic handling experience, for this reason the testers in the preliminary acquisition phase were expert naval users, i.e. students of the Genoa Naval Academy. Their ability to follow a predefined path and their experienced *sense of presence* were analysed in order to understand if the immersive and non-immersive VR systems are comparable or if one of them ensures better performance and a more realistic experience. Moreover, the prolonged use of immersive systems might produce *cybersickness*, which is an effect we want to avoid firstly because it discourages people from participating in multiple sessions and secondly because it can negatively affect performance and hinders learning. Thus, in order to monitor participants' *cybersickness* state, some physiological measurements were recorded during the experimental sessions and people were asked to fill the SSQ after each VR exposure. The fact of testing the system on expert users, on the one hand, biases results, as they have a higher threshold for motion sickness, on the other hand, allows to discerns between the effect of the simulated boat motion and the effect of the visualization technology.

5 The involved partners are the Company Cetena S.p.A. and the Department DIBRIS of the University of Genoa, Italy

3.3.1 *Material and Methods*

Two different setups have been designed and implemented: the traditional simulation non-immersive Virtual Reality system, based on three 27 inch standard monitors, disposed vertically side by side in order to mimic the view from the ship command bridge, and the immersive VR system, based on the Oculus Rift (see Figure 59). In both cases, the user is required to wear a safety helmet, with a HTC Vive tracker attached on it, for the purpose of tracking his head position and rotation.

(a) (b)

Figure 59: Industrial use case: the two visualization modalities of the ship handling simulator: (a) non immersive with monitors and (b) immersive with the HMD.

The physical reproduction of a ship command panel completes the setup. People can handle the virtual ship by rolling a real rudder and moving the accelerator knobs. It is worth noting that, in the non-immersive case, the command panel is completely visible to the user, while, in the immersive case, even if the user has the haptic and force feedback of the panel, he cannot see it, so, a slider showing rudder rotation is shown in the virtual environment.

The simulation is set offshore, in a area of about 50 km x 50 km around Genoa harbour. The scenario is taken from a 3D graphics database inspired to real coastal areas. Rendering is not detailed to avoid to distract the user with an excess of visual information, while maintaining a good level of involvement, determined by the recognition of familiar scenes.

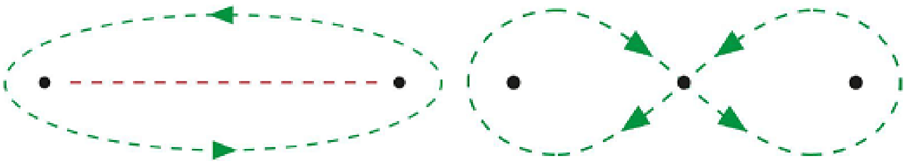
Different simulation parameters, e.g. the duration of the simulation, the boat type and the path, can be set by the experimenter. In particular, it is possible to choose between two ships, whose physical behaviour has been reproduced in a highly detailed way: a patrol boat and a coastguard, afterwards referred to as Fast and Slow boat, respectively (see Figure 60). The patrol boat is a small light vessel (15 m long and 20 t of weight), that can reach 50 knots of speed. The coastguard, instead, is a larger ship (20 m long and 50 t of weight) that can reach a maximum speed of 20 knots. The aesthetics and the operating mechanisms of these two ships have been modelled and implemented as faithfully as possible.

Moreover, experimenter can assign one of the two paths: the Ellipse, the simpler one, and the Eight path, the more complex (see Figure 61), as it implies

(a) (b)

Figure 60: Industrial use case: pictures of the two kind of boats modelled for the ship handling simulator. (a) The Slow ship, a coastguard, and (b) the Fast ship, a patrol boat.

a greater freedom of movement and, thus, a higher probability of losing the original route. Participants are required to have a better spatial mapping and control over the ship, being aware of its turning rate, i.e. time and space required to veer. The path is visible to the user as a green route just above the sea. Weather conditions are an additional feature of the system: in the Eight path, sea is calm and plain, while in the elliptic path, sea is more rough.



(a) (b)

Figure 61: Industrial use case. The two paths, (a) Ellipse and (b) Eight.

3.3.2 *Experiments*

During the experimental session, a mixed experimental design was used: within groups for boat type and visualization modality and between groups for the path shape. In fact, all participants accomplished the task both with the slow and fast vessel with either the monitor and the HMD, but half of them has to follow the Ellipse route and half the Eight one. The order of execution is fixed: participants start with the monitor simulation and the slow boat (Monitor Slow), then the fast boat (Monitor Fast); after this, they wear the Oculus Rift and accomplish the task in the immersive VE with the slow (HMD Slow) and fast (HMD Fast) vessel. Each trial lasts 5 minutes, while the experiment has a total duration of 45 minutes.

Kind of boat	Number of people
Fast smaller than 15 m	8
Fast longer than 15 m	2
Slow smaller than 40 m	3
Slow longer than 40 m	2
None	4

Table 8: Industrial use case: boats usually driven by participants.

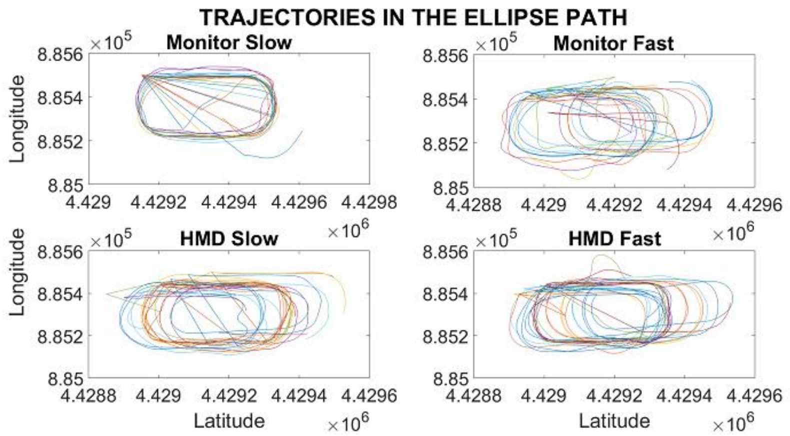
In the beginning, the experimenter explained the modality and the purpose of the test, the different tasks, the setup and the instrumentation used, whereas subjects had to sign a written consent and the privacy policy and fill in an anonymous module giving personal information (age, genre, previous experience in simulation, VR environments, and boat driving). Afterward, two different sensors were attached to the participant: the Scosche Rhythm armband, for heart rate recordings and the Mindfield eSense Skin Response sensors for skin conductance. The first was connected to a smartphone Galaxy S4 via Bluetooth and exploited the BLE Heart Rate Monitor software to memorize and send 1 sample/second, whereas the second was connected to the phone by wire and used a proprietary software to record and send 5 samples/second. Sensor choice and position have been thought in order not to interfere with participants movements or cause discomfort, reducing *sense of presence*. Next, participant was asked to fill in the first Simulator Sickness Questionnaire (SSQ Pre) and to accomplish the four tasks. After each trial, he had to complete a separate SSQ, in order to control his state of malaise. Participants were, furthermore, told that they could interrupt the experiment whenever they wanted. A brief introductory tutorial phase preceded the use of each new hardware setup and allowed users to familiarize with the interface and visualization system. Finally, volunteers were asked to fill in an IPQ for each trial.

20 volunteer healthy male subjects aged between 20 and 24 years (21.8 $\pm$1.1 years) participated to the experimental session. They had normal or corrected to normal vision. The majority of them were naive towards Virtual Reality (74%), while 5% had already took part to experiments involving VR systems. All of them were expert boat drivers (see Table 8).

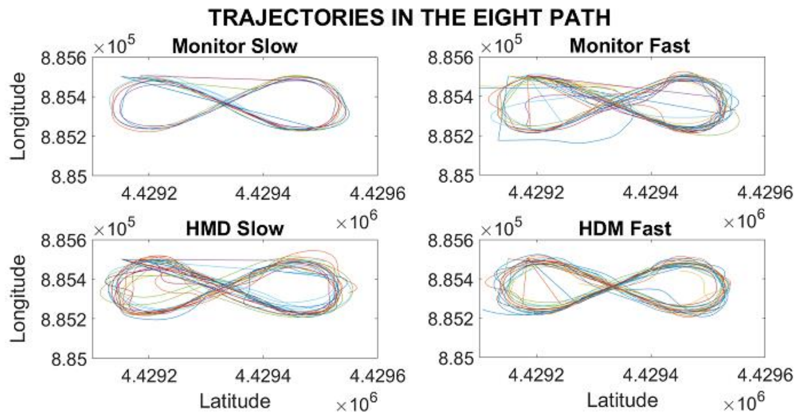
3.3.3 *Data Analysis*

The analysis of the recorded data aims to evaluate and compare an immersive and a traditional non-immersive system, in order to assess if the former can be a valid substitute of the latter in the training of nautical personnel. In particular, we considered performance, through trajectories, *sense of presence*, through IPQ and head rotation, and *cybersickness*, through SSQ, skin conductance and heart rate.

Participants trajectories, obtained from the analysis of the latitude and the longitude of the virtual boat during task execution, are quite accurate: the original path shape is easily identifiable in both visualization systems, especially concerning the Eight path data (Figure 62). In the Ellipse path case, the multiple decentralized trajectories are due to problems during scene initialization and data acquisition. Unluckily, as we do not have access to the actual paths' coordinates, we can do only a coarse and qualitative evaluation.



(a)



(b)

Figure 62: Industrial use case: trajectories of all participants with the Ellipse(a) and Eight path (b).

Sense of presence was measured considering both answers given to the IPQ and the tendency of people to freely explore the VE, by rotating their head. Data from IPQ were collected and analysed, firstly, based on the kind of trial, secondly, taking into account also the path, thirdly considering people's previous experience with VR and ship handling. In general, IPQ rates given to the three subscales that compose the questionnaire (Experienced Realism, Spatial Presence and Involvment) are good for both visualization modality but better for the HMD, indicating a higher *sense of presence* (Figure 63 a). Moreover, trials with the Fast boat have less Spatial Presence but greater Experienced Realism if compared to the trials with the Slow ship, maybe because of the realistic real-time physical behaviour reproduction and the response speed of the vessel to user's commands. A between group Wilcoxon test was performed and only the difference of Involvement and Spatial Presence parameters in the monitor and HMD trials has been found to be statistically significant. This means that the use of the Oculus Rift allows the user to feel more involved and present in the virtual simulated environment. Considering the Presence Factor (Figure 63 b), trials with the HMD obtained better results moreover HMD Fast grade is statistically different from Monitor Slow and Fast grade.

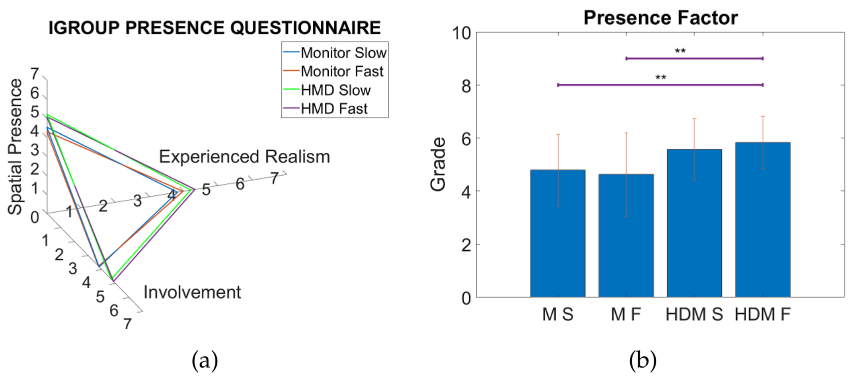
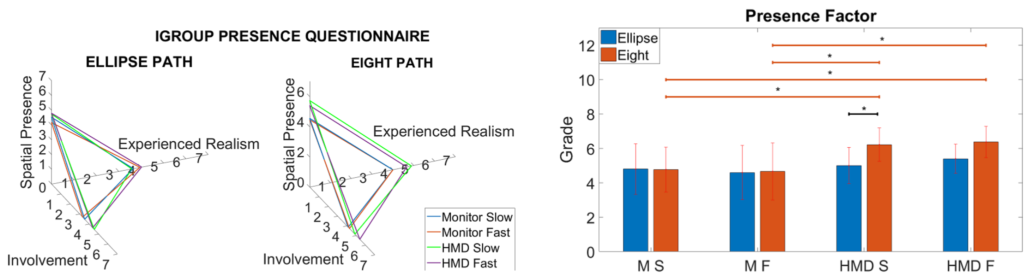


Figure 63: Industrial use case: results of the four IPQs organized considering the trial. (a) Results divided based on the three evaluation subscales. (b) Mean of grades of the Presence Factor. M = monitor, HDM = head-mounted display, S = slow, F = fast. * p-value<0.05 e ** p-value<0.02.

In the path shape analysis, differences between monitor and HMD are more evident in the Eight case than in the Ellipse case (Figure 64 a). In fact, in the Ellipse path only the difference between Monitor Fast and HMD Slow and Fast for Involvement is statistically significant ($p<0.05$); whereas, in the Eight path all Spatial Presence rates referred to monitor and headset are statistically significant ($p<0.05$) (see Table 9). The Presence Factor, shown in Figure 64 b, is better with the HMD, confirming previous results. This trend is more evident in the Eight path, were results with both visualisation modalities are significantly different ($p<0.05$).

We further noticed that, people who had already tried VR tended giving higher score to the simulation systems, even if differences between experts and not experts are not statistically significantly different. Finally, the level of expertise of participants on ship handling had no effect on the results.

Considering head rotation angles, it is worth noting that the tendency to explore the VE is subjective: people can be more or less prone to exploration. For each participant, head rotation angles were extracted and their histograms were calculated. This way, it is possible to highlights users' preferential head rotation angle and distribution across a trial. Referring to Unity coordinate system, yaw is the rotation around the vertical axis (Y), pitch is the rotation around lateral axis (X) and roll is the rotation around the sagittal axis (Z) (Figure 67). Only yaw is relevant for this analysis as it represents the tendency

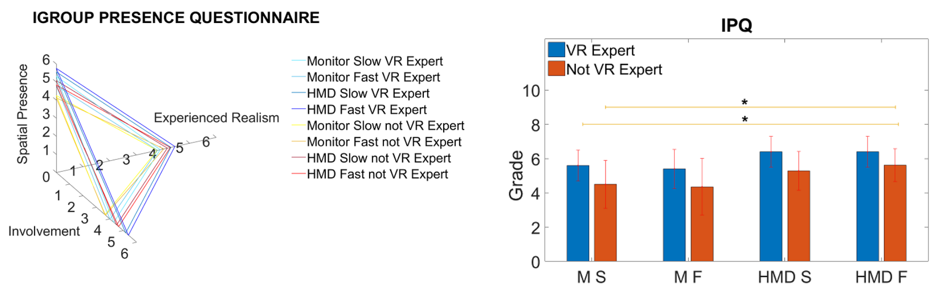


(a) (b)

Figure 64: Industrial use case: results of the four IPQs organized considering the trial and the path. (a) Results divided based on the three evaluation subscales. (b) Mean of grades of the Presence Factor. M = monitor, HDM = head-mounted display, S = slow, F = fast. * p-value<0.05 e ** p-value<0.02.

		SP	I	ER
Ellipse	**M F-HMD S**		0.0254	
	M F-HMD F		0.0409	
Eight	**M S-HMD S**	0.0030		
	M S-HMD F	0.0427		
	M F-HMD S	0.0110		
	M F-HMD F	0.0472		
Cross	**HMD S-HMD S**	0.0178		0.0261

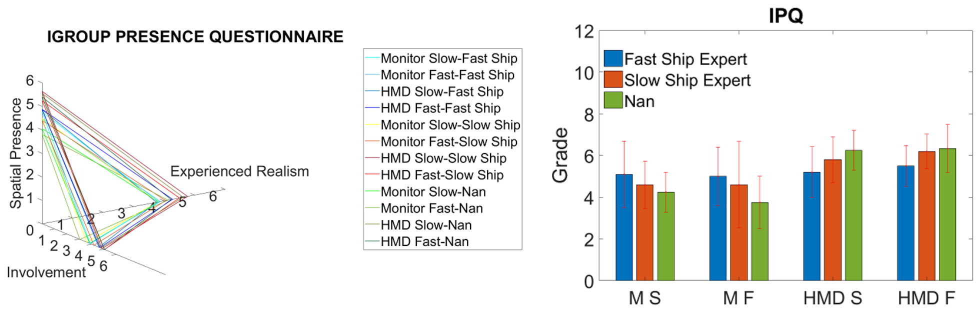
Table 9: Industrial use case: p-value obtained making a within group Wilcoxon test and comparing Spatial Presence (SP), Involvement (I) and Experienced Realism (ER) grades in the IPQs considering the two path separately. Cross refers to the between group Wilcoxon test, made comparing Ellipse and Eight path trials.



