Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Control and Computer Engineering (35th cycle)

Interaction in Virtual Reality Simulations

Davide Calandra

* * * * *

Supervisor
Prof. Fabrizio Lamberti

Doctoral Examination Committee:
Prof. Claudio Pacchierotti, Referee, IRISA, Université de Rennes, Inria, France
Prof. Michele Russo, Referee, Sapienza Università di Roma, Italy
Prof. Sofia Seinfeld, Universitat Politècnica de Catalunya, Spain
Prof. Bill Kapralos, OntarioTech University, Canada
Prof. Guido Marchetto, Politecnico di Torino, Italy

Politecnico di Torino
July 18, 2023

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .
Davide Calandra
Turin, July 18, 2023

Summary

Virtual Reality (VR) has emerged as a powerful technology for creating immersive and engaging simulations that enable users to interact with computer-generated environments in a natural and intuitive way. However, the design and implementation of effective interaction methods in VR remain challenging. The lack of proper haptic feedback, and the need to rely on input devices such as controllers or gestures, for example, can result in awkward or unnatural interactions, reducing the perceived level of realism and the immersion related to the VR experience. At the same time, the employment of poorly designed interaction paradigms may impair usability, reduce the the sense of presence, and even cause unpleasant effects related to the so called cybersickness.

This doctoral thesis, which covers a subset of the the research work performed in the three-year Ph.D. period, aims to address these challenges by investigating the role of interaction in VR simulations. The investigated topics range from the study of locomotion interfaces in VR, to the use of haptic interfaces for simulating passive and haptic tools applied to real life training use cases and the exploration of further forms of Human-Computer Interaction (HCI) and Human-Human Interaction (HHI) through voice and body gestures, also in the context of multi-user shared simulations.

Results obtained in the considered case studies cover a wide number of relevant aspects, such as realism, usability, and engagement of VR simulations, among others, ultimately leading to a validation of proposed approaches and methodologies. In this way, the thesis contributes to the understanding of how to design and evaluate interaction paradigms in VR simulations in order to enhance aspects related to User eXperience (UX), with the goal of letting users successfully achieve the intended simulation objectives.

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to all those who have contributed to my doctoral thesis.

First and foremost, I would like to thank my supervisor, Prof. Fabrizio Lamberti, for his impeccable guidance, constant support, and invaluable advice throughout the entire research process. His expertise, motivation, and encouragement have been instrumental in shaping the focus and direction of my research, and I am truly grateful for his unwavering commitment to my success.

I owe a great debt of gratitude to the faculty, staff, students, and colleagues of the Department of Control and Computer Engineering at Politecnico di Torino, for fostering an intellectually stimulating and supportive academic community where research ideas can flourish and thrive. Their collaborative efforts and diverse perspectives have provided me with a rich and stimulating research environment.

I would like to extend my thanks to my family and friends, who have been a constant source of love, encouragement, and support throughout my academic journey. Their unwavering belief in my abilities and passion for my research has been a driving force in my motivation and success.

Last but not least, I would like to express my gratitude to all the research participants who generously gave their time and insights to participate in my study. Their contributions have been invaluable in shaping my research and adding depth and nuance to my findings.

In conclusion, I am deeply indebted to all those who have supported and encouraged me on this journey, and without whom this thesis would not have been possible. Thank you all for your invaluable contributions, support, and encouragement.

I would like to dedicate this thesis to my dear family and friends, and in particular to my daughter Martina.

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong"

— Richard P. Feynman

Contents

List of Tables 10

List of Figures 11

1 Introduction 13
1.1 Background and Motivation . . . 13
1.2 Thesis Structure . . . 16

2 Locomotion and Haptic Interfaces 19
2.1 Evaluation Testbed for Locomotion . . . 22
2.1.1 Background . . . 23
2.1.2 Materials and Methods . . . 26
2.1.3 Experiment . . . 35
2.1.4 Use Case . . . 41
2.1.5 Results and Discussion . . . 43
2.2 Passive Haptic (PH) Interfaces for Interactive Simulations . . . 49
2.2.1 Background . . . 51
2.2.2 Materials and Methods . . . 55
2.2.3 Experiment . . . 62
2.2.4 Results and Discussion . . . 67
2.3 Consumer Devices for Haptic Simulation . . . 74
2.3.1 Background . . . 76
2.3.2 Materials and Methods . . . 78
2.3.3 Experiment . . . 82
2.3.4 Results and Discussion . . . 83
2.4 Considerations and Remarks . . . 88

3 Voice and Body Language 91
3.1 Voice-based Interfaces for Navigation . . . 93
3.1.1 Background . . . 94
3.1.2 Materials and Methods . . . 95
3.1.3 Experiment . . . 99

3.1.4 Results and Discussion 101
3.2 Avatar Representation for Multi-User Training Simulations 105
3.2.1 Background 107
3.2.2 Materials and Methods 108
3.2.3 Experiment 113
3.2.4 Results and Discussion 115
3.2.5 Embodiment 115
3.3 Emotion Conveyance in Social VR Avatars 119
3.3.1 Background 120
3.3.2 Materials and Methods 122
3.3.3 Experiment 124
3.3.4 Results and Discussion 126
3.4 Considerations and Remarks 128

4 Conclusions 133
4.1 Summary of Contributions 133
4.2 Limitations and Future Developments 136

Glossary 139

Bibliography 143

List of Tables

2.1 NRs taken into account for LET-VR 32
2.2 Metrics established to assess the UX with the locomotion technique provided for the specific task, categorized based on the relevant NRs they contribute to 34
2.3 The practice score metric results include the percentage of trainees who performed a specific action correctly. 68
2.4 Breakdown analysis of the compound configuration presented in Praticò et al. [40]. 80
2.5 Objective results related to the screwdriving task. 84
2.6 Subjective ranking of configurations (custom). 87
3.1 Details of the navigation tasks considered in the testing scenario. . . 99
3.2 Outcomes concerning the objective metrics. 101
3.3 Ranking by preference of the three interfaces. 104
3.4 Subjective outcomes concerning the direct comparison section of the questionnaire. 118

List of Figures

1.1 Various types of VR haptic interfaces identified by Wee et al. [3]. . 15
2.1 Various types of VR locomotion techniques identified by Templeman et al. [21]. 20
2.2 Classification of haptic devices based on the type of feedback provided. 21
2.3 Architecture of Locomotion Evaluation Testbed VR (LET-VR). . . 23
2.4 LET-VR's tasks (grouped into scenarios) and training scene 27
2.5 Locomotion paradigms evaluated with LET-VR 41
2.6 Objective metrics scores for the four evaluated techniques (with unitary weights). 45
2.7 Subjective per-scenario metrics scores for the four evaluated techniques with unitary weights. 46
2.8 Subjective overall metrics scores for the four evaluated techniques with unitary weights. 47
2.9 Wildland scenario depicted in the VR simulation. 58
2.10 Considered firefighting equipment (shovel, McLeod, and broom). . . 59
2.11 Effect of the firefighting tools on the simulation. 62
2.12 Objective evaluation phase of the experiment 65
2.13 Subjective results have been obtained based on the IMMS [101]. . . 70
2.14 Subjective outcomes concerning the AttrakDiff questionnaire [103]. . 71
2.15 Subjective outcomes concerning the VRUSE questionnaire [67] (V+VR trainees only). 72
2.16 Hardware and scenario used for the experiment. 79
2.17 Steps of the simulated procedure 81
2.18 Real ES procedure and evaluated haptic configurations. 83
2.19 Subjective outcomes concerning the post-real experience (with the ES) questionnaire. 84
2.20 Subjective outcomes concerning the UEQ. 85
2.21 Subjective outcomes concerning the SIM-TLX [123]. 85
2.22 Subjective outcomes concerning the VRUSE [67]. 86
3.1 Voice-based hands-free navigation techniques in the locomotion taxonomy [21]. 92
3.2 Example of workflow of the Speech-only (S) technique. 96

3.3 Example of workflow of the Speech with Gaze (SG) technique. . . . 97
3.4 Example of workflow of the and Speech with Gaze & Descriptions (SGD) technique. 98
3.5 Disambiguation panel used to solve ambiguities and navigation tasks implemented for the experiment. 100
3.6 Subjective outcomes concerning the SASSI [143]. 102
3.7 Subjective outcomes concerning the custom section of the questionnaire. 103
3.8 Own mirrored avatars shown to participants prior to the experience and other user avatars during the VR experience of the two evaluated modalities. 111
3.9 Screenshots of the VR simulation taken during the experiments. . . 113
3.10 Subjective outcomes concerning the Embodiment Questionnaire [160] sub-scales. 116
3.11 Subjective outcomes concerning the NMPS questionnaire [161] subscales. 118
3.12 VR setup selected for the experimental activity. 122
3.13 The stage area within the VE where participants can walk to spectate the avatar during the acting. 125
3.14 Subjective outcomes concerning the questionnaires for each scene. . 131