(a) (b)

Figure 65: Industrial use case: results of the IPQs organized considering people experience with VR setups. (a) Results divided based on the three evaluation subscales. (b) Mean of grades of the Presence Factor. M = monitor, HDM = head-mounted display, S = slow, F = fast. * p-value<0.05.

of participants to look around. Figure 68 shows results referred to two participants who accomplished the task with Ellipse (a) and Eight (b) paths. In



(a) (b)

Figure 66: Industrial use case: Results of the IPQs, considering people experience with ship handling. (a) Results divided based on the three evaluation subscales. (b) Mean of grades of the Presence Factor. M = monitor, HDM = head-mounted display, S = slow, F = fast.

the monitor case, rotations are scattered around a central value, which corresponds to the initial head rotation when the application starts; while in the HMD case, especially in the Eight path, participants tended to turn their head, in order to receive more information from the surrounding environment and better follow the path. This is due to the higher difficulty of the task.

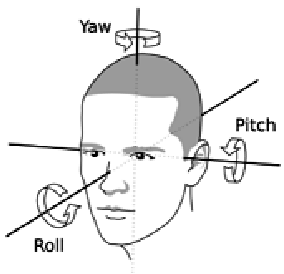
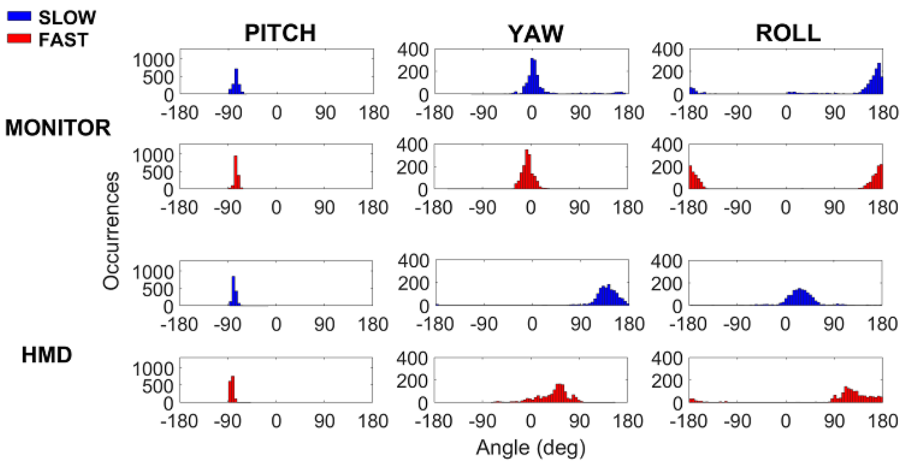


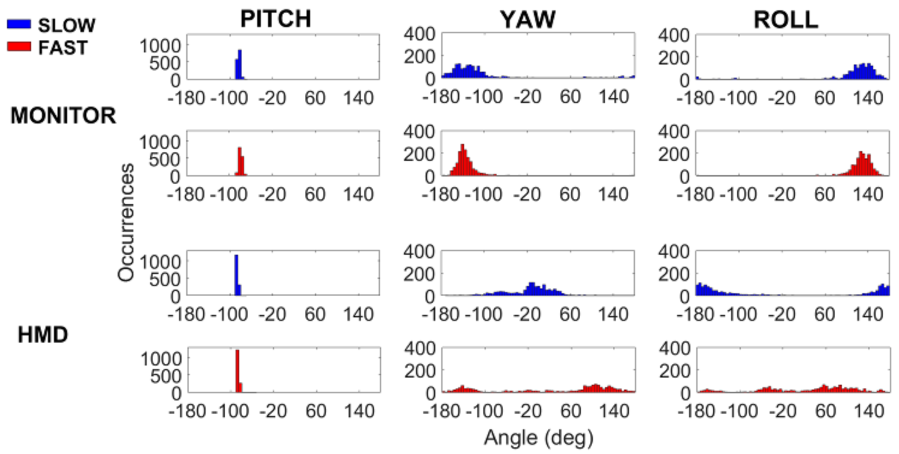
Figure 67: Head rotation angles schema: Yaw is the rotation around the vertical axis (y), Pitch is the rotation around lateral axis (x) and Roll is the rotation around the sagittal axis (z).

Skin conductance and heart rate were recorded as a measure of well-being and change in participant emotional state. Again, measurements were analysed considering first the trial (Monitor Slow, Monitor Fast, HMD Slow, HMD Fast) and then both the trial and the path. Additionally, also participants previous experience with VR setups and ship handling were taken into account, in order to investigate their influence on *cybersickness*. In particular, we expect that people non-naive towards VR would be less excited, anxious and prone to *cybersickness*. Nonetheless, students with practical experience in ship handling should be more confident about the task.

Results obtained from the first analysis of skin conductances are shown in Figure 69 a. In the monitor case, in general, skin conductance is stable and constant, even if slightly higher in the trial with the faster ship, probably because of the greater difficulty of the task. This can be due to the fact that participants



(a)

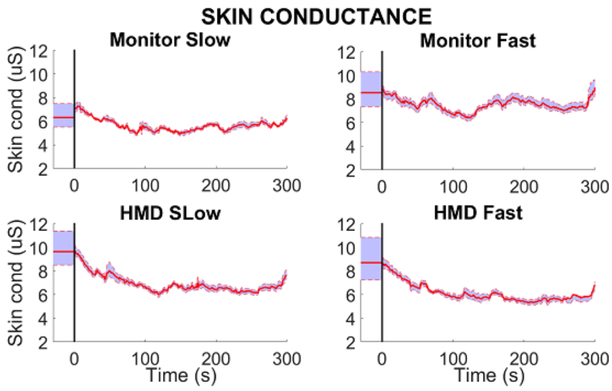


(b)

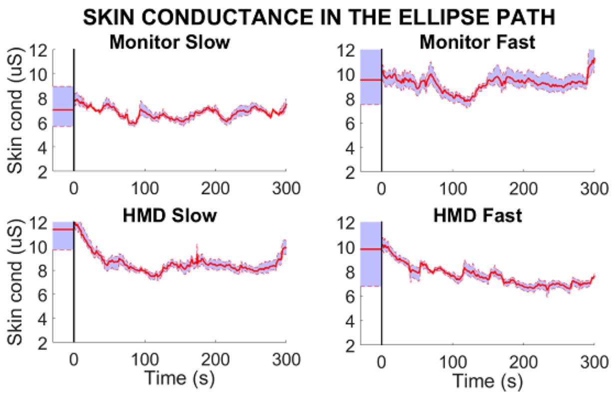
Figure 68: Industrial use case: non normalized histograms of Pitch, Yaw and Roll head rotation angles of participant 1 (a), who did the trial with the Ellipse path, and participant 2, who did the trial with the Eight path (b). Results are divided based on the trial.

face for the first time the fast task and still have little confidence with the setup, hardware and software. This effect, however, is attenuated in the following trials. Future analysis with a counterbalanced presentation of the setups could better explain this trend. In the HMD case, it initially fast decreases and then settles around a stable value, similar to the one recorded during the simulation with the monitor. This descending trend could be associated to people expectation over the trial difficulty, or the excitement of using a new hardware system. These results are confirmed also by the analysis taking into account both the trial and the path, with higher baseline in the Ellipse case (Figure 69 b and c).

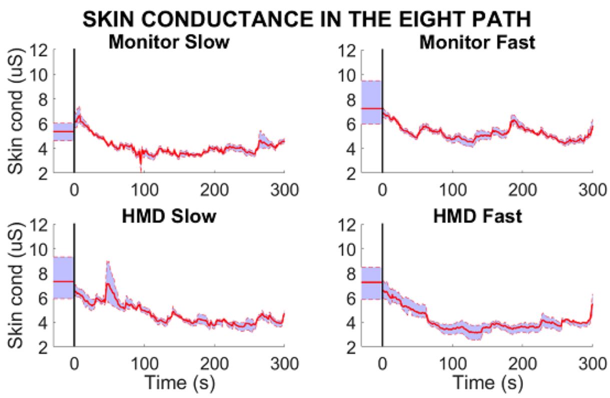
The fact of having already tried VR has a mild influence on participants, both in the baseline and general trend. As shown in Figure 70, subjects new to



(a)



(b)



(c)

Figure 69: Industrial use case: average skin conductance considering first the trial (a) then the trial and the path, Ellipse (b) or Eight (c).

Virtual Reality show higher value of skin conductance, probably due to their state of excitement.

Finally, results obtained grouping participants based on their previous experience on ship handling (Figure 71) suggest an influence of the task confidence, in fact non-expert students have higher and more irregular skin conductance.

Considering heart rate, in general, it is regular, with no peaks in none of the

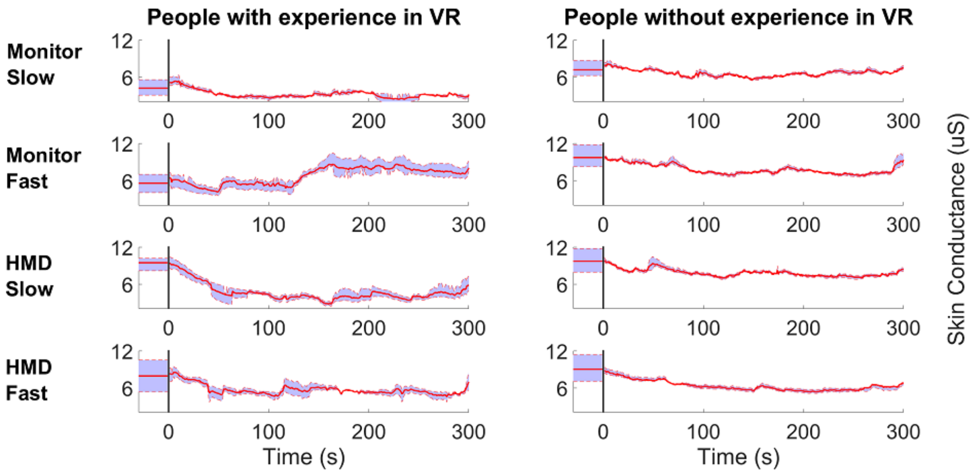


Figure 70: Industrial use case: average skin conductance of participants based on the trial with (left) or without (right) previous experience in VR.

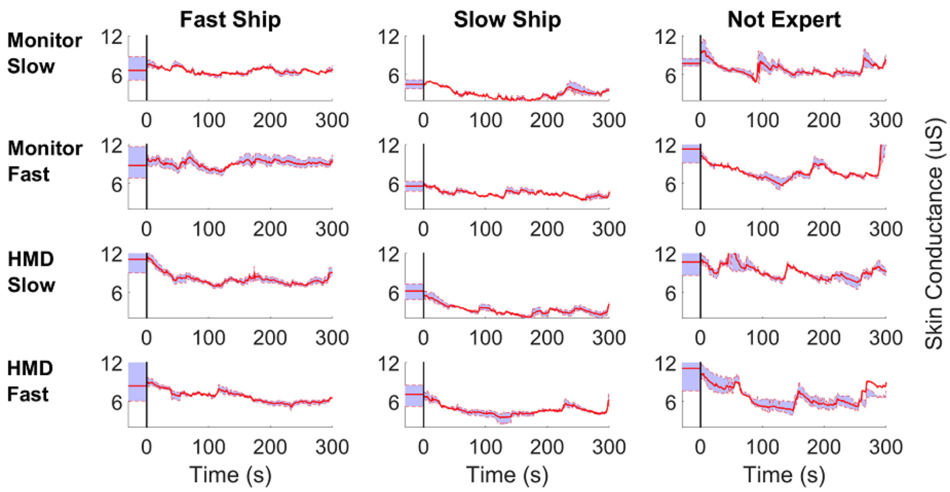
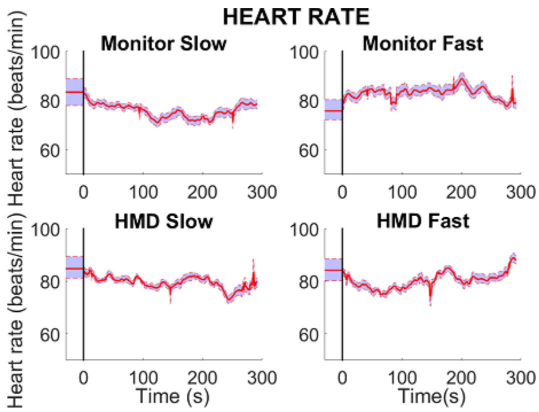


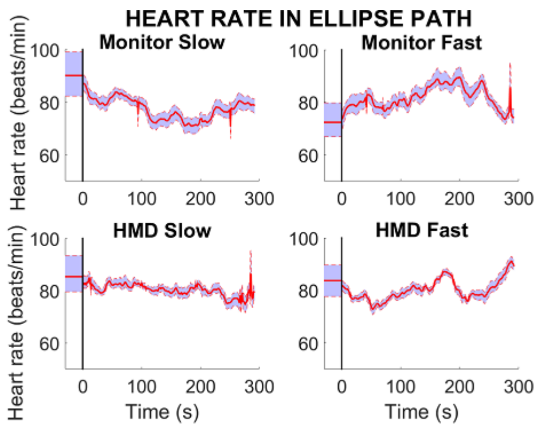
Figure 71: Industrial use case: average skin conductance of participants based on the trial considering their previous experience with different kind of ships.

four trials (Figure 72 a). This indicates that both visualization systems do not introduce particular emotional states that could compromise performance and interfere with learning. The absence of elevated heart rate values (around 100 beat/min) could indicate that participants have perceived the simulations as natural experiences, comparable to real life driving experiences, and is associated to an absence of *cybersickness* symptoms. Slightly higher value in the fast trials are probably caused by the major difficulty of the task and can indicate a higher level of stress. The analysis of data referred to the two paths confirms these results (Figure 72 b and c).

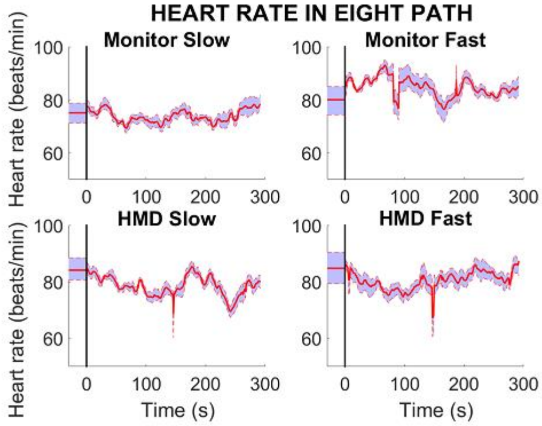
Similarly to skin conductance, data were further analysed based on participants previous experience with VR. As shown in Figure 73, people who have



(a)



(b)



(c)

Figure 72: Industrial use case: average heart rate considering first only the trial (a) then the trial and the path, Ellipse (b) or Eight (c).

never experienced Virtual Reality, tend to have slightly higher heart beat rate in the initial phase of the simulations, probably because of the excitement/anxiety of using a new device. However, both VR naive and non-naive subjects do not show elevated heart rate values (around 100 beat/min), excluding the presence of *cybersickness*.

Finally, considering participants previous experience on ship handling, while expert students maintain standard heart rate values, with the only exception of the first trial with the fast boat, not-experts show higher and more irregular heart beat rates (see Fig. 74). This confirm an influence of task difficulty or perceived difficulty on participants emotional and mental state.

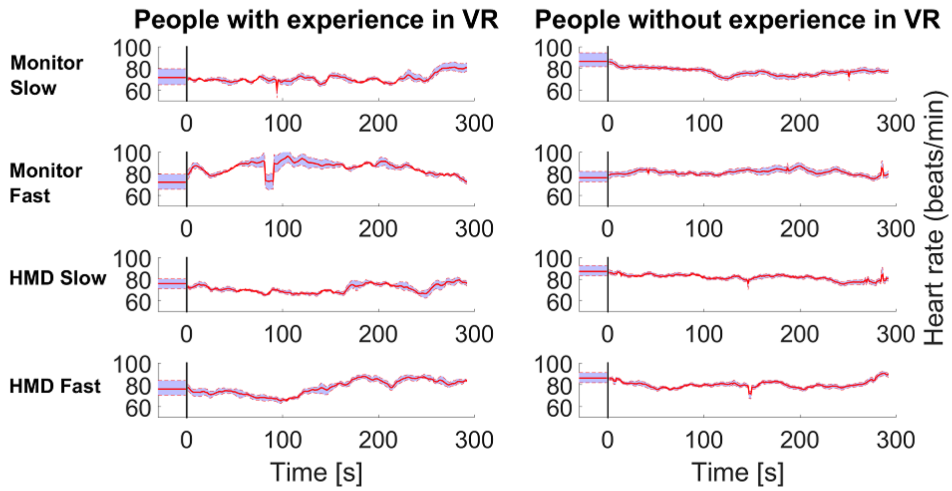


Figure 73: Industrial use case: average heart rate of participants based on the trial with (left) and without (right) previous experience in VR.