Chapter 1

Introduction

1.1 Background and Motivation

More then three decades have passed since the term Virtual Reality (VR) has been coined [1], and since then, advancements in that technology have been remarkable and increasingly rapid. For years this technology, falling today in the broader spectrum called eXtended Reality (XR) along with the sisters Augmented Reality (AR) and Mixed Reality (MR), has been perceived as pure sci-fi speculation. In the last ten years, however, the arrival of ever more mature technological solutions with increasingly practical implications on the consumer market have been witnessed [2]. From the first successful Kickstarter campaign of the Oculus Rift[1], which managed to provide a first affordable high-quality immersive VR Head-Mounted Display (HMD), to the fist appearance of VR systems capable to offer sensor-based tracking capable to support full room-scale movements (e.g., the HTC VIVE[2]), until the recent launch on the market of increasingly cheap, lightweight, feature-ful, all-in-one devices, aimed at mass adoption (e.g., the Meta Quest 2[3]). Along with the rapid evolution of commercial HMDs, industry and academia focused their efforts on studying how to achieve experiences that increasingly look and feel more real [3].

Having had the opportunity to engage with VR firsthand, author witnessed the transformative power it holds. VR simulations offer a unique medium that can transport individuals to immersive VEs, enabling them to interact with digital content and experiences in unprecedented ways. Through these personal encounters with VR applications, the immense potential and existing limitations in the realm

---

[1] Oculus Rift Kickstarter page: `https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game`

[2] HTC VIVE: `https://www.vive.com/us/support/vive/`

[3] Meta Quest 2: `https://www.meta.com/it/en/quest/products/quest-2/`

of interaction have been recognized.

Moreover, during author's experiences with VR, it has been observed that while some simulations excel in providing highly immersive and intuitive interactions, others fall short in delivering a seamless User eXperience (UX). These observations have sparked a deep curiosity to understand the underlying factors that contribute to effective interaction design in VR. By delving into the intricacies of interaction within virtual reality simulations, the aim is to contribute to the development of more natural, intuitive, and immersive UXs.

At an international level, the research community has recognized the significance of interaction in VR and its potential impact on various fields. Numerous studies have explored different aspects of interaction techniques within VR, such as hand tracking, gesture recognition, haptic feedback, locomotion, and User Interface (UI) design. However, despite significant progress, there are still several challenges that remain unresolved.

Interaction in VR can be classified into two main types. Human-Computer Interaction (HCI), which happens between computer systems and human users [4], and Human-Human Interaction (HHI), which takes place between human users and the virtual representation of other humans, usually in form of avatars [5]. In addition to this classification, it is crucial to consider the multidimensional nature of interactions within VR environments.

One essential aspect of interaction in VR is locomotion within the virtual space. The way users navigate and move in VEs [6] plays a vital role in their engagement and comfort. In this context, the issue of "incompatible spaces" is often encountered while attempting to create natural interaction in interactive simulations. In fact, although VEs offer unrestricted movement and infinite walking, spatial limitations exist within the physical environment where the simulation takes place [7]. In this context, literature proposed a wide range of VR locomotion techniques, typically designed to make use of users' spatial orientation skills and minimize the physical distance between their body movements and the corresponding movements in the VE. However, these solutions are usually characterized by very different strengths and weaknesses, and selecting the most suitable technique for a given use case is not straightforward.

A second fundamental aspect is related to the interaction with virtual elements within the VE [8]. The ability to grasp and manipulate virtual objects convincingly contributes to the sense of agency and can enable more complex interactions and tasks within VR simulations and applications. However, despite current commercial VR systems already managed to reach a stunningly excellent visual and audio feedback quality [9], a complete stimulation, involving the other senses (touch, taste, and smell), is still far from being achieved [3], due to the immaturity of the relative interaction technologies. To simulate the use of real-world objects with specific physical characteristics within the VE is necessary to provide accurate haptic feedback and mimicking the properties and behavior of real objects. This problem has been addressed in the literature by proposing a wide variety of solutions [3] (Figure 1.1). These devices can be categorized as *hand-held* (e.g., hand controllers), *wearable* (e.g., gloves, exoskeletons, thimbles), *physical props* (physical replicas of real-world objects used as interfaces with the virtual world), *encountered types* (based on devices that provide on-demand haptic feedback, such as drones or robotic arms), and *mid-air* (based on ultrasound). However, despite this wide variety of solutions, there is still plenty of room for improvement, e.g., regarding the aspects of multimodality and the realism of the simulation [3].

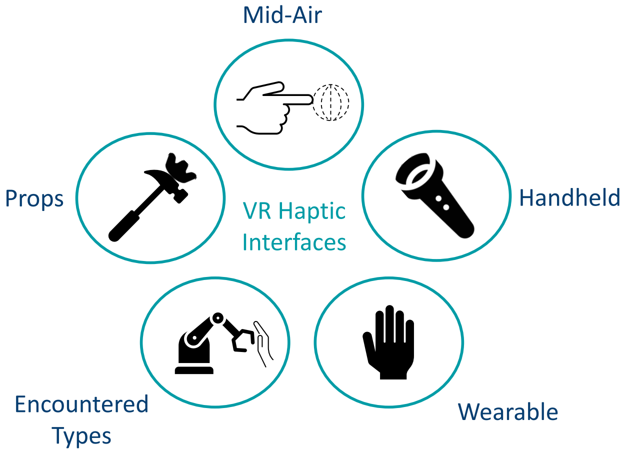


Figure 1.1: Various types of VR haptic interfaces identified by Wee et al. [3].

In addition to locomotion and haptic interfaces, another crucial aspect of interaction in VR is the integration of voice and body language. These modalities enable users to communicate and express themselves naturally within the VE, further enhancing the immersive and realistic experience.

Voice interaction plays a fundamental role in enabling users to engage with the virtual world through spoken commands, conversations, or verbal interactions with both virtual and other human-controlled characters. By leveraging voice recognition technology, VR systems can interpret and respond to user vocal inputs, allowing for intuitive and hands-free interaction. Voice commands can be used to perform actions, trigger events, or navigate within the virtual space, providing a seamless and efficient means of control. Body language interaction complements voice interaction by incorporating gestures, postures, and movements into the VR experience. Through motion capture devices or specialized controllers, users can express themselves physically, mimicking real-world actions and gestures. This enables more nuanced communication and interaction with virtual entities, fostering a sense of presence and embodiment within the VE. By combining voice and body language interaction, users can engage in natural and multimodal communication, simulating real-life social interactions [5]. This integration opens up possibilities for realistic simulations, training scenarios, and collaborative experiences. Users can convey emotions, demonstrate intentions, and establish non-verbal cues, enhancing the sense of realism and social presence in VR. However, limited research has been conducted to measure the impact of avatar representation from psycho-sociological and perceptual perspectives in XR applications [10].

On the basis of these premises, it is easy to understand how the topic of interaction in VR simulations is still considered a hot research theme, characterized by lingering open problems and still requiring new ideas and further rigid investigations.

In summary, interaction in VR encompasses locomotion, manipulation of virtual objects, simulation of real-world interactions, and the representation of users and others within the virtual environment. By addressing these aspects, researchers and developers can create immersive and realistic VR experiences that provide a compelling and engaging user interaction.

The contingency of the state of research and the rapid advancements in VR technology underscore the importance of investigating interaction in VR simulations. As the field continues to evolve, it is crucial to address the gaps in our understanding of effective interaction techniques. By delving into this area, the aim of this thesis is to contribute to the development of novel interaction paradigms that leverage the full potential of VR while overcoming its current limitations.

The outcomes of this research can have profound implications across various domains. In education, immersive VR interactions can enhance learning outcomes by enabling students to engage in realistic simulations and virtual laboratories. In healthcare, natural and intuitive interactions can facilitate training, rehabilitation, and telemedicine applications. Furthermore, the entertainment and gaming industries can benefit from more accessible and engaging VR experiences, leading to new forms of storytelling and interactive gameplay.

The motivations behind this Ph.D. thesis stem from personal experiences and the state of research at an international level. Through this research, the aspiration is to contribute to the advancement of interaction techniques in VR, ultimately enabling more immersive, intuitive, and transformative virtual experiences. By addressing the existing challenges and exploring emerging possibilities, this work has the potential to shape the future of VR applications, benefiting various sectors and positively impacting the lives of individuals worldwide.