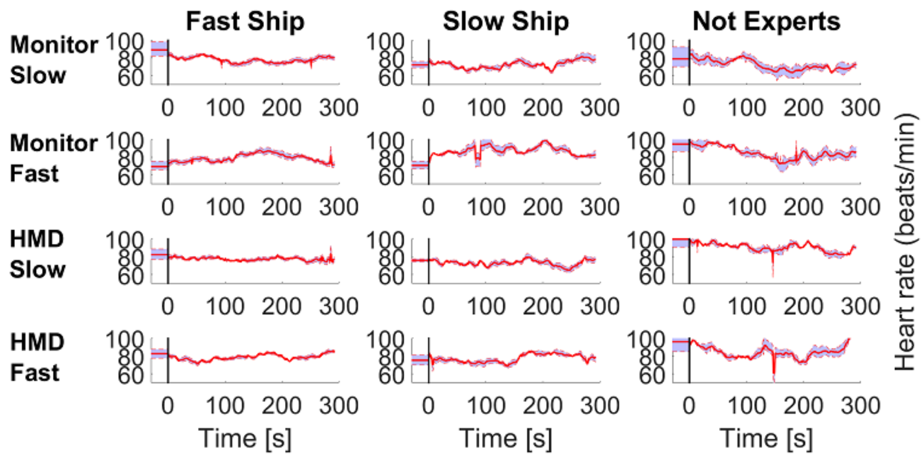


Figure 74: Industrial use case: average heart rate of participants based on the trial considering their previous experience with different kind of ships.

Apart from physiological measurements, also SSQ responses were taken into account in order to evaluate *cybersickness*. Data were divide considering, firstly, the trial and, secondly, both the trial and the path.

	N	O	D
Pre-Post HMD S			0.0030
Pre-Post HMD F	0.0029	0.0119	0.0059
Post M S-Post M F	0.0111	0.0438	
Post M S-Post HMD S	0.0123	0.0032	0.0004
Post M S-Post HMD F	0.0006	0.0009	0.0026
Post M F-Post HMD S			0.0147
Post M F-Post HMD F		0.0475	0.0124

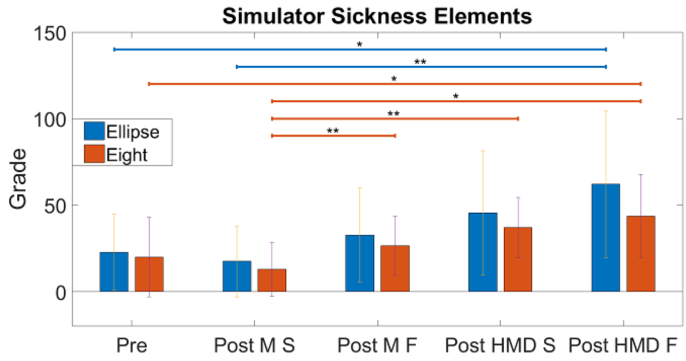
Table 10: Industrial use case: p-value obtained making a between group Wilcoxon test and comparing Nausea (N), Oculomotor (O) and Disorientation (D) grades in the SSQs.

The analysis on SSQ grades referred to the Ellipse and Eight paths (see Figure 75) shows an increase in the *cybersickness* between consecutive trials. *Cybersickness* is higher in the trial with elliptic path, probably because the sea was more rough. The worst sickness value, so, could depend on a combined influence of sea conditions, kind of hardware used (immersive or non-immersive) and ship velocity, as the use of the HMD alone does not explain the increase of malaise. A Wilcoxon test was performed in order to determine the statistical significance of these results. In particular, in the between groups analysis, comparing data referred to the two paths, the null hypothesis was never rejected, while the within group analysis revealed interesting correlations. In the Ellipse case, the differences between the total values of sickness in the Monitor Slow trial and in the HMD Fast trial are statistically significant ($p<0.02$), as the differences between the initial and final total grades ($p<0.05$). While in the Eight path, the results obtained in the first and final questionnaire ($p<0.05$) and in the trial Monitor Slow and the following tests are statistically significant($p<0.02$). Therefore, in the first case, factors majorly influencing sickness seem to be the boat velocity and the hardware used for simulation, with the HMD negatively affecting participants well-being. Whereas, in the second case, the velocity of the boat plays a fundamental role: curved and irregular trajectories and sudden direction changes, notwithstanding, cause malaise more easily than regular linear path.

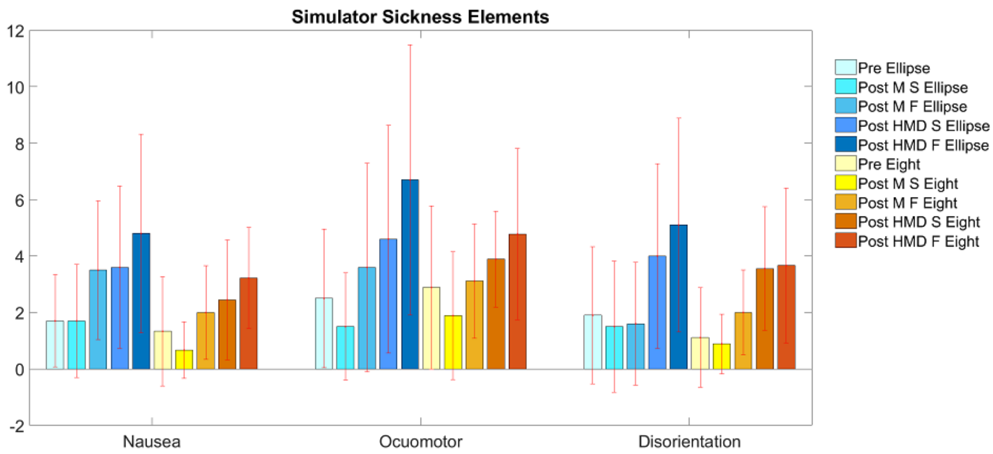
These considerations are confirmed by the evaluation of the three major symptoms of cybersickness. In general, for both path conditions, they tend to increase during the experimental session, in particular Oculomotor grades, which is consistent with results found in the literature.

3.3.4 *Outcome*

The aim of this work is the application of new technologies for immersive VR to ship handling simulation as a substitute of standard non-immersive systems. The simulator system is composed by a software component, the VR



(a)



(b)

Figure 75: Industrial use case: results of SSQs organized considering the trial and the path. (a) Total grade. (b) Results are divided in the three subcategory of the SSQ (Nausea, Oculomotor and Disorientation). M = monitor, HDM = head-mounted display, S = slow, F = fast.

framework, and a hardware part, the visualization modality and the physical reproduction of a ship command panel, with rudder and knobs, used for interaction.

As a preliminary study, expert naval users were asked to perform a manoeuvring task. Three conditions were considered, the kind of visualisation setup (immersive, HMD, and non-immersive, Monitor), the path (Ellipse and Eight) and the boat type (Slow and Fast), and both quantitative (head rotations, trajectories and physiological measurements) and qualitative (SSQ and IPQ) parameters were recorded and analysed.

In order to assess the usability of immersive VR simulation systems for nautical personnel training purposes, we have to demonstrate that they provide a realistic driving experience both in terms of handling a ship, following a predefined path and *sense of presence*, do not introduce any form of severe *cybersickness*, even after a prolonged use. Indeed, the analysis on the raw trajectories, even if coarse, clearly shows that expert drivers are able to follow the predefined path in a quite accurate manner in all conditions. *Sense of pres-*

ence is high in both visualization modality but better for the immersive setup. Moreover, it is worth noting that one of the main feature of HMDs, is that user is surrounded by a 360 degrees VE, that he can freely and naturally explores by turning his head. This can be very useful in the case of complex paths, as the analysis of yaw rotation angles in the case of the Eight path has demonstrated. Finally, considering *cybersickness*, skin conductance and heart rate remain constant and stable in the four trials, indicating that both setups do not introduce anxiety, stress or particular emotional or malaise states that could compromise performance and, eventually, learning of new skills. The only exception are the Fast tests, where skin conductance and heart rate tend to be slightly higher, maybe because of the perceived difficulty of the task. From the analysis of answers given to the SSQ, instead, it has been observed a mild increase of *cybersickness*, especially in the immersive setup. However, as the execution order was fixed, first Monitor then HMD, it can not be stated if this increase was due to the setup used, or to a physiological fatigue as the experimental session proceeds.

Further acquisitions, on non expert drivers or senior drivers, are required in order to better understand the usability of the systems on a large scale, especially considering motion sickness and seasickness, and their actual usefulness in providing long term learning of new skills.

3.4 Cognitive Research Use Case

One of the greatest challenges in Affective Computing (AC) is the artificial elicitation of emotions. Ideally, real-life setting should be used; unfortunately, reproducing ideal conditions is time consuming, as it is not possible to reliably predict the amount of time needed to gather specific emotional states, and very challenging from a technical point of view, due to the need of using multiple sensors and to maintain the synchronization across different data streams. Previous works proposed different solutions ranging from passive methods, based on the presentation of pre-validated sets of pictures [49] or videos [13], to active methods, exploiting, for example, interactive scenarios in lab conditions [119]. Passive methods usually result in low intensity reactions, while interactive scenarios imply difficulties in controlling the experimental procedure and the type of elicited emotions [86]. Immersive VR, thus, could represent a valid alternative, as it allows a better control and manipulation of the emotion-eliciting stimuli, ensures the replicability of the experimental conditions among participants and produce strong emotional reactions thanks to the interactive component.

The collaboration with Casa Paganini-InfoMus Research Center aimed at creating an ecological multimodal dataset for detection and recognition of emotional states. The project was composed by two main parts: (i) the design and implementation of a VR game eliciting specific emotions and levering a well-established psychological theory, the Roseman's appraisal theory; the creation of a software platform collecting and synchronizing the multimodal data, physiological (heart rate, muscle contraction), kinematic (acceleration), visual (video of the user and of the VE during interaction) and audio (user's respiration). My role consisted in the modification of the Ironman game, described in 3.2.1.2, in order to follow Roseman's theory requirements for emotion elicitation of joy and frustration.

3.4.1 *Material and Methods*

3.4.1.1 *The appraisal theory based VR Game*

In affective sciences, emotion elicitation refers to the use of emotional stimuli to evoke affective responses. Research in this field often relies on appraisal theories, focusing on the cognitive processes preceding the elicitation of an emotion (appraisal) and considering emotion as a process, rather then a state. According to [94], appraisal "mediates between the stimulus and the emotional response, and it is immediate and often unconscious". Thus, appraisal precedes cognitive labeling, simultaneously stimulating both physiological arousal and the emotional experience, triggering and differentiating emotional episodes, determining the intensity and the quality of actions, physiological responses, behaviors and feelings [94, 144].

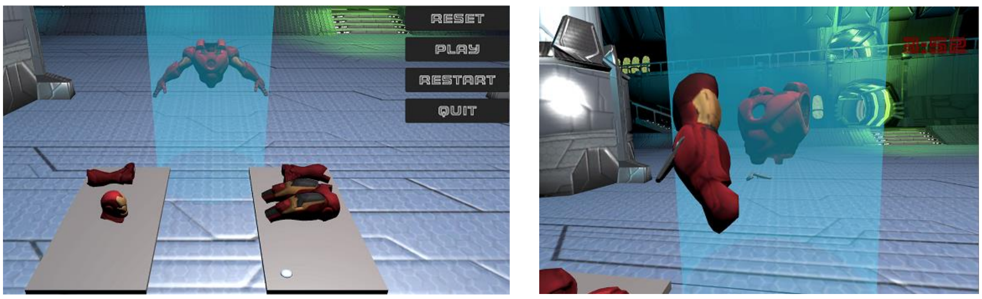
Roseman's theory [139] is part of the appraisal theories. The author conceptualizes the emotions as an organized system of coping responses to situational circumstances, which are based on five appraisal components to determine the occurrence of 14 discrete emotions. The considered components are:

- situational state: i.e. the consistency of the event with someone's motives;
- probability: the certainty of uncertainty of the outcome of the event;
- agency: whether an event is self-caused, other-caused, circumstance-caused;
- motivational state: consistency of the event with the motive of obtaining reward (appetitive) or avoiding punishment (aversive);
- power: referring to a person's control (strong or weak) over the situation.

For instance, the event of receiving a prize would elicit pride, as it is consistent with one's motives of being rewarded, certain, and appraised as something depending on the person's ability.

The Ironman game, described in Section 3.2.1.2, was chosen as a basis for the experiment because it exploits relatively simple game controls and avoids cognitively demanding tasks and elaborate narratives [108]. Moreover, its futuristic environment, similar to the one used in many commercial games, and its soundtrack allow the player to feel engaged and involved in the VE and distract him from the data collection process going on during the playing session. Considering results shown in 3.2.3.4 we preferred using controllers for interaction.

The game, however, had to be adapted to recreate the situational circum-



(a) (b)

Figure 76: Two views of the modified Ironman game: (a) the desktop application for the experimenter and (b) the VR view of the player.

stances that, according to Roseman's appraisal theory, would elicit a positive and a negative emotion, respectively joy and frustration. All modifications were discussed and implemented in collaboration with Casa Paganini's psychologists. Frustration and joy are both circumstance-caused, i.e. they depend on an external factor. This has been translated in the imposition of a time constraint. In the frustration case, the constraint would prevent the player from finishing to assembly the Ironman suit, even if he had the capacity to finish

Emotion	Appraisal variables in theory	Appraisal variables in VR
Frustration	Circumstance-caused, strongly uncontrollable events inconsistent with personal appetitive motives	The task is simple but time constraints make it impossible for the players to win the game
Joy	Circumstance-caused, strongly uncontrollable events inconsistent with personal appetitive motives	Having enough time to complete the task, players satisfy their desire to win the game

Table 11: Cognitive research use case: appraisal variables related to frustration and joy, according to Roseman's theory, and the corresponding features in the VR game.

the task; whereas in the joy case, time would be sufficient to satisfy his desire to accomplish the task (see Table 11). It is worth noting that, while pride is self-caused, joy is due to external circumstances, this is why time has a strong importance in success achievement.

The Ironman game was, thus, modified adding a feature: the option of choosing between two playing conditions, normal or manipulated. In the normal condition, the user has a fixed time to assembly the suit, 2, 3 or 4 minutes according to the experimenter choice. These options were defined based on the results in Section 3.2.3.4, which highlight that the mean time for task completion is around 3 and 4 minutes, for adults and teenagers. In the manipulated condition, whenever the player is very close to accomplish the task (when 11 of the 13 suit parts have been correctly positioned), the duration of the game is suddenly shorten, a countdown appears and a voice announces that only 10 seconds are left, making it nearly impossible to successfully complete the game.

The general structure of the Ironman game, its dual nature (VR and desktop application) and the VR scenario were maintaied but as shown in Figure 77, some additional blocks have been added to meet the needs of our collaborators. In particular, from the player perspective, after the introductory scene (Intro), with the game instructions, a demo scene has been added, to familiarize with the game interaction modality and Ironman exoskeleton structure. Here, some pieces of the suit are pre-assembled and player as to complete the puzzle. From the supervisor point of view, instead, the settings scene was modified in order to allow him to set the trial duration (2, 3 or 4 minutes) and the desktop view has been simplified (Figure 76 b), leaving only the reset, play/pause, restart and quit option. Moreover, the experiment have full control over the gameflow, he can chose the game mode, "Manipulated Time" or "Normal Time", after the demo or after each trial execution, or quit the game at any time, in case the player feels sick.

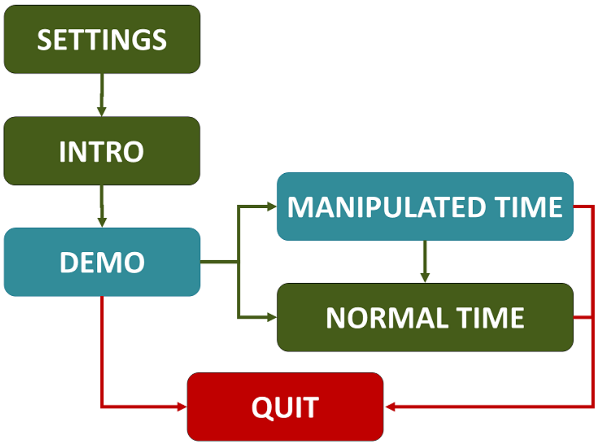


Figure 77: Flowchart of the modified Ironman game. Green blocks were already part of the game described in Section 3.2.1.2, while blue one have been added in the new version.

3.4.1.2 *The multimodal recording system*

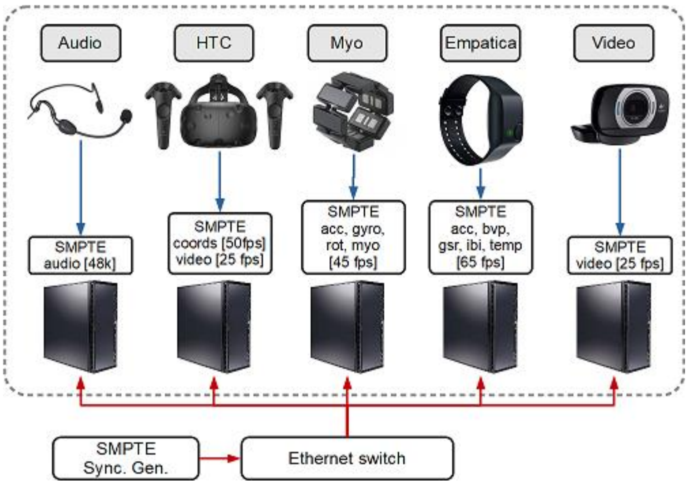


Figure 78: Cognitive research use case: multimodal recording system.