1.2 Thesis Structure

The doctoral thesis is divided into four chapters. While the first chapter serves as an introduction, the two central chapters address a specific aspect of the topic investigated during the Ph.D. program. Each of these two chapters is then split in sections. In particular, the first section of each chapter provides an overview of the context in which the research is situated, whereas the second chapter delves into the literature review related to the topic. The study methodology is described in the third section. The fourth section, in turn, presents the obtained results. A considerations and remarks section is then used to summarize the main findings.

Chapter 2 focuses on two of the most important elements of Human-Machine Interaction (HMI) in VR, i.e. locomotion and haptic simulation.

Locomotion is a major open problem for VR. The number of proposed devices and techniques for moving around inside Virtual Environments (VEs) is constantly growing, but these solutions are usually characterized by very different strengths and weaknesses. Hence, the choice of the most suitable technique for a given use case is not straightforward. In this context, a novel evaluation testbed for locomotion techniques and devices in VR is presented, with the aim to provide an effective way to compare different alternatives and identify the most suitable technique for a given set of requirements.

Similarly, also haptics represents a hot and widely studied topic in the field of HMI in VR. The standard interfaces commonly used to interact with virtual objects are the handheld VR controllers, which lacks sophisticated ways to provide a realistic haptic feedback [11]. However, the number and the variety of haptic solutions is constantly growing [3]. In this context, two investigations on haptics were carried out. In the former, the employment of Passive Haptic (PH) interfaces an interactive training simulation in VR (targeting forest firefighting) is compared with a traditional training approach. In the latter, focused on the use of configurations of Active Haptic (AH) and PH technologies, the trade-off between performance (in terms of UX) and complexity of two haptic solutions based on consumer devices are evaluated against a scenario regarding the simulation of an active electromechanical hand tool (a screwdriver).

Chapter 3 shifts the investigation onto two important HMI and HHI means, i.e. voice [12] and body language [13], which play a fundamental role for supporting novel VR interaction paradigms. The use of speech is firstly studied as input to support hands-free navigation in the context of a training-oriented simulation scenario. Then, the combined use of voice and body language is studied in the context of two different use cases. The first study investigates the effects of avatars' body representation in a multi-user simulation in the field of emergency training. In the second study, different techniques for conveying emotions in shared VEs are evaluated in the context of a social VR-oriented scenario.

Finally, Chapter 4 discusses the conclusions of the research and the possible future directions for development.

Chapter 2

Locomotion and Haptic Interfaces

The work described in this chapter has been formerly published in [14]–[16].

The use of VR has become increasingly popular in a wide range of applications, including gaming, education, and therapy. One of the challenges in VR is to provide users with a natural and immersive experience, which requires addressing several issues related to locomotion and interaction with virtual objects. This chapter focuses on these two topics, which are crucial aspects for enabling users to move around and manipulate virtual elements in a seamless and intuitive way.

In VR environments, locomotion is a frequently performed task [8]. While real walking can provide a natural and intuitive way to navigate in such environments [17]–[19], its use is often limited by the size of the tracked area in the VR system (typically the size of a room) or requires expensive technologies to cover larger settings [20]. Various approaches have been proposed to enable effective locomotion in VR, each with different hardware requirements, costs, and levels of usability and performance. According to the taxonomy defined by Templeman et al. [21], locomotion techniques can be broadly categorized as “magical” or “mundane” (Figure 2.1). Magical techniques enable users to move in ways that are not feasible in the real world. The most common technique of this type is *teleportation* (or simply *teleport*) [22], whereas other metaphors (like *world-in-miniature* [23] and *hand-based manipulations* of the VE [24]) have limited diffusion. On the other hand, mundane techniques rely on real-world metaphors and can be either vehicle-centric or body-centric. Vehicle-centric techniques refer to the methods that utilize virtual vehicles [25]–[27], while body-centric techniques aim to recreate physical movements such as walking, running, or swimming[28]. Body-centric techniques can be further classified based on their approach, which may involve *repositioning*, *proxy gestures*, or *redirected walking* [29].

Identifying the best solutions for specific application domains can be regarded

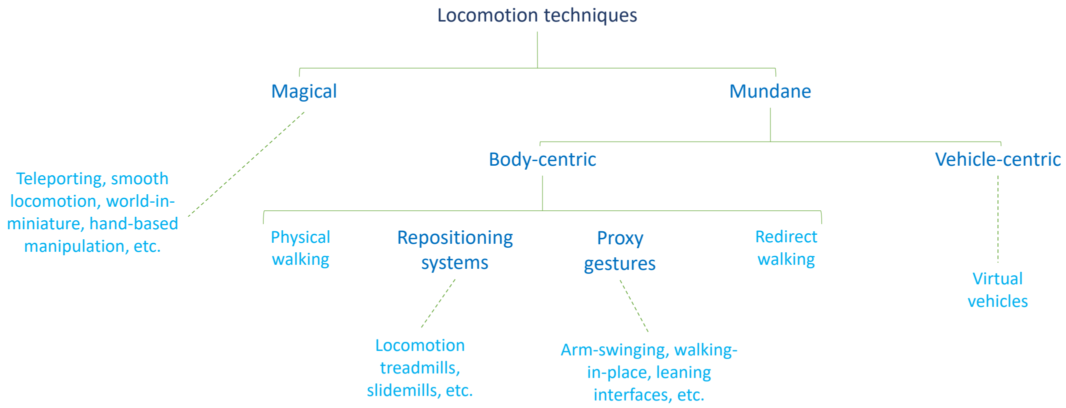


Figure 2.1: Various types of VR locomotion techniques identified by Templeman et al. [21].

as a challenging task due to the absence of a standardized methodology for evaluating and comparing these approaches. To tackle this problem, a new evaluation testbed has been developed in the context of the research reported in this doctoral thesis. The testbed draws on the results of several studies reported in existing literature, and its goal is to enable a thorough examination of the design possibilities of locomotion methods in VR. Along with the testbed, an experimental protocol supporting the collection of objective and subjective measures is presented, paired with a scoring system to rank locomotion approaches based on a weighted set of requirements. An example of use of the testbed is demonstrated by selecting the optimal technique for a famous VR application.

Locomotion techniques are crucial in VR experiences as they allow users to navigate and explore VEs. However, the sense of presence and realism in VR can be further enhanced by incorporating haptic feedback and enabling users to have tactile sensations during hand interactions. Hand interaction in VR refers to the user's ability to manipulate and interact with virtual objects using their hands. By incorporating haptic feedback into hand interfaces, users can not only see and hear the VE but also feel and touch it. This integration of haptics in hand interactions introduces a heightened degree of realism and interactivity, enhancing the overall VR experience. Users can feel the texture, shape, and weight of virtual objects, allowing for more natural and intuitive interactions. However, although high-quality visual and audio feedback are commonly available in VR experiences, the widespread availability of high-quality haptic devices, especially at the consumer level, is still limited [3]. To accommodate the wide range of haptic stimuli, it may be necessary to take advantage of multiple commercial off-the-shelf devices, to combine different functionalities into a single solution, which can help address any missing features in more advanced devices.

The human somatosensory system is capable of perceiving haptic stimuli that can be broadly categorized into two main types: tactile and kinesthetic [3], [30]. Tactile stimuli refer to the perception of 2D form, local shape (e.g., curvature), textures, light pressure, and vibrations detected by the skin. In contrast, kinesthetic stimuli are associated with the identification of 3D shape and the perception of weight, firm pressure, proprioceptive features, forces, and movements detected at the level of joints, tendons, and muscles [30], [31].

Existing haptic devices are designed to provide tactile and kinesthetic stimuli in two ways: active and passive (Figure 2.2). Active Haptic (AH) devices use motors, actuators, and rumblers to actively apply forces to the user, while Passive Haptic (PH) devices apply resistance to the user's motion solely through the mechanical properties of materials or friction/braking mechanisms [32], without actively exerting forces on the user. Therefore, haptic devices are commonly characterized depending on the kind of feedback they are capable of delivering: Active Tactile Feedback (ATF), Active Kinesthetic Feedback (AKF), Passive Tactile Feedback (PTF), and Passive Kinesthetic Feedback (PKF) [30], [33]. Additionally, there is the so-called pseudo-haptic feedback, which is a technique capable of simulating haptic sensations without the need for specific devices, by using only visual feedback and sensory illusion [34]–[37].

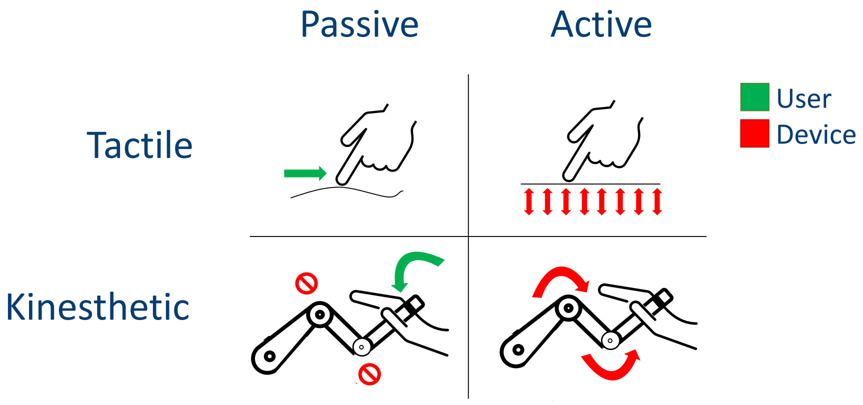


Figure 2.2: Classification of haptic devices based on the type of feedback provided.

In this context, a first study was conducted to investigate advantages that may arise from utilizing VR and PH interfaces when applied to real emergency training situations. PH are physical props, often low-fidelity in nature, that can be used in combination with the visual information provided by a VE to enhance user feedback, conveying information to users via their physical characteristics, such as weight, shape, and other measurable attributes [38], [39]. To this purpose, a PH-based VR Training Scenario (VRTS) enhanced with a real-time fire spreading logic was implemented in partnership with the firefighting unit of Piedmont Region. The effectiveness of the VRTS was assessed as a supplementary method to conventional video-based training. To conduct the evaluation, a user study was carried out with trainees of an existing training course. The study divided the trainees into two groups; one group received video-based lessons from the course, whereas the other group underwent practical training with the developed VRTS. Exexperimental results indicate that incorporating the VR training with PH interfaces resulted in enhanced procedural learning, motivation, and perceived quality of the overall learning experience.

Moving the investigation from PH to AH devices, a second study was built upon the results of a previous work that examined the effect of utilizing two types of haptic setups, one leveraging consumer-level VR gloves and the other exploiting user-prepared props. The selected use case was the simulation an active electromechanical tool, specifically an Electric Screwdriver (ES) [40]. The prior investigation concluded that the integration of vibrotactile gloves and a personalized mockup of the screwdriver was the most effective solution in various aspects. Nonetheless, the previous study did not separate or identify the specific contribution of each individual component of the setup towards the overall User eXperience (UX). Hence, this new study aimed to conduct a detailed analysis of the reference configuration by recognizing simplified configurations that could be achieved using elements composing the original setup. The performance-sophistication trade-off of these configurations was then assessed through a further user study.

2.1 Evaluation Testbed for Locomotion

Although prior works showed that the physical walking is perceived as the most natural form of locomotion [17]–[19], this particular technique cannot be employed in case of VEs surpassing the physical boundaries of the user's surrounding space. To overcome this limitation, a wide number of literature works investigated the possibility to design *stationary* locomotion techniques, which allows 3D movement in the VE regardless of the free space around the user [41]. These alternatives differ in terms of their hardware requirements, performance levels, and costs. In addition, determining the most suitable method for a particular application scenario can be challenging. The literature reveals a significant number of studies that use unique evaluation methods to demonstrate the appropriateness of novel locomotion techniques [6], [42]–[44]. Hence, researchers do not have a standard methodology to investigate this issue.

To this purpose, a comprehensive testbed that can be used to compare different VR locomotion techniques was developed in this research (Figure 2.3) . The testbed, developed with Unity (2018.4 for the OpenVR version, and 2020.3 for the OpenXR one), is organized in two parts. Firstly, a *methodology* to enable the experimental examination of various techniques from multiple Point Of Views (POVs) is presented. The proposed methodology is based on prior research and gathers both objective and subjective measures of user performance and experience while performing various locomotion-related tasks in different scenarios. Secondly, a *scoring system* is introduced, which merges the acquired empirical data with a set of predefined criteria and evaluates the selected locomotion methods, ranking them based on their suitability for specific use cases. Finally, an example is presented to showcase how the testbed can be utilized to compare various common techniques, based on the requirements of a famous VR application. The testbed, labeled Locomotion Evaluation Testbed VR (LET-VR), is available as open-source [45] and can be expanded by other researchers.

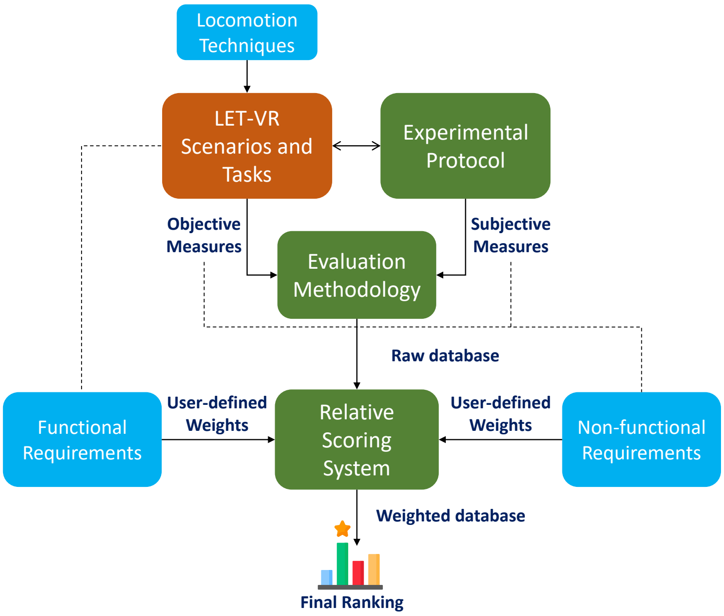


Figure 2.3: Architecture of Locomotion Evaluation Testbed VR (LET-VR).

2.1.1 Background

VR has become widely accessible to different users thanks to recent technological advancements, but studies comparing and analyzing various locomotion interfaces are not new in literature. For instance, over two decades ago, Bowman et al. [6]

proposed a framework to compare different techniques for locomotion in immersive VEs. However, the framework was limited to head tracking and interaction through a spatial input device, such as a 3D mouse, because of the limited technology available at that time. The primary focus was on controlling the direction and velocity of both absolute and relative displacements.

Recent research has evaluated the naturalness and effectiveness of various locomotion interfaces in VR. In Nilsson et al. [42], participants were requested to use four different interfaces, i.e., keyboard, Arm-Swinging (AS), Walking-In-Place (WIP), and hip movement, which involves swinging one's hips to move in the VE, and both the sense of presence and the so called Unintended Positional Drift (UPD) were measured. A different study evaluated the spatial orientation and distance estimation ability of individuals using both AS and WIP techniques and compared them to those achieved during real walking. However, these studies did not include repositioning systems, which is another major approach to VR locomotion mentioned previously.

The previous studies discussed in the earlier sections focused on evaluating various body-centric locomotion techniques in VR. However, the other major approach to locomotion in VR, which is repositioning systems, was not taken into consideration by them. In response to this gap, during the PhD period two studies were carried out to compare a passive repositioning system (a SlideMill, SM) with AS and WIP in terms of usability and cybersickness [46], [47]. SM, unlike VR TreadMill (TM), have a static surface under the user's feet, which can make the interface feel less natural and immersive compared to TMs [48]. In the study, participants were required to complete a challenging task in a realistic VR environment that simulated an emergency situation. The researchers collected both objective and subjective measurements. Unfortunately, since the experiment involved a single task with multiple actions, authors could not isolate the specific contribution of the locomotion technique to the success of the task and individual operations.

A comparable outcome was obtained by Nilsson et al. [49]. The goal was to compare three different WIP interfaces based on marching, wiping, and tapping gestures regarding naturalness, sense of presence, and UPD. In the evaluation, participants were asked to follow a given path for a specific duration of time. However, since the task was very simple, authors were unable to observe any differences in terms usability or performance among the interfaces. As a result, they suggested that more complex tasks, such as object avoidance, movement precision, and promptness in starting and stopping movement, should be used in future evaluations.

In Whitton et al. [43], a precision task was employed to assess the impact of the locomotion technique design on performance. The task involved approaching to specific targets on walls as quickly as possible without hitting them. The study compared the use of natural walking, WIP, and a JoyStick-based (JS) interface with three different visual conditions: a HMD in a computer-generated environment, unrestricted natural vision in a real environment, and field of view-restricted natural vision in a real environment. Authors of the study created five experimental setups to investigate the effect of interface design on performance. They measured the final distance to the target and a time-to-collision metric, which was calculated by dividing the target-user distance by the user's velocity.

The above studies have primarily focused on evaluating different locomotion techniques in VE. However, as mentioned before, it is worth noting that in certain applications, such as video games, locomotion might not be the primary objective, but rather a secondary task that is performed in conjunction with other main tasks, such as locating and interacting with virtual elements.