Casa Paganini's collaborators implemented a multimodal recording system by exploiting the EyesWeb XMI open research platform, a modular application for the design and development of real-time multimodal systems [169]. As reported in Figure 78, several machines are connected through a wired network, on which a SMPTE timecode signal is constantly transmitted via UDP packets and acts as a synchronization clock between the machines, each having it's own internal clock. The wired network is a local gigabit Ethernet connection which ensures a high speed transmission of the UDP packets containing the SMPTE timecodes. Each machine connected to the system acts as an independent recorder and manages the communication with one of the sensors. Furthermore, as it is distributed over a network of wire-connected workstations, more modules can be added without introducing latency. In the current setup, five devices were considered: a head-mounted wireless microphone to register the audio of respiration; the HTC Vive, providing its own tracking system; the MYO device[6], capturing forearms electromyographic signal; Empatica E4 bracelet[7], used to collect photoplethysmographic signals for blood volume pressure evaluation, as an indicator of heart rate, electrodermal activity and skin temperature; a webcam recording user while playing. Also, a second video stream is recorded in EyesWeb XMI by capturing a portion of the screen of the machine running SteamVR.

3.4.2 *Preliminary Experiments*

A preliminary data collection was carried out using the developed framework and the outcome constitute the EmoVR multimodal corpus. The aim was understanding if the modified Ironman game could be used for the purpose of eliciting specific emotions, asking people to accomplish the assembly task both in the Manipulated and Normal Time mode. For this reason, along with our game, other commercial games available on SteamVR and known in the literature to elicit specific emotions, were used. Half of them focused on fear of height, the other half on awe [36, 37].

In total 5 tasks were selected, as shown in Table 12. Five participants took part in the data collection (3 females) and had to perform the 5 tasks in a fixed order. Interactive applications were interleaved with video contents requiring a passive participation only, in order to alternate tasks with high arousal, with less agitating sessions.

After each data collection stage, participants were asked to self-report their affective state from a list of 16 labels (see Table 12), including most of the emotions from Roseman's theory. Each participant could report more than one emotion per task.

3.4.3 *Data Analysis*

From participants self-reports (see Table 13), we can state that the proposed VR applications were able to elicit a large spectrum of positive and negative emotions confirming the previous literature. In particular, frustration and joy were successfully elicited in 3 out of 5 participants. Other reported emotions were pride and surprise; whereas only one participant did not report any emotion for our game. It is important to notice that the same event can subjectively result in different emotions being elicited. Therefore, although used stimuli were carefully designed, the elicited emotions may differ. According to Roseman theory, however, joy is differentiated from pride by the agency component: the first is circumstance-caused while the second is a self-caused events. Thus, this

6 https://support.getmyo.com/hc/en-us
7 https://www.empatica.com/en-eu/research/e4/

Name	ID	Description	Emotions
Kitty Rescue game [8]	T1	Rescue the kitten lost in a sky-scraper in construction (i.e., virtual height exposure)	fear, joy, satisfaction
Set of videos from [36]	T2a	Combination of YouTube clips	amusement
	T2b	Video of hens wandering	neutral
	T2c	Drone video of mountains	awe
	T2d	Video sequence of tall trees	awe
Ironman Game	**T3**	**See Section 3.4.1.1**	**frustration joy**
Shinrin-yoku [9]	T4	Relax in a virtual forest	awe
RideOp [10]	T5	VR luna park	fear

Table 12: Cognitive research use case: list of tasks used for data collection.

Task	T1	T2				T3	T4	T5
		a	b	c	d			
awe/delight	0	0	0	1	1	**0**	3	3
surprise	1	1	0	0	2	**1**	1	0
hope	0	0	0	1	0	**0**	0	0
joy	1	3	0	1	2	**1**	2	1
relief	2	0	0	1	0	**0**	2	1
fear	5	0	1	0	1	**0**	0	4
frustration	2	0	0	0	0	**3**	0	0
anger	0	1	1	0	0	**0**	0	0
pride	0	0	0	0	0	**1**	0	0
guilt	0	0	0	0	0	**0**	0	0
regret	0	0	0	0	0	**0**	0	0
sadness	1	0	1	0	0	**0**	0	0
distress	1	0	0	0	0	**0**	0	0
no emotion	0	1	3	1	1	**1**	0	0
other emotion	0	0	1	0	0	**0**	0	0

Table 13: Cognitive research use case: self-reported emotions for each task. Numbers represent the amount of people who experienced the corresponding emotion in each task.

model gives the plausible explanation of the answer "pride", as it is reasonable to assign the agency to oneself when winning a game.

3.4.4 *Outcome*

Two were the main contributions of this work: the design and implementation of the first immersive VR game inducing frustration and joy, based on the Roseman's appraisal theory, and the creation of a software platform for the collection of synchronized multimodal data, exploiting a novel combination of modalities, i.e., physiological data, kinematic data, video recordings and audio.

Preliminary experimental results confirm that it is possible to successfully induce a spectrum of positive and negative emotions in VR scenarios, suggesting the feasibility of using VR-based methods to collect affect-related data. In particular, our game succeded in eliciting frustration in 3 subjects on 5 and joy/pride in 2 subjects on 5. This encourage us in the creation of new VR ad hoc implemented contents, for the elicitation of other emotional states. Taking into account, however, that elicitation of certain emotions could be trivial (joy/pride).

Finally some limitations in the use of a simplified self-reporting questionnaire also emerged. For this reason, in the future, more sophisticated validated tools will be exploited, e.g., GRID [59], asking players to evaluate their experience in terms of appraisal variables, instead of emotional labels.

3.5 Cognitive Assessment of Elderly Use Case

Nowadays, the increase of the population average age determines a greater incidence of neurodegenerative diseases related to aging [64, 129], like dementia. Actually, dementia is not a specific disease but a descriptive term for a set of symptoms that can be caused by various disorders affecting the brain. One of the most common and significant symptoms of dementia is the loss of memory, usually accompanied by difficulty in: (i) following a conversation or finding the right words to express a concept; (ii) decision making or problem solving; (iii) carrying out complex actions (for example cooking a meal); (iv) assessing distances and orientations. Dementia is commonly preceded by a pre-dementia stage, named Mild Cognitive Impairment (MCI), where elderly people have a cognitive decline greater than the one expected for their age. Such decline might not interfere notably with Activities of Daily Living (ADL).

Until a few decades ago, dementia was considered a severe condition, while today thanks to researches in the neuropsychological field, we have a better comprehension of this pathology, of its different forms and possible therapeutic approach, thus rising a great interest for non-pharmacological treatments and screening tools for early diagnosis. Currently, these tools primarily include classical paper-pencil tasks for assessing the cognitive decay, such as the Mini-Mental State Examination (MMSE) [58], the Short Portable Mental Status Questionnaire (SPMSQ) or Pfeiffer test [124] and General Practitioner Cognitive Assessment of Cognition (GPCog) [27]. These tests investigate different cognitive areas (orientation, registration, attention and calculation, recall, language and praxis), through a series of questions. They are short and take some minutes to be executed. Nevertheless, they have some limitations: they present a low level of sensitivity for mild degrees of impairment and a dependence on cultural background.

In the last few years, researchers have tried to devise alternative methods, able to overcome these limitations but many of these approaches are still in the form of proofs-of-concept. Lots of the new approaches focus on the simulation of ADLs to evaluate the behaviour of patients in different daily situations. Non-immersive and immersive VR platforms, thus, play an important role, as they allow to reproduce complex environmental and social situations that can stimulate the subject in a similar way to the corresponding context, to modulate the intensity and duration of the experience according to the needs of patients and to evaluate psychopathological reactions and cognitive functions more immediately than the classic paper-pencil tests. In this way, screening instruments with a greater ecological validity can be obtained.

By considering non-immersive VR systems, in the literature we can find many different applications for assessment and training: a music video game with the aim of increasing positive emotions [16]; a tablet-based cooking game where the patient plans the steps to do a recipe [99]; a virtual loft where daily activities are simulated and each task can evaluate executive functions (reasoning and planning), attention (selective and divisive), memory (short-term and long-term) or visuospatial orientation [189]; ECO-VR, a virtual apartment where the patient can accomplish different tasks involving two or more cognitive functions [121]; RehabCity, simulating a district where four buildings can be visited (a supermarket, a post office, a bank and a pharmacy) and by accessing each individual building, it is possible to perform tasks of different complexity [170]; a virtual supermarket for tablet, where the main objective is to buy the products in the shopping list and pay the due amount [190].

By moving towards immersive VR, [123] created a VR application for neurocognitive assessment simulating the daily activity of shopping in a supermarket. When subjects wear the headset, they are immersed in a virtual supermarket, they can explore the scenario using a joystick and interact with objects thanks to a special tracked glove. A virtual shop was also proposed to measure episodic memory for older adults [44]. Results show that memory performance in the VR task are positively correlated with performance on a traditional memory task.

In the current use case, a serious game similar to the one proposed by [190] has been implemented on two platforms, PC and tablet. The aim is developing an application for the assessment of different intellectual functions substituting traditional tests, like the MMSE, and evaluating the ability to concentrate and pay attention, the reasoning, the visual perception and orientations. The two solutions were proposed to a group of young people, of healthy elderly and of elderly affected by mild cognitive impairment (MCI). The aim was both evaluating the best interface to be used and validating the system by comparing our results with those obtained with standard approaches. Outcomes can be used as a starting point for the design VR-based serious games and exergames substituting standard questionnaire-based methods, thus improving both clinical evaluation of performance and patients' motivation.

3.5.1 *Material and Methods*

The serious game consists of a simple virtual supermarket with two shelves and a fruit counter, as shown in Figure 79. The scenario was created using Unity 3D. Items in the scene were made with SketchUp or downloaded from the free library 3D Warehouse [11]. Two versions of the game were developed: one runs on a 13,3-inch MacBook Pro with a 2,5 GHz Intel Core i5 processor and Intel HD Graphics 4000 1536MB graphics processor; the other runs on a 10,1-inch Samsung Tab 4 with a 1,2 GHz Quad-Core processor and Android 5.0.2 operating system.

Two sequential tasks are proposed: first, the participant has to add to his virtual cart all the 10 items listed in a shopping list, which is randomly generated at the beginning of the game (Figure 79 left); then he is asked to pay the exact total amount (Figure 79 right).

The interface is designed in order to be simple and intuitive. The user can visualize a shelf per time and cannot freely navigate the scene, he can use

11 https://3dwarehouse.sketchup.com/

Figure 79: A view of the Shopping Task (left) and Payment Task scene (right).

the arrows at the bottom of the screen to turn the camera and change the shelf. Only when he is in front of the desired shelf, he can interact with the objects on it. By clicking or tapping on an item, a pop-up window with the product name is displayed and the player is asked if he/she wants to add that article to the cart or not. Three buttons on the upper part of the screen give the user more options: the Cart button (top left corner) allows to see the list of bought products and eventually delete items; the Shopping List button (top right corner) shows or hides the shopping list; the Payment button (top center) enables the payment screen and disables the shopping scene.

3.5.2 *Experiments*

The game was tested first on 32 volunteer healthy subjects with an age range of 21–78 years (mean 39.8 ± 18.8 years), divided into Under60 (22 people in an actual age range of 21–49 and average age 28.1 ± 6.4) and Over60 (10 people in an age range of 60–78 and average age 65.7 ± 6.6 years); then on 30 elderly volunteers (ages between 60 and 94 years, mean age 74.6 ± 10.5), both healthy and affected by mild cognitive impairment. 13 of them were from two retirement homes for elderly, 10 were recruited by the project Mo.Di.Pro., in collaboration with the hospital Galliera in Genoa, and 7 were volunteers recruited by the authors.

In the first experimental session, a repeated measures counterbalanced experimental design was used: all participants were asked to try the application with both devices. Half of them started with the tablet and the other half with the computer. Each experiment consists of a training phase, with the demo scene, i.e. a simplified version of the main scene, a first trial with one of the device and the second trial with the other device. The second trial usually began immediately after the end of the first one. After each trial, participants were asked to remember and write all the bought items. In the end, they had to express and motivate their preference for the interaction device, PC or tablet.

In the second experimental session, participants were asked to play the virtual supermarket game only using the tablet, because touchscreen is an intuitive and natural interface and because in an assistive context (i.e. a hospital or a retirement home) the use of portable device may be more practical. During the experiment, subjects were first asked to provide some personal information, e.g., age and schooling, in an anonymous way, and to submit two paper-pencil questionnaires, the Pfeiffer test and the GPCog. Then, after a Training phase, the actual trial began.

Performance in both trials have been evaluated considering the execution times in the two tasks (shop and pay), the Shopping Score (SS), the Payment Score (PS) and the number of remembered items. The Shopping Score (Equation 6) takes into account the number of the bought items that are (CI) or are not (WI) in the given list and the number of items deleted from the cart (DI).

$$SS = CI - \alpha * WI - \beta * DI \quad (6)$$

Different weights (α and β) are associated to different errors.
The Payment Score (Equation 7), instead, is set to 0 if the payed amount is incorrect and 10 if it is correct, otherwise it is computed, taking into account the number of times the player has reset the total (E) multiplied for an empirical constant ($\gamma = 0.5$).

$$PS = 10 - \gamma * E \quad (7)$$

In both cases, a low score is related to an impairment in solving the task.

3.5.3 *Data Analysis*

In the first experiment, three different analysis were performed: based on the device used (tablet or PC); based on the order of execution; based on participants' age, taking into account both the device and order of trials. We also evaluated the statistical significance of the compared data groups by performing a two-sample t-test for all of the three cases, in order to assess if differences in results were actually linked to different conditions or not.

Comparing performance, considering both the two devices and the order of execution (Table 14) it can be noticed that there are no differences for the SS and PS, while execution times are slightly better with the PC and in the second trial, maybe because of learning. However, standard deviations are large indicating an internal variability of the data. Furthermore, the average number of remembered items does not considerably vary during the second trial despite subjects were aware they had to memorize the shopping list. Many subjects complained that the first time they had to write down the list, their bad results were due to the fact that they did not know they had to remember items in the list; while the second time they got confused and risked to write items from the first list. So results are not directly related to concentration but to subjective short term memory capability.

Significant differences were found between the performance of Over60 and Under60 groups regarding the mean execution time of both tasks (see Figure 80): as expected, Under60 group times are much lower ($p < 0.05$). Considering each group separately, we can notice that in the younger group, results with the two devices are comparable, while in the elderly group performance looks slightly (but not significantly) better in the case of the tablet. Also variability is lower in the case of Under60 group, maybe because more subjects were

	Tablet	PC	First Trial	Second Trial
Shopping Score/10	9.6 ± 0.9	9.6 ± 0.6	9.5 ± 1.0	9.8 ± 0.5
Payment Score/10	9.9 ± 0.2	9.9 ± 0.1	9.9 ± 0.2	9.9 ± 0.1
Shopping Time [s]	185.5 ± 114.7	191.4 ± 136.9	196.8 ± 132.1	180.0 ± 119.8
Payment Time [s]	18.8 ± 10.3	20.8 ± 16.1	20.8 ± 13.2	18.8 ± 13.8
# Remembered Items/10			8.6 ± 1.4	8.4 ± 1.5

Table 14: Cognitive assessment use case: mean and standard deviation of evaluated parameters considering the device used (tablet or PC) and the order of execution (First and Second Trial).

considered, or because they are more confident with technologies. Average Shopping and Payment Scores are particularly high for both groups, 9.2 ± 1.2 and 9.8 ± 0.4 for the Shopping Scores respectively of Over60 and Under60 group; 9.9 ± 0.3 and 9.9 ± 0.1 for the Payment Scores.

Considering the age group and the order of execution (Figure 81), average shopping time slightly decreases for Over60 group (from 347.6 ± 147.6 to 319.3 ± 120.) as well as for Under60 group (from 128.3 ± 26.2 to 116.7 ± 38.6), but the difference is not significant. The average number of remembered items in the first and second trial remains quite constant inside the two groups, but it is significantly higher ($p < 0.05$) in the Under60 group with respect to the Over60 group (7.8 ± 1.5 vs 7.3 ± 1.9 for Over60, 9.0 ± 1.1 vs 8.9 ± 0.9 for Under60).

Finally, in general, 50% of participants preferred using the PC because the screen was larger and the objects were more visible, the 30% of them preferred the tablet for the its intuitiveness and the last 20% enjoyed both devices. Preferences did not match the objective data.

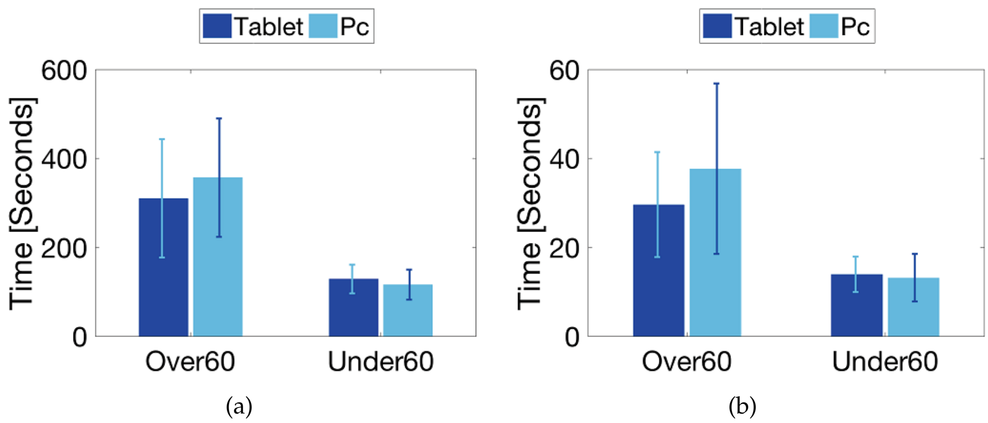


Figure 80: Cognitive assessment use case: mean and standard deviation of times required to complete the Shopping (a) and the Payment (b) Task with the two devices and for the two age groups.