For example, in Pai and Kunze [50], a user a user study was carried out to assess the efficacy of AS in comparison to WIP for a task involving grasping. The study focused on aspects like immersion, motion sickness, workload, and physical demand. The participants were required to navigate through six different paths while holding a virtual object that had to be released into a basket at the end of each path. The work in Ferracani et al. [7] presented a framework to assess naturalness and effectiveness of four locomotion interfaces including WIP, AS, Tap (a gesture that enables users to move in the desired direction by tapping their index finger) and Push (another gesture where the user closes and opens their hand while dragging it, resembling the motion of moving a lever). Users were instructed to navigate along predetermined paths while avoiding obstacles and interacting with virtual objects by moving them within the VR during the tests.

In Lapointe and Savard [51], the authors conducted an experiment to compare three different bimanual locomotion interfaces for desktop virtual walkthroughs. These techniques relied on a JS with a varying number of Degrees-Of-Freedoms (DOFs) to control the user's position in the VE with their dominant hand, along with a classical mouse to control gaze with the other hand. In the experiment, the participants completed a primed search task where they had to reach a target as quickly as possible, with the target position already known to them. Following the experiment, the participants were asked to complete a questionnaire to assess the usability, fatigue level, accuracy, and speed of the three interfaces.

In Loup and Loup-Escande [52], users were asked to travel long distances and search for specific objects in a realistic VE to compare the effectiveness of the teleportation technique and AS in terms of efficacy, cybersickness, UX, and cognitive load. While the previous works mainly focused on the performance of the locomotion interface and the user's perceived experience, some studies analyzed the impact of the locomotion method on psychological aspects. For example, in Schuemie et al. [44], both objective and subjective, were defined to examine how locomotion techniques affect VEs intended for the treatment of phobias, specifically the acrophobia (fear of heights). In Suma et al. [53], two user studies were conducted to examine the impact of four different locomotion techniques on users' ability to gather and recall information in a complex maze with objects scattered around. The locomotion methods included real walking in both a real and virtual maze wearing a HMD, moving-where-looking (using a handheld trigger to move in the gaze direction), and moving where-pointing (direction given by a tracker mounted on a handheld device).

As seen, various approaches have been developed to evaluate and compare VR locomotion interfaces, each focusing on highlighting specific features of the interface being studied. These experiments typically involved users performing different tasks with different degrees of complexity and evaluated using a diverse range of objective and subjective measures. Building upon these findings and utilizing existing literature, the aim was to propose a comprehensive evaluation testbed that could be used to study VR locomotion interfaces across a wide range of tasks commonly found in VR applications and games, from multiple perspectives. To demonstrate its capabilities and evaluate its effectiveness, the testbed was exploited to study three consumer-ready gesture-based interfaces that were chosen for their expected high level of UX in terms of naturalness, intuitiveness, and comfort.

2.1.2 Materials and Methods

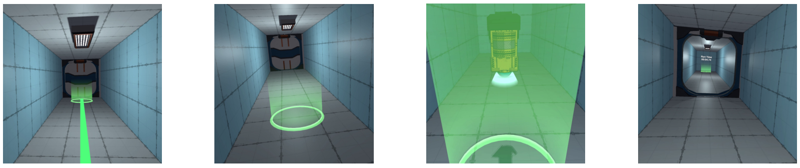
The investigation began with an examination of previous experimental studies in the literature aimed at identifying a collection of essential and frequently encountered locomotion tasks in VR. Additional tasks that are less frequently considered in the literature but are common in typical VR applications were also included. These tasks were considered as *Functional Requirements* (FRs), as they represent the functionalities that a VR locomotion technique should be able to provide. The tasks were grouped into five scenarios based on their shared features, and their execution was evaluated using a set of metrics, which were either objective or subjective. These metrics referred to general characteristics that a VR locomotion technique should offer to the users of a given VR application and were referred to as *Non-functional Requirements* (NRs). The FRs and NRs are detailed in the following, including the works they were derived from if applicable. Guidelines for conducting the experiments will also be given.

Tasks and Scenarios

The five essential aspects that an efficient VR locomotion technique should support include straight movements, direction control, decoupled movements, agility, and object interaction, each addressed by the devised scenarios. Figure 2.4 displays screenshots of the required operations (supported functionalities) in the scenarios.

The first scenario (*Straight movements*, or *S1*) is intended to evaluate a particular technique in the simplest locomotion situations where there are no directional changes required. Consequently, it concentrates on various tasks that one may encounter while moving in a straight line, such as walking casually, halting at specific locations, following a given target object, and running quickly for a short period

🟥 *S1. Straight movements*



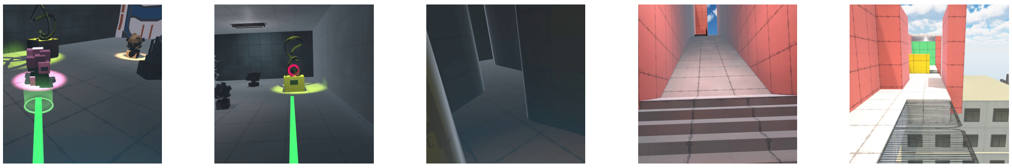
(a) 🟧 *S1.T1. Straight line walking*

(b) 🟧 *S1.T2. Over/Under-shooting*

(c) 🟥 *S1.T3. Chasing*

(d) 🟥 *S1.T4. Sprinting*

🟩 *S2. Direction control*



(e) 🟩 *S2.T1. Multi-straight line walking*

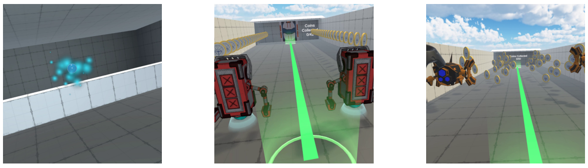
(f) 🟩 *S2.T2. Backward walking*

(g) 🟩 *S2.T3. Curved walking*

(h) 🟩 *S2.T4. Stairs & ramps*

(i) 🟩 *S2.T5. Fear*

🟦 *S3. Decoupled movements*

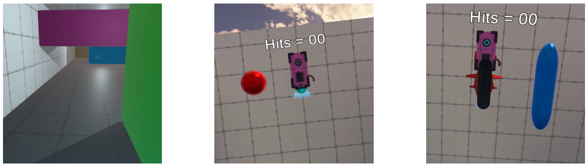


(j) 🟦 *S3.T1. Decoupled gaze*

(k) 🟦 *S3.T2. Stretched-out hands*

(l) 🟦 *S3.T3. Decoupled hands*

◼ *S4. Agility*

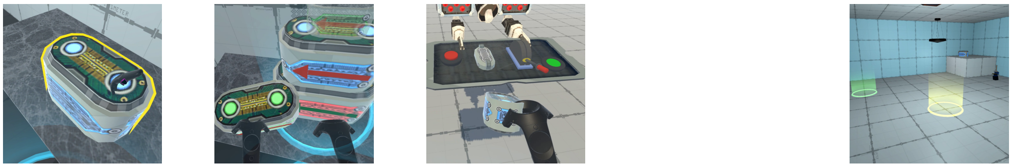


(m) ◻ *S4.T1. Dynamic agility*

(n) ◼ *S4.T2. Stationary agility*

(o) ◼ *S4.T3. Evasion*

🟪 *S5. Interaction with objects*



(p) 🟪 *S5.T1. Grabbing*

(q) 🟪 *S5.T2. Manipulation*

(r) 🟪 *S5.T3. Interaction in motion*

(s) *Training*

Figure 2.4: LET-VR's tasks (grouped into scenarios) (a–r) and training scene (s).

of time. Similar tasks are often present in VR applications where users embody a character, and they require a certain degree of physicality (e.g., [46], [47]). The first task in S1, called *Straight line walking* (*S1.T1*) and depicted in Figure 1a, requires the user to follow a green path on the floor and reach a destination in front of them (similar to Nabiyouni et al. [54]). To successfully complete Task *S1.T2*, *Over/Under-shooting*, the user must stop at three green circular areas with decreasing radius, with the next area only appearing after reaching the previous one. This task is a simplified variation of the one introduced by Whitton et al. [43]. In Task *S1.T3*, called *Chasing*, the user must follow a moving object (i.e., a droid) while remaining inside a green cylinder-shaped volume behind it. As the droid changes speed periodically, this task assesses the user's ability to adapt. Finally, in the *Sprinting* task (*S1.T4*), the user must cover a long path as quickly as possible, which involves both reaching a stationary target at distance and running. Such tasks are frequently encountered in VR games (e.g., [55]).