In the second experiment, analysis are conducted considering the participants health state as the independent variable: results are divided in those referred to healthy subjects (Healthy) and the elderly affected by mild cognitive impairment (MCI). Considered parameters are, in general, better for the

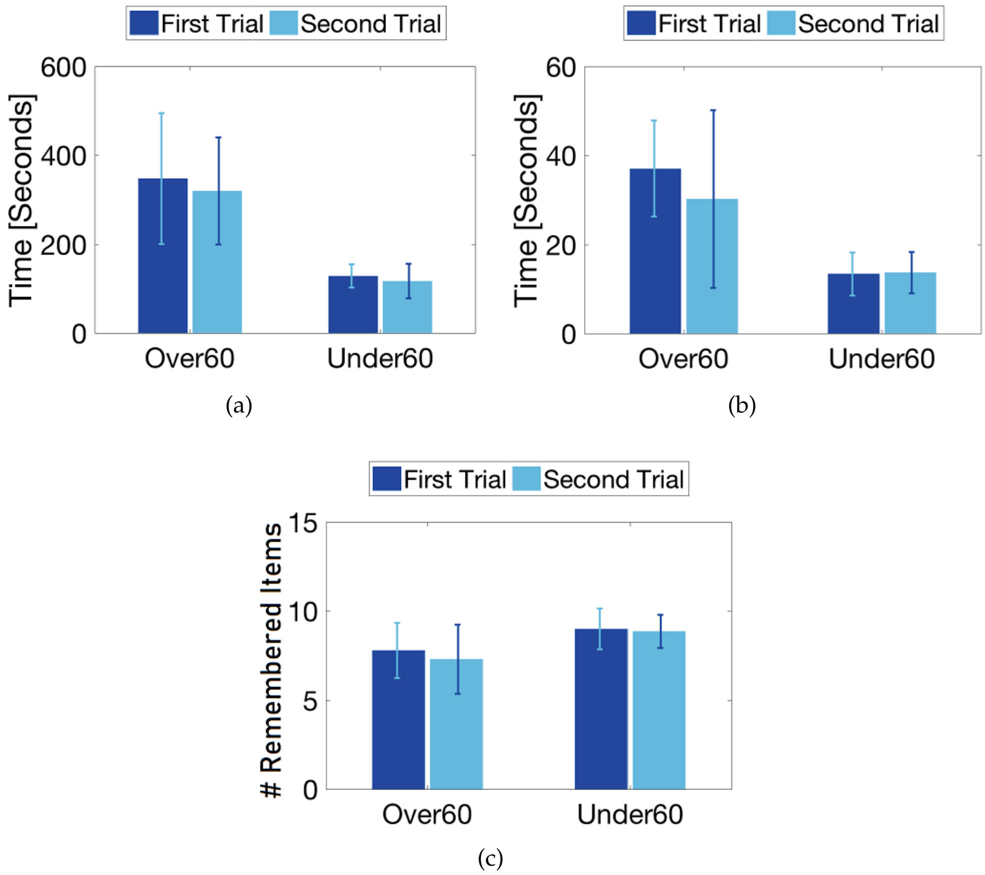


Figure 81: Cognitive assessment use case: mean and standard deviation of times required to complete the Shopping (a) and the Payment (b) Task and of the number of remembered items (c), considering the order of execution and the age group.

Healthy group, as shown in Figure 82: scores are higher, times are shorter and the number of remembered items is greater. These differences have been demonstrated to be statistically significant through a t-test statistical analysis ($p<0.05$). In particular, the lower score of MCI users is due to the fact that people bought less items, with respect the given list, maybe because of the fatigue or annoyance.

While all the Healthy users were able to complete the tasks easily, people affected by MCI found the task quite difficult. For example, they had difficulties in using euro currency or in understanding the paying mechanism and the interface, e.g. where to click in order to add money. Finally, Healthy people can remember an average of 8 items over 10, whereas MCI users have a mean of 3.60 ± 1.65 remembered items. This is especially true for those who obtained a low rating in the GPCog test, which assesses memory.

As the main goal of this research is the design of exergames for the assessment of elderly able to substitute traditional pencil-paper tests, results obtained in the supermarket game were compared to Pfeiffer and GPCog tests outcome and the Pearson correlation coefficient has been computed. A moder-

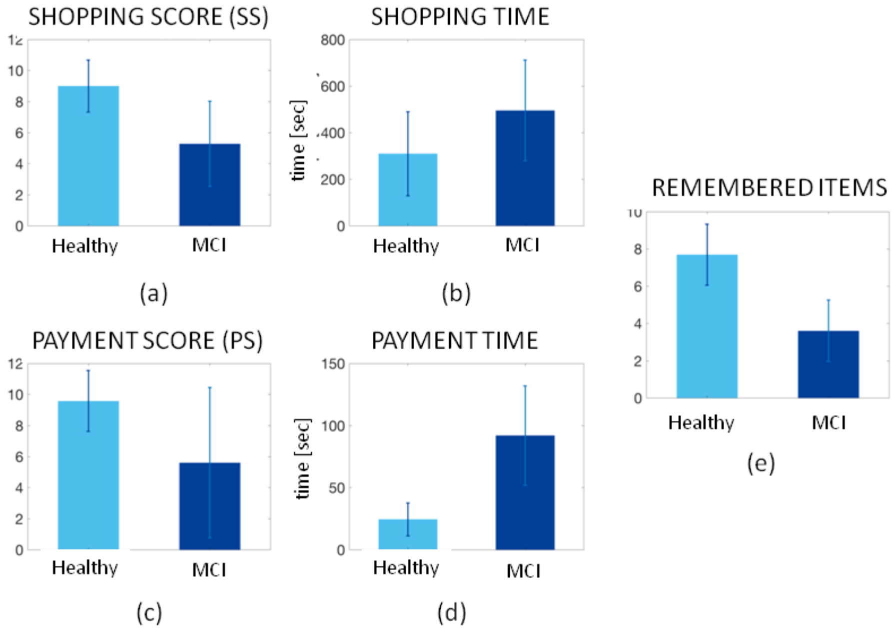


Figure 82: Cognitive assessment use case: mean and standard deviation of the Shopping Score (a), the Payment Score (c), the tasks completion times (b,d) and the number of remembered items (e) of Healthy and MCI subjects.

	Pfeiffer	GPCog
Shopping Score	-0.6	0.57*
Payment Score	0.04	0.62*
Shopping Time	0.29	0.59*
Payment Time	0.42	0.84**
# Remembered Items	0.16	0.78**

Table 15: Cognitive assessment use case: Pearson correlation coefficients among the parameters computed for the supermarket game and the Pfeiffer and GPCog tests. *$p<0.05$; **$p<0.001$.

ate (Shopping Score, Shopping Time and Payment time) and strong (Payment Time and number of Remembered Items) correlation have been found with the results of the GPCog test, suggesting a possible application of our non-immersive VR test, simulating an ADL, on the evaluate the cognitive status of elderly people. However, the absence of correlation with the Pfeiffer test indicates that further studies are necessary, in order to obtain a more complete evaluation.

3.5.4 *Outcome*

This work fits with the current research on the design of new methodologies for the assessment of the cognitive status of the elderly using serious games approaches based on ADL simulations. In particular, our focus is on non-immersive VR technologies. A serious game, inspired to a simple daily activity, i.e. doing the shopping, has been designed and two different versions have been implemented, for tablet and PC. A first series of experiments was conducted on Under60 and Over60 healthy volunteer subjects, in order to understand the best interface to be used both in terms of preferences and performance, evaluated by measuring the execution times of the proposed tasks (shop and pay), the Shopping Score, the Payment Score and the number of remembered items. As expected, young people performed significantly better than elderly, they were faster and received higher scores. Moreover, results referred to the Over60 group highlight an influence of learning over performance, comparing first and second trial, but no influence of the device used. For these reasons, we can state that both devices are potentially valid for the development of a serious game. However, considering preferences, PC still remains the favourite device for interaction, maybe because elderly have more confidence with PC and mouse than with tablet and touchscreen or because tablet screen is too small for them.

Starting from these results, in a second session of experiments the supermarket game with the tablet was tested on healthy and mild cognitive impaired subjects. Tablet was chosen because, in the previous experiment, performance have been proven to be independent from the device used and because in assistive context portable devices are more practical. MCI subjects showed difficulties in task completion and obtained results were compared with standard pencil-paper test for memory, the GPCog, and reasoning, the Pfeiffer, revealing a correlation only with the first one.

In future, the layout and the usability of the supermarket game will be improved and a third interaction modality will be added, namely immersive Virtual Reality. VR will allow obtaining a more natural and ecological experience. Nonetheless, based on gamification principles, different scenarios and tasks inspired to multiple daily activities could be implemented. It would be also interesting collaborating with doctors in order to acquire data from cognitive impaired and/or elderly people. The cooperation with the medical staff will provide us a better understanding of the problem of assessment of elderly cognitive status, so we will be able to add new features and create an ad hoc tool for doctors.

4

Conclusions

Having as the final goal the improvement of systems for simulation and training, through a better understanding of how humans perceive and process information and the design of a natural and intuitive interface, in this part the finding of the research work done in the field of perception and Natural Human Computer Interaction is summarized and discussed.

Virtual Reality is an innovating technology which, in the last decade, has had a widespread success thanks to the release of new low cost devices, such as the head-mounted displays. If, in the past, due to the expensive equipment required, VR was a niche product mainly used for military, aerospace or flight simulations, its recent diffusion has certainly contributed to the multiplication and diversification of its domains of application and has caught the attention of industries and researchers, particularly interested in the potential that this technology offers. Nowadays, VR is used in diverse areas, ranging from data visualization to serious games and edutainment, from applications in the medical field, for diagnosis, operation planning and minimal invasive surgery simulation, physical and cognitive rehabilitation, treatment of psychological problems, assessment and training of elderly people, to Industry 4.0, for modelling and products design, complex engines maintenance, assembly and prototyping process, teaching security for work places fatality reduction, training for risky situation managing. In particular, our interest is on simulation and training in the industrial and medical context.

Although, the literature on this topic is rich, the current work mainly focuses on the general mechanisms underling perception/action loop in VR. On the one hand, we want to understand how humans gather and process all the information presented, through the evaluation of the visual system bandwidth and of the Visual Working Memory capacity. On the other hand, since interface has to be a sort of transparent layer allowing trainees to accomplish the task without directing any cognitive effort on the interaction itself, we investigate the best interaction modality, specifically for selection and manipulation tasks, in terms of naturalness, intuitiveness, stability and efficiency. A better comprehension of the functioning of the perception/action loop in Virtual Reality is crucial in order to improve the design and implementation of applications for training and simulation in immersive VR, through the definition of guidelines delineating the best way to present information in VR, in order to take advantage of the immersivity offered by headsets, and to interact with them.

To this aim we have developed ad hoc frameworks and methodologies. The software frameworks consist in the creation of VR scenarios, where the experimenter can eventually choose the modality of interaction and set some features of the scene, guaranteeing experiments repeatability and controlled conditions. The methodology includes the evaluation of performance and User Experience, considering both quantitative and qualitative metrics derived from the collection and the analysis of heterogeneous data, as physiological and inertial sensors measurements, timing and self-assessment questionnaires.

Considering the part of our research regarding the evaluation of the VWM capacity, in the literature it is possible to find several studies about this topic. The majority of them, however, utilize 2D images or videos displayed on standard desktop screen monitors, whereas a systematic approach exploiting VR technologies does not exist yet, at least to our knowledge. For this reason, we first decided to adapt a well-established method, the *whole-display* change blindness *one-shot* paradigm, to be used in Virtual Reality. We considered the

influence of five different conditions on the visual system bandwidth, i.e. the kind of change (appearance, disappearance and no change), the number of objects to be remembered (4, 6, 8, 10, 12), the spatial layout (vertical and horizontal), the occupied horizontal Field of View (40°, 80°, 120°) and the observation time (0.3 and 0.9 s), and we devised two experiments, namely the *change detection* and *change localization* test. We used the HTC Vive headset for visualization and the HTC Vive controllers for interaction. In the *change detection* experiment, each controller was marked with a virtual label, "Yes" or "No", and participants had to press the trigger of the controller corresponding to the right answer; while in the *change localization* test, people used only one controller to raycast the position in which they though a change occurred.
Concerning FOV 40° and 80°, where stimuli are inside the HMD's FOV, results confirm findings of the previous literature, i.e. people are able to recognize a change with 0.75 accuracy when up to 8 elements are in the scene. Data acquired on FOV 40° allow us to create a baseline for the analysis of results obtained with larger FOVs and fill the gap between non immersive 2D and immersive 3D experiments. Same results are obtained also in the FOV 120° when subjects are given enough time, 0.9 s, to turn their head and build a gist of the entire scene. In the trial with FOV 120° and short observation time, instead, the upper limit is 6 items. The kind of change did not affect results, maybe because of an absence of pictorial cues, such as occlusions or shades. Also layout seems not influencing performance, even if people often complained that the trials with vertical wall were uncomfortable and more fatiguing, maybe because the semi-cylindric wall had a ray of 70 cm and was perceived as being too close.

Considering our research on Natural Human Computer Interaction, instead, multiple solutions are currently available, however as they are task dependent, we first had to define the desired task. So, taking into consideration that our final application domain is simulation and training in the industrial and medical field, we decided to focus on selection and manipulation of objects and we chose two state of the art solutions for those tasks, a touchful one, i.e. the HTC Vive controllers, and a touchless vision-based one, i.e. Leap Motion. The first ensure more stability, reliability and effectiveness, thanks to their inertial sensors and external tracking; whereas the second allows interacting with virtual objects using natural and intuitive hand poses and gestures, thus improving *sense of presence*, *ownership* and *agency*. Nonetheless, in vision-based solutions, hands detection and recognition problems are still common issues. For this reason, we implemented our own grab function, based on the definition of three hand-object states (no collision, collision and grab action) to which we associated three different behaviours and colours of the object.

Controllers and Leap Motion-based interfaces were tested and compared in terms of performance and preferences. To this purpose, we designed and implemented two different scenarios, the Shape Sorter, where participants have to grab 12 objects, different in colour and shape, and put them in the corresponding hole, and the Ironman game, where users are asked to assemble the movie character suit. The two tests differ for the width of the interaction space and for the level of fineness of manipulation involved, e.g. in the second case, larger movements are required and the target position has to be reached with higher accuracy.

We, furthermore, devised five different experiments. In Experiment 1, we used the Oculus Rift DK2, suitable for sitting and standing application, for visualization and the Leap Motion, with our grab function, for interaction. We tested this interface in the Shape Sorter task, considering two different game areas, the peripersonal and near action space. Results suggest that, in order to ensure a stable interaction in the near-action space, another tracking system is required. For this reason, since then, experiments have been conducted with the HTC Vive, allowing for standing applications and room-scale setups. In Experiment 2, we compared the HTC Vive controllers and Leap Motion, first with the standard algorithm found in the Unity Asset Core then with our grab function, in the Shape Sorter task. From the analysis of recorded data controversial outcomes emerge: even if the touchful interface ensures a more stable and effective interaction, evidenced by better performance, people like the Leap Motion and appreciate the idea of interacting with the VE using their own hands. Moreover, the algorithm we implemented contributed in improving performance with Leap Motion. We, thus, decided to repeat the previous experiment using only the Leap Motion with our grab function and testing the influence of different feedback. In fact, the lack of haptic feedback, which guides actions during interaction with the real world, is one of the main problem related to the use of touchless interface. We defined four conditions (None, Visual1, Visual2 and Audio) and tested them in the Shape Sorter task (Experiment 3) and in the Ironman game (Experiment 4). Visual and audio feedback modalities have been implemented as a substitute of the haptic one, which would require specialized hardware. Finally, in order to have a baseline for the analysis of results obtained in Experiment 4, we devised an in-the-wild experiment (Experiment 5), asking naive participants to accomplish the complex task only with HTC Vive controllers.

HTC Vive controllers have in general better results, however, our grab algorithm improves both Leap Motion interaction, ensuring more comparable performance, and preferences, which increase from 13% to 31%. Different feedback modalities obtain similar results, indicating that the improvements brought by our algorithm depend more on the restrictive control on objects behaviour and on the distinction of hand-object states, than on the visual feedback provided in the first implementation. Considering preferences, in the Shape Sorter scenario there is not an evident preference between Audio (38%) and Visual 1 and 2 (32%), while in the Ironman scenario visual feedback modalities received 50% of preferences and Audio only 11%. This means that, when Audio feedback during interaction is the only sound source in the scene, it is preferable using this non-visual channel to convey information, in order not to overload the visual one; whereas, when multiple sound source are present in the scene, Audio is quite ineffective.