The second scenario (*Direction control*, or *S2*) is aimed at assessing the effectiveness of locomotion techniques in performing tasks that involve changing directions in various ways. These changes may be required by the task goals or the simulated environment's constraints. The tasks devised for this scenario involve changing direction while moving forward, backward, along curved paths, uphill under different conditions, and while facing obstacles such as a dangerous chasm. This scenario encompasses tasks that may be present in simulations of natural environments, such as forests or mountains, as well as city environments with multi-level buildings that have obstructions and staircases. The first one, called *Multi-straight line walking* (*S2.T1*), requires the user to follow a broken line-shaped path. In the Nabiyouni et al.'s [54] investigation, which inspired this task, participants simulated visiting a museum and moving between artworks. In this version, the user must navigate through the artworks and pause in front of each one until the lights turn off before proceeding to the next objective (like in Paris et al. [56]). The emphasis is on managing the direction rather than the speed, with new sections of the path appearing only when the current artwork is completed. The following task, *Backward walking* (*S2.T2*), asks the user to exit the museum space by walking backwards while maintaining eye contact with the final artwork until they reach the door situated behind it. This task was proposed in Ferracani et al. [7] and is part of a larger set of tasks in the proposed testbed. After completing this task, the user must traverse a tunnel that requires continuous direction adjustments while moving, similar to Nilsson et al. [49]'s (*Curved walking*, *S2.T3*). The fourth task, *Stairs & ramps* (*S2.T4*), examines the management of direction from a different POV. Initially, the user is required to climb both a ramp and a staircase, and finally decide whether to take a staircase or a ramp for the final ascent. This task aims to assess the degree of disruption caused by movements that affect the user's POV, as staircases generate a sussultatory motion and ramps create a more uniform movement that can affect the UX (and the cybersickness), depending on the locomotion paradigm employed. Lastly, the *Fear* task (*S2.T5*) investigates the degree of confidence in controlling direction with the given technique. The user is required to navigate a dangerous area on the roof of a tall building, with the potential risk of falling due to incorrect movements. Interactions with starcases and dangerous zones were previously investigated in Schuemie et al [44].

The purpose of the third scenario (*Decoupled movements*, or *S3*) is to assess the user's ability to decouple the management of walking and its orientation from the movements of the hands and gaze, which is particularly important in situations where the user is required to perform gaze or hand interactions. Enabling unrestricted movement for the user and avoiding interference from the interaction method can help avoid any adverse effects on the UX. The first task (*Decoupled gaze*, *S3.T1*), the user is instructed to walk in a straight line while maintaining focus on a moving object (a floating blue ball) beside them. This ability is commonly used in VR experiences [57] to gather information, maintain focus on a specific location, avoid obstacles, and more. In the second one (*Stretched-out hands*, *S3.T2*), the user needs to control two droids on either side of them to collect two groups of coins rotating in mid-air while walking straight. This task evaluates locomotion methods that use hand-held controllers, which could limit the user's arm movements. The final task (*Decoupled hands*, *S3.T3*) is comparable to the previous one, but requires the user to control two flying drones to gather coins positioned at varying heights and distances from the path. Users must appropriately move their hands in all six DOF to control the two drones and achieve the goal. The capacity to exert accurate control over motion while engaging in other interactions is critical in many VR applications and video games, such as shooting [58] or collecting objects [59].

The focus of the fourth scenario (*Agility*, *S4*) is to evaluate the virtual and physical agility of the locomotion technique under investigation, with a particular focus on agility during unconstrained and constrained movement. These skills are essential when simulating VR experiences that necessitate swift avoidance responses to unexpected threats. The first task (*Dynamic Agility*, *S4.T1*) requires users to traverse a room by steering clear of variously-shaped blocks that move towards them, adjusting their path and ducking when necessary [60]. In the second one (*Stationary Agility*, *S4.T2*), users face a droid that cyclically fires projectiles at them from varying (pseudo-random) directions and speeds. Users must avoid these threats by moving their upper body only, making it possible to gauge the limitations imposed by the investigated technique on stationary motion that do not involve actual locomotion, as required in applications like [61]. The final task (*Evasion*, *S4.T3*) is a variant of the previous one, in which the user must evade a sequence of bigger projectiles that are basically human-sized capsules. Here, the user must employ locomotion to quickly dash to the right or left, as it is no longer feasible to avoid the bullets merely by shifting their position on the spot.

The final scenario (*Interaction with objects*, *S5*) aims to assess how the locomotion technique affects accurate interaction with objects in the VE. This is an essential aspect for applications that simulate procedures requiring the use of virtual tools. The scenario explores pick-transport-place interactions with objects in diverse conditions, such as the number of objects, the presence of obstacles, and occluded sight. It also assesses manipulation and placement of objects requiring accurate motion and repeated displacements, as well as the effect of intricate interactions executed while moving. The first task, called *Grabbing* (*S5.T1*), requires the user to pick up objects and position them in different target locations on the basis of their colors, sizes, and shapes, similarly to what was done in Pai et al. [50] and Ferracani et al. [7]. At the beginning, the task involves moving one object at a time. However, later on, the user must pick up two objects simultaneously, one for each hand. The second part of the task necessitates the user to transport objects while navigating through a maze and evading contact with the walls, like in Suma et al. [53]. In the second task, called *Manipulation* (*S5.T2*), the user must exhibit fine-grained control of locomotion and manipulation to relocate objects that are scattered throughout the environment to specific locations and assembling a tower by accurately stacking them. This kind of interaction is used in applications like [62]. The last task, *Interaction in Motion* (*S5.T3*), is the most challenging one. This task evaluates how the locomotion technique affects the user's ability to interact with objects while in motion. The user is presented with a moving object, i.e. a drone holding a panel, and must perform four interactions in a specific order. This task aims to examine how the user's movement affects their ability to complete essential object interactions while in motion. The moving robot travels at a speed similar to the user's walking speed and constantly moves away from them while maintaining a safe distance. The user must perform a series of interactions on the robot, including pressing buttons, inserting a battery into a socket, pushing a lever, and pressing another button, in a specific order. To accomplish this, the user must adjust their speed accordingly to perform the interactions while staying in proximity to the robot.

In the upcoming discussion, it will be explained that the previously mentioned scenarios were designed to be flexible; this means that, in a given study, researchers can select specific scenarios that address the aspects they are interested in investigating. However, before conducting the actual test, users must be trained to master the particular locomotion technique being studied. For this purpose, a scenario has been created that includes challenges designed to help users learn the technique with the aid of human guidance. Additionally, online videos demonstrating the training process are available online[1]. It is important to note that in order to make LET-VR scenarios adaptable to each user's characteristics and the chosen locomotion technique, certain tasks (e.g., S3.T3, S5.T3) have been tailored to adapt to the

---

[1] LET-VR Videos:
`https://www.dropbox.com/sh/xccjauld6q6ovz4/AACBSSUYE7EHfAQkrM0HS88xa?dl=0`

particular configuration under evaluation. This is achieved by utilizing a calibration file which contains details regarding the user's head height and the maximum distance between their hands. By using this information, LET-VR can account for the user's stature as well as any constraints imposed by the locomotion technique (e.g., user's stance, interface strapping, etc.). The calibration file can be generated either manually or automatically via a process provided in the training scenario. In the automatic method, the user is directed to a calibration spot and asked to hold a specific pose (standing upright with arms extended) for a few seconds before concluding the training.

Measures

LET-VR is intended for evaluating NRs of a VR locomotion technique (listed in Table 2.2) on particular tasks by using either objective or subjective measurements.

To analyze certain NRs, various objective metrics have been defined for each task. Some of these metrics are sourced from previous studies, while others have been specifically designed for particular tasks to address specific aspects that require investigation.

The objective metrics are categorized into three groups, which correspond to the three major NRs that may be relevant for a locomotion technique: *Operation Speed* (*OS*), *Accuracy* (*AC*), and *Error-Proneness* (*EP*). Table 2.1 shows the metrics associated with each category, with the exception of the *StairsChoice* metric, which is marked as *OT* (*Other*) because it represents a subjective measure that is indirectly obtained through an objective measure.

Certain metrics are deemed *elementary* since they directly reflect the corresponding requirement, particularly in tasks where the user has a single objective. For instance, in *S1.T2*, the distance from the target is a direct measure of accuracy.

A recurrent accuracy metric is the path deviation. This metric was inspired by a similar metric used in a study by Whitton et al. [43], but was slightly modified to better suit the evaluation's purpose. The original metric was intended for tasks where the user could not stop or walk backwards, but in some of the tasks in LET-VR, such as S1.T1, remaining still after having deviated too far from the target path is considered an error. Therefore, for user **i**, path deviation was defined as:

$$PathDev_u = \int_0^{CompletionTime_u} PathDev_u(t)dt \tag{2.1}$$

n which $PathDev_u(t)$ represents the degree of deviation from the prescribed trajectory over time, and is calculated as the area between $PathDev_u(t)$ and the temporal dimension. This metric is considered an elementary measure.

Other metrics are referred to as *cumulative* because the task requirement is determined by different independent measures. For example, in task S2.T1, EP is evaluated based on metrics such as the initial angular error, the recall time, and the estimated path length.

In conclusion, metrics are classified as *compound* when the user's objective is intricate and the requirement for the task involves the combination of multiple interdependent measures. For instance, in S3.T3, where the user is required to walk straight while collecting floating coins, the path deviation metric could be used to measure AC related to the first goal, while the percentage of coins collected to measure the second goal. However, the combined metric should not be directly proportional to both measures.

Table 2.1: NRs taken into account for LET-VR:
🟥 objective per-task 🟧 objective per-scenario
🟨 subjective per-scenario ⬜ subjective overall

Accuracy (AC)	Subjective Units of Discomfort [44] (SUD)
Operation Speed (OS)	Self-motion compellingness
Error-Proneness (EP)	Acclimatisation
Physical Effort (PE)	Control
Input sensitivity	Presence
Input responsiveness	Learnability
Ease of use	Intuitiveness
Perceived errors	Comfort
Appropriateness	Enjoyability
Satisfaction	Overall system usability
Mental effort	Cybersickness: Nausea [63]
Perceived Physical Effort (PE)	Cybersickness: Oculomotor [63]
Naturalness	Cybersickness: Disorientation [63]
V/R Phys. str. similarity	Cybersickness: Total [63]

For what it concerns the *compound* metrics, for S2.T2 and S3 AC requirement, a normalized path deviation is initially defined for user u:

$$NormPathDev_u = \frac{PathDev_u}{MaxDistance \cdot CompletionTime_u} \tag{2.2}$$

in which $MaxDistance$ denotes the farthest distance reachable by the user on either side of the path. This quantity is then merged with the proportion of time during the task when the user was visually fixated on the intended target located behind it, as represented by $LookAtTimePerc_u$:

$$ACBkw_u = LookAtTimePerc_u(1 - NormPathDev_u) \tag{2.3}$$

In a similar manner, for task S3.T1, the value $NormPathDev_u$ is integrated with the ratio of time that the user was moving their visual focus from the walking direction, as denoted by $GazeUncRate_u$:

$$ACGazeUnc_u = GazeUncRate_u(1 - NormPathDev_u) \tag{2.4}$$

As for task S3.T2, the normalized path deviation is fused with the fraction of time that the user maintained their arms extended while walking, represented by $StrcTimePerc_u$:

$$ACStrc_u = StrcTimePerc_u(1 - NormPathDev_u) \tag{2.5}$$

Concerning task S3.T3, the quantity $NormPathDev_u$ is multiplied by the percentage of coins gathered by the user, denoted by $CoinsPerc_u$:

$$ACHandsUnc_u = CoinsPerc_u(1 - NormPathDev_u) \tag{2.6}$$

Finally, to quantitatively capture an additional NR that characterizes the physical exertion associated with a particular locomotion interface, user's heart rate is measured before and after each scenario to evaluate changes (Physical Effort metric, PE). This measurement can be obtained through an optical sensor or similar devices. While heart rate variability can pose challenges and noise in measurement, it has been previously demonstrated as a reasonably reliable indicator of the energy expenditure related to a specific physical activity [50], [64]–[66].

The rest of NRs are evaluated subjectively through a questionnaire that consists of three sections. The first section is given before the experiments begin. The second section is administered after each scenario, with minor variations in the questions based on the scenario. These questions measure the "per-scenario" metrics. The third section includes questions that contribute to determining the "overall" metrics. Subjective metrics, like objective metrics, can be either elementary or cumulative. The questions are in the form of statements that users express agreement or disagreement with, and they are scored on a Likert scale from 1 to 5. For symptoms related to cybersickness, the scale ranges from 0 (none) to 3 (severe). The questionnaire, available online[2], is structured as follows:

- The first section begins with questions designed to assess the participant's prior experience utilizing technologies that are associated with the experimental activity. It then includes all the questions from the Simulator Sickness Questionnaire (SSQ) [63], which is used to measure the intensity of cybersickness symptoms before the start of the experience, if present.

- In the second section, after completing each scenario, the participants are asked to assess the Input sensitivity, Input responsiveness, Ease of use, Perception of errors, Appropriateness, and Satisfaction of the technique used based on adapted questions from the VRUSE questionnaire by Kalawsky [67]. Additionally, the participants are requested to rate the level of Mental and Perceived

---

[2] LET-VR Questionnaire: `https://github.com/VRatPolito/LET-VR/tree/master_public/Experimental%20Material/Questionnaire`

Table 2.2: Metrics established to assess the UX with the locomotion technique provided for the specific task, categorized based on the relevant NRs they contribute to. For cumulative metrics, the contributing elements share the same color:
🟧 Operation Speed (OS) 🟨 Accuracy (AC)
🟦 Error-Proneness (EP) 🟩 Other (OT)

Scenario	**Task**	**Per-task metric**
🟥 *S1*	🟧 *T1*	Time to complete the task (s).
		Path deviation, determined by calculating the integral of the distance between the actual path and the intended path, with respect to time (m·s).
		Number of collisions between the user and the walls (#).
	🟧 *T2*	Time to complete the task (s).
		Distance between the user and the center of the target when they come to a stop [43] (m).
		Number of times where the user leaves the intended target destination (#).
	🟥 *T3*	• Percentage of time during which the user was inside the reference area while following the moving robot (%). • Mean distance between the user and the center of the reference area during the task of following the moving robot (m).
		Number of times where walking was interrupted (#).
	🟥 *T4*	Time to complete the task (s).
		Number of collisions between the user and the walls (#).
🟩 *S2*	🟩 *T1*	Time to complete the task (s).
		• Difference in angles between the target and the walking direction after one meter [56] (deg). • Distance covered to reach the target [56] (m). • Time spent in determining the new direction and beginning to walk after the lights were switched off [56] (s).
	🟩 *T2*	Time to complete the task (s).
		Path deviation associated with the percentage of time the user looked at the target in front of them while walking backwards (%).
		Number of times where the gaze was shifted away from the target (#).
	🟩 *T3*	Time to complete the task (s).
		Number of times where walking was interrupted (#).
	🟩 *T4*	Selection made by the user regarding whether to use a ramp (0) or stairs (1) (0/1).
	🟩 *T5*	Time to complete the task (s).
		Path deviation calculated with respect to the rim [44] (m·s).
		Number of times the user fell from the roof (#).
🟦 *S3*	🟦 *T1*	Time to complete the task (s).
		Path deviation associated with the percentage of time during which the user looked at the target located to their side while walking (%).
		Number of times where walking was interrupted (#).
	🟦 *T2*	Time to complete the task (s).
		Path deviation associated with the percentage of time during which the user maintained their arms in a stretched position while walking (%).
		Number of times where walking was interrupted (#).
	🟦 *T3*	Time to complete the task (s).
		Path deviation associated with the percentage of coins that the user collected (%).
		Number of times where walking was interrupted (#).
⬛ *S4*	⬜ *T1*	Time to complete the task (s).
		Number of collisions between the user and the moving elements [7] (#).
	⬛ *T2*	Number of collisions between the user and projectiles (#).
	⬛ *T3*	Number of collisions between the user and projectiles (#).
🟪 *S5*	🟪 *T1*	Time to complete the task (s).
		• Number of times the user let a grabbed objects fell from hands (#). • Number of collisions between the user and the walls (#). • Number of collisions between the grabbed items and the walls [7] (#).
	🟪 *T2*	Time to complete the task (s).
		• During the setup phase, average level of positional and rotational accuracy for items that were placed on the table (%). • During the assembly phase, average level of positional and rotational accuracy for items that were placed on the table (%).
	🟪 *T3*	Time to complete the task (s).
		Percentage of time during which the user stayed in close proximity to the moving robot (%).
		Number of times that the user either performed an interaction in the incorrect order or dropped a grabbed object (%).

PE required to use the technique on the basis of the dimensions outlined in the ISO 9241-400 standard [68]. The degree of naturalness in the walking gesture and the Virtual/Real (V/R) physical strain similarity (i.e., the similarity between the physical strain in the virtual and real worlds) are also evaluated using questions defined in Nilsson et al. [49]. For the Fear task (S2.T5), the participants are asked to rate level of Subjective Units of Discomfort (SUD) as defined in the Acrophobia Questionnaire (AQ) proposed by Schuemie et al. [44].