Finally, three use cases were considered: an industrial use case, a cognitive research use case and a cognitive assessment application use case.

The industrial use case, in collaboration with CETENA S.p.A., consisted in the assessment of the usability of an immersive Virtual Reality ship handling simulator system as a valid substitute of non-immersive systems for training of nautical personnel. To this purpose, expert naval users were asked to perform a manoeuvring task and three conditions were manipulated: the kind of visualisation setup (immersive with HMD and non-immersive with Monitor), the path (Ellipse and Eight) and the boat type (Slow and Fast).

Results confirm that immersive VR systems can substitute non immersive one, as they allow expert drivers to follow the predefined path in a quite accurate manner, feel a higher *sense of presence*, taking advantage of the possibility to rotate their head to explore the 360 degrees VE surrounding them, and do not cause severe considering *cybersickness*.

The cognitive research use case, in collaboration with Casa Paganini-InfoMus Research Center, consisted in the implementation of an immersive VR game inducing frustration and joy, based on a well-established psychological theory, the Roseman's appraisal theory. This will be used in future, together with a software platform for the collection of synchronized multimodal data, to create an ecological multimodal dataset for detection and recognition of specific affective states, eliciting emotions in immersive VE.

During a preliminary experimental session, five volunteers were asked to play our game together with other games and applications, known in the literature to elicit specific emotions. Results confirm that it is possible to successfully induce a spectrum of positive and negative emotions in VR scenarios, suggesting the feasibility of VR-based methods to collect affect-related data. In particular, our game succeeded in eliciting frustration in 3 subjects on 5 and joy/pride in 2 subjects on 5.

The cognitive assessment application use case, consisted in the development of a system for the cognitive decay assessment of elderly, in particular dementia, able to substitute standard paper-pencil tests, in the context of a project my research team is working on. Classical assessment methods, in fact, have some limitations: they present a low level of sensitivity for mild degrees of impairment, depend on the patient cultural background and, as they are perceived as an exam, they can cause stress and frustration during their filling. Instead, new technologies, in particular immersive and non-immersive VR, allow to simulate daily life activities (ADL), which are usually impaired by dementia, in the form of a game, putting the patient more at ease.

A serious game recreating the situation of doing the shopping, has been designed and implemented on two different platform, tablet and PC. A first series of experiments was conducted on Over60 healthy volunteer subjects, in order to understand the best interface to be used both in terms of preferences and performance. An Under60 group, constituting the control group and the baseline for the analysis of results, was tested too. In the Over60 group, device used had no influence on performance even if PC still remains the favourite device for interaction, maybe because elderly have more confidence with desktop screen and mouse than with tablet and touchscreen or because tablet screen was too small for them. However, tablets are more practical in assistive context for their portability. In a second session of experiments, the supermarket game was tested on healthy and mild cognitive impaired subjects. MCI subjects show difficulties in task completion and obtained results are compared with standard pencil-paper tests for memory, the GPCog, and reasoning, the Pfeiffer, revealing a correlation with the first one.

To sum up, considering the final goal of our research, i.e. the design and implementation of applications for training and simulation, in context such as Industry 4.0 and the medical field, VR has been found to be a valid tool able to simulate desired situations in a realistic and involving way, eliciting user's *sense of presence*, without causing severe *cybersickness*. Moreover, VR simulations have many advantages with respect to real-life one: they allow to immerse the user in dangerous and potentially fatal scenarios or unethical situations while guaranteeing the immunity of the trainee himself and of third parties; they ensure active learning in a multi-sensory artificial environment; being simulations parameters settable and scenarios controllable, they ensure repeatability and the monitoring of people state and improvements; they allow to trigger or avoid triggering specific emotions, in order to voluntarily create potentially stressful or relaxing situations.

Considering the ability of trainees to perceive and process information, comparable cognitive mechanisms are activated in VR and real-life settings. A simulation should optimize the use of human Visual Working Memory, taking into account that it is limited to around 6/8 items. When people are given enough time to build the gist of a scene, they are able to retain in memory up to 8 elements, even when no more visible. As human horizontal FOV is $220°$ while VR headset FOV is $90°$, this means that information and instructions can be easily distributed around the user, in order not to overload the current view of the scene hindering interaction with the VE or task accomplishment.

Finally, considering interaction itself, when selection and manipulation tasks are simple and do not require fine movements, controllers and Leap Motion ensure comparable performance; whereas when tasks are complex, the touchful solution turns out to be more stable and efficient, also because the touchless interface requires a higher control over the objects behaviour, reducing scene naturalism. Nonetheless, the idea of interacting with the virtual environment using natural poses and gestures is usually attractive, so new strategies to improve bare hands interaction are required.

5

Perspective

The results obtained lay the foundations for further researches both on the way humans perceive information coming from the surrounding environment and on the interaction with the environment, which are examined in this last part.

Virtual Reality is a constantly evolving technology, new headsets, with innovative features, better resolution, larger FOVs, and VR related devices, such as tools trying to make interaction with virtual objects more and more natural and efficient, keep being released every year and researchers and entrepreneurs from different fields keep being attracted by the potentiality that this technology offers. It is therefore reasonable to think that VR, in the next future, will continuously improve and will more and more emerge as a system able to replace real-life experience. To this aim, the perception/action loop will have to be investigated and boosted. In this perspective, the current work represents a first step in that direction, however, further researches are required. Results obtained by our research offer several interesting perspectives for the theoretical study of the perception/action loop but also for more practical applications.

Considering the work on VWM capacity, for example, we now have a systematic methodology, a software framework and a baseline for the interpretation of results obtained with stimuli displayed in VR environments. Indeed, it will be possible continuing to gradually enlarge the Field of View, up to 360 degrees, while proportionally increasing the observation time, and evaluate if the FOV/time relation will be still valid or if other factors will intervene to degrade performances. Moreover, objects features and the semantic of the scene will be modelled in order to create and test the influence of more or less complex scene over performances. By the analysis of acquired data it will also be possible to define a model of WM functioning in Virtual Reality. This will have an impact both on an applicative and theoretical point of view. On the one hand, in fact, it will allow understanding how to arrange information and instructions in the 360 degrees VE, so that it will be easier for trainees to perceive and process them, designing and testing also different vertical layouts, in order to make them less fatiguing and uncomfortable than in the experiments presented in the current research. On the other hand, it could be used to help understanding how Working Memory works in real-life scenarios. It is worth noting, in fact, that, although VR allows reproducing Change Blindness experiments which would be impossible to recreate in the real world, the strong limitation of this technology is the limited Field of View, around 90°, which is smaller than the human one, around 220 °. Thus, nowadays, with the currently available headset, the two conditions can not be directly compared.

Finally considering that the majority of simulations should take place in familiar environments, as laboratories, yards, garages, surgical rooms or houses, and will be constituted by the repetition of multiple sessions, it will be necessary to consider not only the effect on performance and learning rates of the semantic of the scene, but also of Short Term and Long Term Memory.

Nonetheless, considering interaction, ideally in future, we will want to interact with the virtual environment using our own body, natural and intuitive gestures and poses and receiving realistic and coherent feedback. Moreover, in the industrial and medical context, the use of specific tools is often required. Thus, it would be interesting projecting and evaluating Mixed Reality solutions, where the user can handle real items, while seeing their virtual representation, and use them to interact with the virtual environment. A bare-hand interface would be ideal for this purpose, while controller would prevent trainees from handling other objects. For this reason, first, it will be necessary to improve our grab function to allow for finer manipulations of items. Recursive or predictive filters, taking into account not only the current hand-object state, as in our grab function, but also joints position recorded in previous frames, in order to estimate future states, could be an interesting option to investigate to make interaction more robust to tracking issues.

Moreover, starting from the obtained results, it would be possible to create new VR scenarios, involving selection and manipulation tasks. The complexity level of the proposed task, however, should be better defined, distinguishing between the difficulties due to interaction or motorial components and to the cognitive load. Indeed, the interpretation of results referred to the two different interaction modalities, controllers and Leap Motion, should be clearer. Finally, findings presented in this work could be used to implement and test specific applications for training and simulation in the industrial or medical field, for example assembling/disassembling or repairing an engine, diagnosing a system or performing a maintenance operation, simulating simple surgical operations, as sutures, medical manoeuvres or daily activities for the upper limbs rehabilitation. Even if a fine and realistic manipulation could be difficult to obtain, it would be still possible to design sequences of simple selection and manipulation actions, which the trainee has to follow and memorize.

Bibliography

[1] Cassandra N Aldaba, Paul J White, Ahmad Byagowi and Zahra Moussavi. 'Virtual reality body motion induced navigational controllers and their effects on simulator sickness and pathfinding'. In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE. 2017, pp. 4175–4178.

[2] George A Alvarez. 'Representing multiple objects as an ensemble enhances visual cognition'. In: *Trends in cognitive sciences* 15.3 (2011), pp. 122–131.

[3] George A Alvarez and Patrick Cavanagh. 'The capacity of visual short-term memory is set both by visual information load and by number of objects'. In: *Psychological science* 15.2 (2004), pp. 106–111.

[4] Ferran Argelaguet and Carlos Andujar. 'A survey of 3D object selection techniques for virtual environments'. In: *Computers & Graphics* 37.3 (2013), pp. 121–136.

[5] L Arms, Dianne Cook and Carolina Cruz-Neira. 'The benefits of statistical visualization in an immersive environment'. In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE. 1999, pp. 88–95.

[6] Mirjam Augstein, Thomas Neumayr and Thomas Burger. 'The Role of Haptics in User Input for People with Motor and Cognitive Impairments'. In: *Studies in health technology and informatics* 242 (2017), pp. 183–194.

[7] Malika Auvray and J O'Regan. 'L'influence des facteurs sémantiques sur la cécité aux changements progressifs dans les scènes visuelles'. In: *L'année Psychologique* 103.1 (2003), pp. 9–32.

[8] Chiara Bassano, Giorgio Ballestin, Eleonora Ceccaldi, Fanny Isabelle Larradet, Maurizio Mancini, Erica Volta and Radoslaw Niewiadomski. 'A VR Game-Based System for Multimodal Emotion Data Collection'. In: 2019.

[9] Chiara Bassano, Manuela Chessa, Luca Fengone, Luca Isgró, Fabio Solari, Giovanni Spallarossa, Davide Tozzi and Aldo Zini. 'Evaluation of a Virtual Reality System for Ship Handling Simulations'. In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP,* INSTICC. SciTePress, 2019, pp. 62–73.

[10] Chiara Bassano, Manuela Chessa and Fabio Solari. 'A Study on the Role of Feedback and Interface Modalities for Natural Interaction in Virtual Reality Environments.' In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (2020), to appear.

[11] Chiara Bassano, Fabio Solari and Manuela Chessa. 'Studying Natural Human-computer Interaction in Immersive Virtual Reality: A Comparison between Actions in the Peripersonal and in the Near-action Space.' In: *VISIGRAPP (2: HUCAPP)*. 2018, pp. 108–115.

[12] Solari Fabio Chessa Manuela Bassano Chiara. 'The Limits of Visual Perception in Immersive Virtual Reality: A Change Blindness Study 42nd European Conference on Visual Perception (ECVP) 2019 Leuven'. In: *Perception* 48.2_suppl (2019), p. 220.

[13] Y. Baveye, E. Dellandréa, C. Chamaret and L. Chen. 'LIRIS-ACCEDE: A Video Database for Affective Content Analysis'. In: *IEEE Transactions on Affective Computing* 6.1 (2015), pp. 43–55. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2015.2396531.

[14] Mercedes Barchilon Ben-Av and Dov Sagi. 'Perceptual grouping by similarity and proximity: Experimental results can be predicted by intensity autocorrelations'. In: *Vision research* 35.6 (1995), pp. 853–866.

[15] Knud Benedict, Matthias Kirchhoff, Michael Gluch, Sandro Fischer, Michele Schaub, Michael Baldauf and Sebastian Klaes. 'Simulation Augmented Manoeuvring Design and Monitoring-a New Method for Advanced Ship Handling'. In: *TransNav, International Journal on Marine Navigation and Safety od Sea Transportation* 8.1 (2014).

[16] Samuel Benveniste, Pierre Jouvelot and Renaud Péquignot. 'The MINWii Project: Renarcissization of patients suffering from Alzheimer's Disease through video game-based music therapy'. In: *International Conference on Entertainment Computing*. Springer. 2010, pp. 79–90.

[17] Ned Block. 'Perceptual consciousness overflows cognitive access'. In: *Trends in cognitive sciences* 15.12 (2011), pp. 567–575.

[18] Jelte E Bos, Sjoerd C de Vries, Martijn L van Emmerik and Eric L Groen. 'The effect of internal and external fields of view on visually induced motion sickness'. In: *Applied ergonomics* 41.4 (2010), pp. 516–521.

[19] Cristina Botella, Javier Fernández-Álvarez, Verónica Guillén, Azucena García-Palacios and Rosa Baños. 'Recent Progress in Virtual Reality Exposure Therapy for Phobias: A Systematic Review'. In: *Current Psychiatry Reports* 19.7 (2017), p. 42.

[20] Stephane Bouchard, Patrice Renaud, Genevieve Robillard and Julie St-Jacques. 'Applications of virtual reality in clinical psychology: illustrations with the treatment of anxiety disorders'. In: *IEEE International Workshop HAVE Haptic Virtual Environments and Their*. IEEE. 2002, pp. 7–11.

[21] Stéphane Bouchard, Geneviève Robillard, Patrice Renaud and François Bernier. 'Exploring new dimensions in the assessment of virtual reality induced side effects'. In: *Journal of computer and information technology* 1.3 (2011), pp. 20–32.

[22] Doug A Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa and Wolfgang Stuerzlinger. '3d user interfaces: New directions and perspectives'. In: *IEEE computer graphics and applications* 28.6 (2008), pp. 20–36.

[23] Doug A Bowman and Larry F Hodges. 'An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments'. In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM. 1997, 35–ff.

[24] Doug A Bowman, David Koller and Larry F Hodges. 'Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques'. In: *Virtual Reality Annual International Symposium, 1997., IEEE 1997*. IEEE. 1997, pp. 45–52.

[25] Doug A Bowman, Ryan P McMahan and Eric D Ragan. 'Questioning naturalism in 3D user interfaces'. In: *Communications of the ACM* 55.9 (2012), pp. 78–88.

[26] James R Brockmole and John M Henderson. 'Object appearance, disappearance, and attention prioritization in real-world scenes'. In: *Psychonomic bulletin & review* 12.6 (2005), pp. 1061–1067.

[27] Henry Brodaty, Dimity Pond, Nicola M Kemp, Georgina Luscombe, Louise Harding, Karen Berman and Felicia A Huppert. 'The GPCOG: a new screening test for dementia designed for general practice'. In: *Journal of the American Geriatrics Society* 50.3 (2002), pp. 530–534.

[28] Kirsten Cater, Alan Chalmers and Colin Dalton. 'Varying rendering fidelity by exploiting human change blindness'. In: *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. ACM. 2003, pp. 39–46.

[29] Dennis Charsky. 'From edutainment to serious games: A change in the use of game characteristics'. In: *Games and culture* 5.2 (2010), pp. 177–198.

[30] Jeff Chastine, Nicole Kosoris and Joshua Skelton. 'A study of gesture-based first person control'. In: *Proceedings of CGAMES'2013 USA*. IEEE. 2013, pp. 79–86.

[31] Bernard Cheng, Mallory Ketcheson, Jordan van der Kroon and TC Nicholas Graham. 'Corgi Defence: Building In A Virtual Reality Environment'. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. 2015, pp. 763–766.

[32] Manuela Chessa, Gabriele Balocchi, Michela Busi, Antonio Novellino and Fabio Solari. 'An Oculus Rift based Exergame to Improve Awareness in Disabled People.' In: *VISIGRAPP (4: VISAPP)*. 2016, pp. 770–777.

[33] Manuela Chessa, Chiara Bassano, Elisa Gusai, Alice E Martis and Fabio Solari. 'Human-computer interaction approaches for the assessment and the practice of the cognitive capabilities of elderly people.' In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 66–81.

[34] Manuela Chessa, Matteo Garibotti, Valerio Rossi, Antonio Novellino and Fabio Solari. 'A virtual holographic display case for museum installations'. In: *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*. IEEE. 2015, pp. 69–73.

[35] Manuela Chessa, Guido Maiello, Alessia Borsari and Peter J Bex. 'The perceptual quality of the oculus rift for immersive virtual reality'. In: *Human–computer interaction* 34.1 (2019), pp. 51–82.

[36] Alice Chirico, Pietro Cipresso, Giuseppe Riva and Andrea Gaggioli. 'A Process for Selecting and Validating Awe-Inducing Audio-Visual Stimuli'. In: *Pervasive Computing Paradigms for Mental Health*. Ed. by Nuria Oliver, Silvia Serino, Aleksandar Matic, Pietro Cipresso, Nenad Filipovic and Liljana Gavrilovska. Cham: Springer International Publishing, 2018, pp. 19–27. ISBN: 978-3-319-74935-8.

[37] Alice Chirico, Pietro Cipresso, David B. Yaden, Federica Biassoni, Giuseppe Riva and Andrea Gaggioli. 'Effectiveness of Immersive Videos in Inducing Awe: An Experimental Study'. In: *Scientific Reports* 7 (2017). DOI: 10.1038/s41598-017-01242-0.

[38] Chris G Christou and Heinrich H Bülthoff. 'View dependence in scene recognition after active learning'. In: *Memory & Cognition* 27.6 (1999), pp. 996–1007.

[39] Chris Christou. 'Virtual reality in education'. In: *Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience*. IGI Global, 2010, pp. 228–243.

[40] Joanna C Coelho and Fons J Verbeek. 'Pointing task evaluation of Leap Motion controller in 3D virtual environment'. In: *Creating the Difference* 78 (2014), pp. 78–85.

[41] Michael A Cohen, Daniel C Dennett and Nancy Kanwisher. 'What is the bandwidth of perceptual experience?' In: *Trends in cognitive sciences* 20.5 (2016), pp. 324–335.

[42] Brendan Corbett, Chang S Nam and Takehiko Yamaguchi. 'The effects of haptic feedback and visual distraction on pointing task performance'. In: *International Journal of Human-Computer Interaction* 32.2 (2016), pp. 89–102.

[43] Maxime Cordeil, Benjamin Bach, Yongchao Li, Elliott Wilson and Tim Dwyer. 'Design space for spatio-data coordination: Tangible interaction devices for immersive information visualisation'. In: *2017 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE. 2017, pp. 46–50.

[44] Nick Corriveau Lecavalier, Émilie Ouellet, Benjamin Boller and Sylvie Belleville. 'Use of immersive virtual reality to assess episodic memory: A validation study in older adults'. In: *Neuropsychological rehabilitation* (2018), pp. 1–19.

[45] Nelson Cowan. 'Metatheory of storage capacity limits'. In: *Behavioral and brain sciences* 24.1 (2001), pp. 154–176.

[46] Matthew Coxon, Nathan Kelly and Sarah Page. 'Individual differences in virtual reality: Are spatial presence and spatial ability linked?' In: *Virtual Reality* 20.4 (2016), pp. 203–212.

[47] Carolina Cruz-Neira, Daniel J Sandin and Thomas A DeFanti. 'Surround-screen projection-based virtual reality: the design and implementation of the CAVE'. In: *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. 1993, pp. 135–142.

[48] Eduardo Cuervo. 'BEYOND REALITY: Head-Mounted Displays for Mobile Systems Researchers'. In: *GetMobile: Mobile Computing and Communications* 21.2 (2017), pp. 9–15.

[49] Elise S. Dan-Glauser and Klaus R. Scherer. 'The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance'. In: *Behavior Research Methods* 43.2 (2011), p. 468. ISSN: 1554-3528. DOI: 10.3758/s13428-011-0064-1. URL: https://doi.org/10.3758/s13428-011-0064-1.

[50] Nasser H Dardas and Mohammad Alhaj. 'Hand gesture interaction with a 3D virtual environment'. In: *The Research Bulletin of Jordan ACM* 2.3 (2011), pp. 86–94.

[51] Simon Davis, Keith Nesbitt and Eugene Nalivaiko. 'A systematic review of cybersickness'. In: *Proceedings of the 2014 Conference on Interactive Entertainment*. ACM. 2014, pp. 1–9.

[52] Simon Davis, Keith Nesbitt and Eugene Nalivaiko. '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)*. Vol. 27. 2015, p. 30.

[53] Julia Diemer, Georg W Alpers, Henrik M Peperkorn, Youssef Shiban and Andreas Mühlberger. 'The impact of perception and presence on emotional reactions: a review of research in virtual reality'. In: *Frontiers in psychology* 6 (2015), p. 26.

[54] Julia Diemer, Andreas Mühlberger, Paul Pauli and Peter Zwanzger. 'Virtual reality exposure in anxiety disorders: impact on psychophysiological reactivity'. In: *The World Journal of Biological Psychiatry* 15.6 (2014), pp. 427–442.

[55] Mahzar Eisapour, Shi Cao, Laura Domenicucci and Jennifer Boger. 'Participatory design of a virtual reality exercise for people with mild cognitive impairment'. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–9.

[56] Audrey Feldman and Linda Acredolo. 'The effect of active versus passive exploration on memory for spatial location in children'. In: *Child development* (1979), pp. 698–704.

[57] Rebecca M Foerster, Christian H Poth, Christian Behler, Mario Botsch and Werner X Schneider. 'Using the virtual reality device Oculus Rift for neuropsychological assessment of visual processing capabilities'. In: *Scientific reports* 6 (2016), p. 37016.

[58] Marshal F Folstein, Susan E Folstein and Paul R McHugh. '"Mini-mental state": a practical method for grading the cognitive state of patients for the clinician'. In: *Journal of psychiatric research* 12.3 (1975), pp. 189–198.

[59] Johnny J. R. Fontaine, Klaus R. Scherer and Cristina Soriano. *Components of Emotional Meaning: A sourcebook*. Ed. by Oxford University Press. 2013. DOI: 10.1093/acprof:oso/9780199592746.001.0001.

[60] Steven L Franconeri, George A Alvarez and James T Enns. 'How many locations can be selected at once?' In: *Journal of Experimental Psychology: Human Perception and Performance* 33.5 (2007), p. 1003.

[61] Bernd Frohlich, Jan Hochstrate, Alexander Kulik and Anke Huckauf. 'On 3D input devices'. In: *IEEE computer graphics and applications* 26.2 (2006), pp. 15–19.

[62] Ron George and Joshua Blake. 'Objects, containers, gestures, and manipulations: Universal foundational metaphors of natural user interfaces'. In: *CHI 2010* (2010), pp. 10–15.

[63] Peter J Gianaros, Eric R Muth, J Toby Mordkoff, Max E Levine and Robert M Stern. 'A questionnaire for the assessment of the multiple dimensions of motion sickness'. In: *Aviation, space, and environmental medicine* 72.2 (2001), p. 115.

[64] Konstantinos Giannakouris et al. 'Ageing characterises the demographic perspectives of the European societies'. In: *Statistics in focus* 72 (2008), p. 2008.

[65] Alessandra Gorini, Claret S Capideville, Gianluca De Leo, Fabrizia Mantovani and Giuseppe Riva. 'The role of immersion and narrative in mediated presence: the virtual hospital experience'. In: *Cyberpsychology, Behavior, and Social Networking* 14.3 (2011), pp. 99–105.

[66] Alessandra Grassi, Andrea Giaggioli and Giuseppe Riva. 'The influence of media content and media form in sense of presence: a preliminary study'. In: *Proceedings of the 11th Annual International Workshop on PRESENCE*. 2008, pp. 16–18.

[67] Kenny Gruchalla. 'Immersive well-path editing: investigating the added value of immersion'. In: *IEEE Virtual Reality 2004*. IEEE. 2004, pp. 157–164.

[68] Elisa Gusai, Chiara Bassano, Fabio Solari and Manuela Chessa. 'Interaction in an immersive collaborative virtual reality environment: a comparison between leap motion and HTC controllers'. In: *International Conference on Image Analysis and Processing*. Springer. 2017, pp. 290–300.

[69] Alexey N Gusev, Olga A Mikhailova and Igor S Utochkin. 'Stimulus determinants of the phenomenon of change blindness'. In: *Psychology in Russia* 7.1 (2014), p. 122.

[70] Bruno Herbelin, Frederic Vexo and Daniel Thalmann. 'Sense of presence in virtual reality exposures therapy'. In: *Proceedings of the 1st International Workshop on Virtual Reality Rehabilitation, Lausanne, Switzerland*. Citeseer. 2002.

[71] Hunter G Hoffman, Walter J Meyer III, Maribel Ramirez, Linda Roberts, Eric J Seibel, Barbara Atzori, Sam R Sharar and David R Patterson. 'Feasibility of articulated arm mounted Oculus Rift Virtual Reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients'. In: *Cyberpsychology, Behavior, and Social Networking* 17.6 (2014), pp. 397–401.

[72] Quinate Chioma Ihemedu-Steinke, Rainer Erbach, Prashanth Halady, Gerrit Meixner and Michael Weber. 'Virtual reality driving simulator based on head-mounted displays'. In: *Automotive user interfaces*. Springer, 2017, pp. 401–428.

[73] MA Just, David Stirling, Montserrat Ros, Fazel Naghdy and P Stapley. 'A comparison of upper limb movement profiles when reaching to virtual and real targets using the Oculus Rift: implications for virtual-reality enhanced stroke rehabilitation'. In: *J. Pain Manag* 9 (2014), pp. 277–281.

[74] Hacer Uke Karacan, Kursat Cagiltay and H Gurkan Tekman. 'Change detection in desktop virtual environments: An eye-tracking study'. In: *Computers in human behavior* 26.6 (2010), pp. 1305–1313.

[75] Robert S Kennedy, Norman E Lane, Kevin S Berbaum and Michael G Lilienthal. 'Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness'. In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220.

[76] Behrang Keshavarz and Heiko Hecht. 'Validating an efficient method to quantify motion sickness'. In: *Human factors* 53.4 (2011), pp. 415–426.

[77] Shaiyan Keshvari, Ronald Van den Berg and Wei Ji Ma. 'No evidence for an item limit in change detection'. In: *PLoS computational biology* 9.2 (2013).

[78] Saba Khashe, Burcin Becerik-Gerber, Gale Lucas and Jonathan Gratch. 'Persuasive effects of immersion in virtual environments for measuring pro-environmental behaviors'. In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 35. IAARC Publications. 2018, pp. 1–7.

[79] Saba Khashe, Gale Lucas, Burcin Becerik-Gerber and Jonathan Gratch. 'Buildings with persona: Towards effective building-occupant communication'. In: *Computers in Human Behavior* 75 (2017), pp. 607–618.

[80] Chaowanan Khundam. 'First person movement control with palm normal and hand gesture interaction in virtual reality'. In: *Computer Science and Software Engineering (JCSSE), 2015 12th International Joint Conference on*. IEEE. 2015, pp. 325–330.

[81] Hyosun Kim, Georgia Albuquerque, Sven Havemann and Dieter W Fellner. 'Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling.' In: *IPT/EGVE*. 2005, pp. 191–199.

[82] Kwanguk Kim, M Zachary Rosenthal, David J Zielinski and Rachael Brady. 'Effects of virtual environment platforms on emotional responses'. In: *Computer methods and programs in biomedicine* 113.3 (2014), pp. 882–893.

[83] Young Youn Kim, Hyun Ju Kim, Eun Nam Kim, Hee Dong Ko and Hyun Taek Kim. 'Characteristic changes in the physiological components of cybersickness'. In: *Psychophysiology* 42.5 (2005), pp. 616–625.

[84] Max Kinateder, Enrico Ronchi, Daniel Nilsson, Margrethe Kobes, Mathias Müller, Paul Pauli and Andreas Mühlberger. 'Virtual reality for fire evacuation research'. In: *2014 Federated Conference on Computer Science and Information Systems*. IEEE. 2014, pp. 313–321.

[85] MS Kizil and J Joy. 'What can virtual reality do for safety'. In: *University of Queensland, St. Lucia QLD* (2001).

[86] Jacqueline Kory and Sidney D'Mello. 'Affect Elicitation for Affective Computing'. In: *The Oxford Handbook of Affective Computing*. Ed. by Rafael Calvo, Sidney D'Mello, Jonathan Gratch and Arvid Kappas. Oxford University Press, 2014. ISBN: 9780199942237. DOI: `0.1093/oxfordhb/9780199942237.013.001`.

[87] Sid Kouider, Vincent De Gardelle, Jérôme Sackur and Emmanuel Dupoux. 'How rich is consciousness? The partial awareness hypothesis'. In: *Trends in cognitive sciences* 14.7 (2010), pp. 301–307.

[88] Joseph J LaViola Jr. 'A discussion of cybersickness in virtual environments'. In: *ACM Sigchi Bulletin* 32.1 (2000), pp. 47–56.

[89] Fedwa Laamarti, Mohamad Eid and Abdulmotaleb El Saddik. 'An overview of serious games'. In: *International Journal of Computer Games Technology* 2014 (2014).

[90] Dominique Lamy, Hannah Segal and Lital Ruderman. 'Grouping does not require attention'. In: *Perception & Psychophysics* 68.1 (2006), pp. 17–31.

[91] Hakwan Lau and David Rosenthal. 'Empirical support for higher-order theories of conscious awareness'. In: *Trends in cognitive sciences* 15.8 (2011), pp. 365–373.

[92] Bettina Laugwitz, Theo Held and Martin Schrepp. 'Construction and evaluation of a user experience questionnaire'. In: *Symposium of the Austrian HCI and Usability Engineering Group*. 2008, pp. 63–76.

[93] BD Lawson and AM Mead. 'The sopite syndrome revisited: drowsiness and mood changes during real or apparent motion'. In: *Acta astronautica* 43.3-6 (1998), pp. 181–192.

[94] Richard S Lazarus. 'Progress on a cognitive-motivational-relational theory of emotion.' In: *American psychologist* 46.8 (1991), p. 819.

[95] Jiaofeng Li, Jiehui Qian and Fan Liang. 'Evidence for the beneficial effect of perceptual grouping on visual working memory: an empirical study on illusory contour and a meta-analytic study'. In: *Scientific reports* 8.1 (2018), p. 13864.

[96] Steven J Luck and Edward K Vogel. 'The capacity of visual working memory for features and conjunctions'. In: *Nature* 390.6657 (1997), p. 279.

[97] Steven J Luck and Edward K Vogel. 'Visual working memory capacity: from psychophysics and neurobiology to individual differences'. In: *Trends in cognitive sciences* 17.8 (2013), pp. 391–400.

[98] Pauline Maillot, Alexandra Perrot and Alan Hartley. 'Effects of interactive physical-activity video-game training on physical and cognitive function in older adults.' In: *Psychology and aging* 27.3 (2012), p. 589.

[99] Valeria Manera, Pierre-David Petit, Alexandre Derreumaux, Ivan Orvieto, Matteo Romagnoli, Graham Lyttle, Renaud David and Philippe H Robert. ''Kitchen and cooking,'a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study'. In: *Frontiers in aging neuroscience* 7 (2015), p. 24.

[100] Alice E Martis, Chiara Bassano, Fabio Solari and Manuela Chessa. 'Going to a virtual supermarket: comparison of different techniques for interacting in a serious game for the assessment of the cognitive status'. In: *International Conference on Image Analysis and Processing*. Springer. 2017, pp. 281–289.

[101] Simon McCallum and Costas Boletsis. 'Dementia games: A literature review of dementia-related serious games'. In: *International conference on serious games development and applications*. Springer. 2013, pp. 15–27.

[102] Michael P McCreery, PG Schrader, S Kathleen Krach and Randy Boone. 'A sense of self: The role of presence in virtual environments'. In: *Computers in Human Behavior* 29.4 (2013), pp. 1635–1640.

[103] Mark McGill, Alexander Ng and Stephen Brewster. '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*. ACM. 2017, pp. 5655–5668.

[104] Robert N McLay, Dennis P Wood, Jennifer A Webb-Murphy, James L Spira, Mark D Wiederhold, Jeffrey M Pyne and Brenda K Wiederhold. 'A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder'. In: *Cyberpsychology, behavior, and social networking* 14.4 (2011), pp. 223–229.

[105] Ryan P McMahan, Doug A Bowman, David J Zielinski and Rachael B Brady. 'Evaluating display fidelity and interaction fidelity in a virtual reality game'. In: *IEEE transactions on visualization and computer graphics* 18.4 (2012), pp. 626–633.

[106] Ryan P McMahan, Doug Gorton, Joe Gresock, Will McConnell and Doug A Bowman. 'Separating the effects of level of immersion and 3D interaction techniques'. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 2006, pp. 108–111.

[107] Ryan Patrick McMahan. 'Exploring the effects of higher-fidelity display and interaction for virtual reality games'. PhD thesis. Virginia Tech, 2011.

[108] B. Meuleman and D. Rudrauf. 'Induction and profiling of strong multi-componential emotions in virtual reality'. In: *IEEE Transactions on Affective Computing* (2018), pp. 1–1. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2018.2864730.

[109] George A Miller. 'The magical number seven, plus or minus two: Some limits on our capacity for processing information.' In: *Psychological review* 63.2 (1956), p. 81.

[110] Mark R Mine. 'Virtual environment interaction techniques'. In: *UNC Chapel Hill CS Dept* (1995).

[111] Mark R Mine, Frederick P Brooks Jr and Carlo H Sequin. 'Moving objects in space: exploiting proprioception in virtual-environment interaction'. In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co. 1997, pp. 19–26.

[112] Teresa Monahan, Gavin McArdle and Michela Bertolotto. 'Virtual reality for collaborative e-learning'. In: *Computers & Education* 50.4 (2008), pp. 1339–1353.

[113] Andreas Mühlberger, Heinrich H Bülthoff, Georg Wiedemann and Paul Pauli. 'Virtual reality for the psychophysiological assessment of phobic fear: responses during virtual tunnel driving.' In: *Psychological assessment* 19.3 (2007), p. 340.

[114] Justin Munafo, Meg Diedrick and Thomas A Stoffregen. 'The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects'. In: *Experimental brain research* 235.3 (2017), pp. 889–901.

[115] Eric R Muth, Robert M Stern, Julian F Thayer and Kenneth L Koch. 'Assessment of the multiple dimensions of nausea: the Nausea Profile (NP)'. In: *Journal of psychosomatic research* 40.5 (1996), pp. 511–520.

[116] Eugene Nalivaiko, Simon L Davis, Karen L Blackmore, Andrew Vakulin and Keith V Nesbitt. 'Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time'. In: *Physiology & behavior* 151 (2015), pp. 583–590.

[117] Keith Nesbitt, Simon Davis, Karen Blackmore and Eugene Nalivaiko. 'Correlating reaction time and nausea measures with traditional measures of cybersickness'. In: *Displays* 48 (2017), pp. 1–8.

[118] Sarah Nichols, Sue Cobb and John R Wilson. 'Health and safety implications of virtual environments: Measurement issues'. In: *Presence: Teleoperators & Virtual Environments* 6.6 (1997), pp. 667–675.

[119] R. Niewiadomski, M. Mancini, G. Varni, G. Volpe and A. Camurri. 'Automated Laughter Detection From Full-Body Movements'. In: *IEEE Transactions on Human-Machine Systems* 46.1 (2016), pp. 113–123. ISSN: 2168-2291. DOI: 10.1109/THMS.2015.2480843.

[120] Seizo Ohyama, Suetaka Nishiike, Hiroshi Watanabe, Katsunori Matsuoka, Hironori Akizuki, Noriaki Takeda and Tamotsu Harada. 'Autonomic responses during motion sickness induced by virtual reality'. In: *Auris Nasus Larynx* 34.3 (2007), pp. 303–306.

[121] Camila R Oliveira, Brandel José P Lopes Filho, Michael A Sugarman, Cristiane S Esteves, Margarida Maria BMP Lima, Carmen Moret-Tatay, Tatiana Q Irigaray and Irani Iracema L Argimon. 'Development and feasibility of a virtual reality task for the cognitive assessment of older adults: the ECO-VR'. In: *The Spanish journal of psychology* 19 (2016).

[122] Gangrae Park, Hyunmin Choi, Uichin Lee and Seongah Chin. 'Virtual figure model crafting with VR HMD and Leap Motion'. In: *The Imaging Science Journal* 65.6 (2017), pp. 358–370.

[123] Thomas D Parsons, Susan McPherson and Victoria Interrante. 'Enhancing neurocognitive assessment using immersive virtual reality'. In: *2013 1st Workshop on Virtual and Augmented Assistive Technology (VAAT)*. IEEE. 2013, pp. 27–34.

[124] Eric Pfeiffer. 'A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients'. In: *Journal of the American Geriatrics Society* 23.10 (1975), pp. 433–441.

[125] WA Phillips. 'On the distinction between sensory storage and short-term visual memory'. In: *Perception & Psychophysics* 16.2 (1974), pp. 283–290.

[126] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst and Tadao Ichikawa. 'The go-go interaction technique: non-linear mapping for direct manipulation in VR'. In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 79–80.

[127] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst and Tadao Ichikawa. 'A framework and testbed for studying manipulation techniques for immersive VR'. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 1997, pp. 21–28.

[128] Mores Prachyabrued and Christoph W Borst. 'Visual feedback for virtual grasping'. In: *2014 IEEE symposium on 3D user interfaces (3DUI)*. IEEE. 2014, pp. 19–26.

[129] Martin Prince, Adelina Comas-Herrera, Martin Knapp, Maëlenn Guerchet and Maria Karagiannidou. 'World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future'. In: (2016).

[130] Yeshwanth Pulijala, Minhua Ma and Ashraf Ayoub. 'VR surgery: Interactive virtual reality application for training oral and maxillofacial surgeons using oculus rift and leap motion'. In: *Serious Games and Edutainment Applications*. Springer, 2017, pp. 187–202.

[131] Siddharth S Rautaray and Anupam Agrawal. 'Vision based hand gesture recognition for human computer interaction: a survey'. In: *Artificial Intelligence Review* 43.1 (2015), pp. 1–54.

[132] James T Reason and Joseph John Brand. *Motion sickness.* Academic press, 1975.

[133] Sandeep Reddivari, Jason Smith and Jonathan Pabalate. 'VRvisu: a tool for virtual reality based visualization of medical data'. In: *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE. 2017, pp. 280–281.

[134] Holger Regenbrecht and Thomas Schubert. 'Real and illusory interactions enhance presence in virtual environments'. In: *Presence: Teleoperators & Virtual Environments* 11.4 (2002), pp. 425–434.

[135] Ronald A Rensink. 'Change blindness'. In: *Neurobiology of attention*. Elsevier, 2005, pp. 76–81.

[136] Ronald A Rensink, J Kevin O'Regan and James J Clark. 'To see or not to see: The need for attention to perceive changes in scenes'. In: *Psychological science* 8.5 (1997), pp. 368–373.

[137] Gary E Riccio and Thomas A Stoffregen. 'An ecological theory of motion sickness and postural instability'. In: *Ecological psychology* 3.3 (1991), pp. 195–240.

[138] Albert A Rizzo, Maria Schultheis, Kimberly A Kerns and Catherine Mateer. 'Analysis of assets for virtual reality applications in neuropsychology'. In: *Neuropsychological rehabilitation* 14.1-2 (2004), pp. 207–239.

[139] Ira J Roseman, Ann Aliki Antoniou and Paul E. Jose. 'Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory'. In: *Cognition & Emotion* 10.3 (1996), pp. 241–278.

[140] Barbara Olasov Rothbaum. 'Using virtual reality to help our patients in the real world.' In: *Depression and Anxiety* (2009).

[141] Jeffrey N Rouder, Richard D Morey, Candice C Morey and Nelson Cowan. 'How to measure working memory capacity in the change detection paradigm'. In: *Psychonomic Bulletin & Review* 18.2 (2011), pp. 324–330.

[142] Mattias Roupé, Petra Bosch-Sijtsema and Mikael Johansson. 'Interactive navigation interface for virtual reality using the human body'. In: *Computers, Environment and Urban Systems* 43 (2014), pp. 42–50.

[143] Roy A Ruddle, Stephen J Payne and Dylan M Jones. 'Navigating large-scale virtual environments: what differences occur between helmet-mounted and desk-top displays?' In: *Presence: Teleoperators & Virtual Environments* 8.2 (1999), pp. 157–168.

[144] Klaus Scherer. 'Appraisal considered as a process of multilevel sequential checking'. In: *Appraisal processes in emotion: Theory, methods, research* 92.120 (2001), p. 57.

[145] Martin Schrepp, Andreas Hinderks and Jörg Thomaschewski. 'Applying the user experience questionnaire (UEQ) in different evaluation scenarios'. In: *International Conference of Design, User Experience, and Usability*. 2014, pp. 383–392.

[146] Sarah Sharples, Sue Cobb, Amanda Moody and John R Wilson. 'Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems'. In: *Displays* 29.2 (2008), pp. 58–69.

[147] Thomas B Sheridan. 'Musings on telepresence and virtual presence'. In: *Presence: Teleoperators & Virtual Environments* 1.1 (1992), pp. 120–126.

[148] Hiroyuki Shinoda, Mary M Hayhoe and Anurag Shrivastava. 'What controls attention in natural environments?' In: *Vision research* 41.25-26 (2001), pp. 3535–3545.

[149] Daniel J Simons, Steven L Franconeri and Rebecca L Reimer. 'Change blindness in the absence of a visual disruption'. In: *Perception* 29.10 (2000), pp. 1143–1154.

[150] Daniel J Simons and Daniel T Levin. 'Failure to detect changes to people during a real-world interaction'. In: *Psychonomic Bulletin & Review* 5.4 (1998), pp. 644–649.

[151] Mel Slater, John McCarthy and Francesco Maringelli. 'The influence of body movement on subjective presence in virtual environments'. In: *Human Factors* 40.3 (1998), pp. 469–477.

[152] Daniele Sportillo, Alexis Paljic, Mehdi Boukhris, Philippe Fuchs, Luciano Ojeda and Vincent Roussarie. '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*. 2017, pp. 6–10.

[153] Simon Stannus, Daniel Rolf, Arko Lucieer and Winyu Chinthammit. 'Gestural navigation in Google Earth'. In: *Proceedings of the 23rd Australian Computer-Human Interaction Conference*. ACM. 2011, pp. 269–272.

[154] Frank Steinicke, Gerd Bruder, Klaus Hinrichs and Pete Willemsen. 'Change Blindness Phenomena for Virtual Reality Display Systems'. In: *IEEE Transactions on visualization and computer graphics* 17.9 (2011), pp. 1223–1233.

[155] Richard Stoakley, Matthew J Conway and Randy Pausch. 'Virtual reality on a WIM: interactive worlds in miniature'. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co. 1995, pp. 265–272.

[156] Evan A Suma, Seth Clark, David Krum, Samantha Finkelstein, Mark Bolas and Zachary Warte. 'Leveraging change blindness for redirection in virtual environments'. In: *2011 IEEE Virtual Reality Conference*. IEEE. 2011, pp. 159–166.

[157] Mirko Sužnjević, Matija Mandjurov and Maja Matijašević. '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)*. 2017.

[158] Ola Svenson and Kit Sjöberg. 'Speeds of subitizing and counting processes in different age groups'. In: *The Journal of genetic psychology* 142.2 (1983), pp. 203–211.

[159] Ioannis Tarnanas, Winfried Schlee, Magda Tsolaki, René Müri, Urs Mosimann and Tobias Nef. 'Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study'. In: *JMIR serious games* 1.1 (2013), e1.

[160] Jochen Triesch, Dana H Ballard, Mary M Hayhoe and Brian T Sullivan. 'What you see is what you need'. In: *Journal of vision* 3.1 (2003), pp. 9–9.

[161] Robert Tscharn, Philipp Schaper, Jan Sauerstein, Sarah Steinke, Sebastian Stiersdorfer, Carsten Scheller and Huy Tan Huynh. 'User Experience of 3D Map Navigation–Bare-Hand Interaction or Touchable Device?' In: *Mensch und Computer 2016-Tagungsband* (2016).

[162] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater and Frederick P Brooks Jr. 'Walking> walking-in-place> flying, in virtual environments'. In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co. 1999, pp. 359–364.

[163] Martin Usoh, Ernest Catena, Sima Arman and Mel Slater. 'Using presence questionnaires in reality'. In: *Presence: Teleoperators & Virtual Environments* 9.5 (2000), pp. 497–503.

[164] D Alexander Varakin, Daniel T Levin and Krista M Collins. 'Comparison and representation failures both cause real-world change blindness'. In: *Perception* 36.5 (2007), pp. 737–749.

[165] J.M. Varela and C. Guedes Soares. 'Interactive 3D desktop ship simulator for testing and training offloading manoeuvres'. In: *Applied Ocean Research* 51 (2015), pp. 367 –380.

[166] José Miguel Varela, JM Rodrigues and C Guedes Soares. '3D simulation of ship motions to support the planning of rescue operations on damaged ships'. In: *Procedia Computer Science* 51 (2015), pp. 2397–2405.

[167] Madis Vasser, Markus Kängsepp and Jaan Aru. 'Change Blindness in 3D Virtual Reality'. In: *arXiv preprint arXiv:1508.05782* (2015).

[168] Sara Ventura, Eleonora Brivio, Giuseppe Riva and Rosa M Baños. 'Immersive Versus Non-immersive Experience: Exploring the Feasibility of Memory Assessment Through 360° Technology'. In: *Frontiers in Psychology* 10 (2019).

[169] Gualtiero Volpe, Paolo Alborno, Antonio Camurri, Paolo Coletta, Simone Ghisio, Maurizio Mancini, Radoslaw Niewiadomski and Stefano Piana. 'Designing Multimodal Interactive Systems using EyesWeb XMI.' In: *SERVE@ AVI*. 2016, pp. 49–56.

[170] Athanasios Vourvopoulos, Ana Lúcia Faria, Kushal Ponnam and Sergi Bermudez i Badia. 'RehabCity: design and validation of a cognitive assessment and rehabilitation tool through gamified simulations of activities of daily living'. In: *Proceedings of the 11th conference on advances in computer entertainment technology*. 2014, pp. 1–8.

[171] Lidong Wang, Guanghui Wang and Cheryl Ann Alexander. 'Big data and visualization: methods, challenges and technology progress'. In: *Digital Technologies* 1.1 (2015), pp. 33–38.

[172] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak and Denis Fisseler. 'Analysis of the accuracy and robustness of the leap motion controller'. In: *Sensors* 13.5 (2013), pp. 6380–6393.

[173] Florian Weidner, Anne Hoesch, Sandra Poeschl and Wolfgang Broll. 'Comparing VR and non-VR driving simulations: An experimental user study'. In: *2017 IEEE Virtual Reality (VR)*. IEEE. 2017, pp. 281–282.

[174] Michael Wertheimer. 'A Source Book of Gestalt Psychology.' In: *Psyccritiques* 13.8 (1968).

[175] Mary E Wheeler and Anne M Treisman. 'Binding in short-term visual memory.' In: *Journal of Experimental Psychology: General* 131.1 (2002), p. 48.

[176] Curtis Wilkes and Doug A Bowman. 'Advantages of velocity-based scaling for distant 3D manipulation'. In: *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. ACM. 2008, pp. 23–29.

[177] Christopher J Wilson and Alessandro Soranzo. 'The use of virtual reality in psychology: A case study in visual perception'. In: *Computational and mathematical methods in medicine* 2015 (2015).

[178] Werner Wirth, Tilo Hartmann, Saskia Böcking, Peter Vorderer, Christoph Klimmt, Holger Schramm, Timo Saari, Jari Laarni, Niklas Ravaja, Feliz Ribeiro Gouveia et al. 'A process model of the formation of spatial presence experiences'. In: *Media psychology* 9.3 (2007), pp. 493–525.

[179] Bob G Witmer and Michael J Singer. 'Measuring presence in virtual environments: A presence questionnaire'. In: *Presence: Teleoperators and virtual environments* 7.3 (1998), pp. 225–240.

[180] Geoffrey F Woodman, Shaun P Vecera and Steven J Luck. 'Perceptual organization influences visual working memory'. In: *Psychonomic bulletin & review* 10.1 (2003), pp. 80–87.

[181] Peter Wozniak, Oliver Vauderwange, Avikarsha Mandal, Nicolas Javahiraly and Dan Curticapean. 'Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality'. In: *Optics Education and Outreach IV*. Vol. 9946. International Society for Optics and Photonics. 2016, 99460P.

[182] Jie Xu, Edward Huang, Liam Hsieh, Loo Hay Lee, Qing-Shan Jia and Chun-Hung Chen. 'Simulation optimization in the era of Industrial 4.0 and the Industrial Internet'. In: *Journal of Simulation* 10.4 (2016), pp. 310–320.

[183] Tal Yatziv and Yoav Kessler. 'A two-level hierarchical framework of visual short-term memory'. In: *Journal of vision* 18.9 (2018), pp. 2–2.

[184] Chaoxiong Ye, Hong-Jin Sun, Qianru Xu, Tengfei Liang, Yin Zhang and Qiang Liu. 'Working memory capacity affects trade-off between quality and quantity only when stimulus exposure duration is sufficient: Evidence for the two-phase model'. In: *Scientific Reports* 9.1 (2019), p. 8727.

[185] Mary K Young, Graham B Gaylor, Scott M Andrus and Bobby Bodenheimer. 'A comparison of two cost-differentiated virtual reality systems for perception and action tasks'. In: *Proceedings of the ACM Symposium on Applied Perception*. ACM. 2014, pp. 83–90.

[186] Shumin Zhai. 'Human performance in six degree of freedom input control'. PhD thesis. University of Toronto, 1996.

[187] Fan Zhang, Shaowei Chu, Ruifang Pan, Naye Ji and Lian Xi. 'Double hand-gesture interaction for walk-through in VR environment'. In: *Computer and Information Science (ICIS), 2017 IEEE/ACIS 16th International Conference on*. IEEE. 2017, pp. 539–544.

[188] Zhengyou Zhang. 'Microsoft Kinect sensor and its effect'. In: *IEEE multimedia* 19.2 (2012), pp. 4–10.

[189] Chiara Zucchella, Elena Sinforiani, Cristina Tassorelli, Elena Cavallini, Daniela Tost-Pardell, Sergi Grau, Stefania Pazzi, Stefano Puricelli, Sara Bernini, Sara Bottiroli et al. 'Serious games for screening pre-dementia conditions: from virtuality to reality? A pilot project'. In: *Functional neurology* 29.3 (2014), p. 153.

[190] Stelios Zygouris, Konstantinos Ntovas, Dimitrios Giakoumis, Konstantinos Votis, Stefanos Doumpoulakis, Sofia Segkouli, Charalampos Karagiannidis, Dimitrios Tzovaras and Magda Tsolaki. 'A preliminary study on the feasibility of using a virtual reality cognitive training application for remote detection of mild cognitive impairment'. In: *Journal of Alzheimer's Disease* 56.2 (2017), pp. 619–627.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Genova, Italy
May 2020

Chiara Bassano