- The final section contains questions that relate to the overall experience. Specifically, the participants were asked to evaluate the technique based on its Self-motion compellingness (i.e., the level of physical movement perceived during the experience) and Acclimatisation (i.e., how quickly the immersion in the experience made them forget that they were not actually walking), as defined in Nilsson et al. [49]. Additionally, the participants are required to assess the level/sense of Control, Presence, Learnability, Intuitiveness, Comfort, Enjoyability and Overall system usability related to the locomotion technique (again taken from the VRUSE [67]). Finally, the severity of cybersickness symptoms are re-evaluated using the SSQ, to analyze changes in detail throughout the entire duration of the experience.

2.1.3 Experiment

The set of scenarios was carefully designed to group tasks with similar attributes, ensuring consistency in per-scenario metrics and enabling participants to skip irrelevant testing. For instance, they may opt to skip the last scenario if object interaction is not relevant to their interests. While the scenarios can be reordered, it is not recommended as they were intended to challenge users progressively.

Unlike scenarios, skipping a task is not allowed as it would have an impact on the evaluation of both performance and experience. Metrics have been designed to assess the participant's overall experience with the scenario, and some dimensions require sufficient time to evaluate, such as cybersickness, acclimatization, and PE. For this reason, the questionnaire is given at the end of each scenario rather than after each task.

The participants are expected to test the selected scenarios with only one locomotion technique due to the between-subjects design of LET-VR. This enables the reuse of data from previous experiments when new techniques are added to a study, provided the conditions are reasonably similar, such as demographics information. Therefore, the questionnaire does not include any questions that require participants to compare multiple techniques.

A within-subjects approach is also possible. However, devising approaches to tackle the impact of learning effects would be difficult to formulate. Additionally, the test experience takes approximately 80 minutes to complete, including every scenario and the full questionnaire. As such, using a within-subjects design would significantly increase the time required as the number of locomotion paradigms tested grows, potentially causing mental and physical fatigue for participants. Ultimately, a within-subjects design may be viable for a few techniques, which would restrict the number of evaluated scenarios.

Execution

Before beginning the experiment, the participants are given a brief introduction to the purpose of the study, the experimental protocol, and the VR hardware and software to be used, including the locomotion technique. The participant completes the first section of the questionnaire, and the administrator decides whether to go on with the experiment based on the participant's SSQ scores and their suitability for the study's sample.

The participant is then brought into the VR training enviromnent. This scenario is designed to allow the participant to become familiar with the locomotion interface and to receive any necessary information regarding the tasks they will perform. The participants are explained the purpose of the "destination" objects, which represents the visual indicators that they need to reach to initiate, execute, or complete the majority of tasks. They are also introduced to the "blocked" state, which temporarily disables the locomotion technique and is typically activated after concluding a task, but may also be activated at need by the administrator to provide additional instructions.

In the first training phase, the participants are requested to execute a number of actions also required for the other scenarios, such as walking straight, changing direction, adjusting velocity, and walking in reverse. They are also given the opportunity to try out interactions with objects, such as grabbing objects and bringing them from one location to another. The participants perform these operations repeatedly until they feel confident in using the locomotion technique and controlling the various interactions. Then, they set the calibration data for themselves using the guidance provided in the "administrator script" of the LET-VR, either automatically or manually.

Following the completion of training, the participant will commence the testing phase, which involves the following steps:

1. before entering the scenario with the VR headset on, the participant's heart rate is recorded;
2. the scenario begins and the participant is directed to the first task;
3. the participant's movement is restricted while the task is explained by the administrator;
4. the participant completes the task once movement restrictions are lifted;

5. after completing the task, movement restrictions are reinstated;
6. steps 3-5 are repeated for all tasks;
7. once the scenario ends, the heart rate is recorded;
8. the participant is then requested to complete the second section of the questionnaire under the supervision of the administrator;
9. All the previous steps are repeated for each scenario included in the experiment.
10. the participant completes the final section of the questionnaire under the supervision of the administrator.

In the event of any interruptions due to internal factors such as cybersickness or external factors, the activity will be temporarily suspended. The administrator will determine whether to allow the participant to continue the experiment and how to handle any partial data, (discarding it, filling missing values with the M or worst scores, etc).

Relative Scoring System

Following the methodology outlined before, objective and subjective measurements are collected and stored in a raw dataset for multiple participants who have experienced all the locomotion techniques under study. To enable comparison of these techniques, a scoring system utilizing the Weighted Sum Model (WSM) has been developed as part of LET-VR. The WSM is a type of Multi-Criteria Decision Analysis (MCDA) method that assesses a set of alternatives (in this case, locomotion techniques) based on a range of decision criteria (requirements), each of which is assigned a weight to indicate its relative importance.

To ensure that each metric contributes equally to the overall score, they should be expressed in a common unit. Therefore, a normalization is necessary. In this system, normalization is performed by testing the statistical significance of the difference between the Mean (M) value of a metric for one technique and the M values for all the counterparts being studied. The traditional threshold of 5% is used for significance, although the methodology does not prescribe a specific statistical method. If the p-value is lower than 5%, one point is given to the "best" performing technique. This normalization does not take into account the actual magnitude of the significant differences, but it does not require to define lower and/or upper bounds for each metric.

The user of LET-VR must define a *direction* for each metric, which may be either positive or negative. The direction is utilized to ascertain whether a technique is superior to another one, based on whether the metric's M value is greater or smaller than that of the other technique.

For example, if four locomotion paradigms (P1-4) are tested and significant differences are measured between P1 and P2, P1 and P3, as well as P3 and P4 for a given metric, the best technique in pairwise comparisons will be awarded one point. If the metric has a positive direction, the technique with the highest M value will be deemed the best and receive the points, influencing the selection of the optimal technique. Conversely, if the direction is negative, then the one with the lower M value will be selected as the best and awarded the points.

For instance, if the direction of a metric m is negative, and the M values for P1, P2, P3, and P4 are 1.0, 2.0, 2.5, and 1.5, respectively, then the points (P_m) assigned by the system for that metric will be 2 for P1, 0 for P2, 0 for P3, and 1 for P4.

Points for cumulative metric are calculated as:

$$P_m = \frac{1}{\hat{\#}_e} \sum_{i=1}^{\#_e} P_i \tag{2.7}$$

where P_i are the points of the i-th element, $\#_e$ is the number of elements and $\hat{\#}_i$ is the count of elements that exhibit at least one statistically significant difference.

The scoring system used in LET-VR requires the user to assign weights, ranging from 0 to 1, to the set of FRs and NRs detailed before. The FRs represent scenarios and tasks outlined in Figure 2.4, while the NRs align with objective and subjective metrics collected with the LET-VR application, as detailed in Table 2.1.

To determine an overall score of the considered locomotion techniques, and to rank them for a specific application domain, weights are combined with statistically processed data. A two-level combination is employed in this case because two types of requirements exist, instead of directly combining weights with requirements as in a basic WSM implementation.

Weights can be assigned to FRs with either a coarse (per-scenario) or a fine (per-task) granularity. When weights are defined per-scenario, each scenario's weight is applied to all its tasks. On the other hand, if weights are assigned per-task, the per-scenario weights are calculated automatically by averaging the per-task weights.

Weights for NRs are directly assigned to the corresponding requirements. The exception to this rule is Motion sickness, for which weights can be assigned either to major components identified in the SSQ (i.e., Nausea, Oculomotor symptoms, and Disorientation) or to the Total (SSQ) score.

In order to calculate an overall score for a technique, the first step is to multiply each score associated with NRs by its corresponding weight, resulting in a *weighted NR score*. Then, the weighted NR scores for the objective metrics AC, OS, and EP are multiplied by the per-task FR weights, while the scores for the post-scenario questionnaire and the PE metric are multiplied by the per-scenario FR weights. The FR weights have no influence on the overall metrics, which are based on the final section of the questionnaire and reflect the overall experience. The overall score is obtained by summing up the weighted scores for each individual metric.

To calculate the weighted score associated with task T_i in scenario S_j, the weighted scores of the objective metrics are multiplied by the corresponding fine-grained FR weight for T_i in S_j. The formula is:

$$P_{S_j.T_i} = w_{S_j.T_i} \cdot (w_{OS} \cdot P_{OS_{S_j.T_i}} + w_{AC} \cdot P_{AC_{S_j.T_i}} + w_{EP} \cdot P_{EP_{S_j.T_i}}) \tag{2.8}$$

where the weight assigned to the task is represented by $w_{S_j.T_i}$, while w_{OS}, w_{AC} and w_{EP} are the weights assigned to the OS, AC and EP objective metrics. The points computed for the three metrics on the given task are represented by $P_{OS_{S_j.T_i}}$, $P_{AC_{S_j.T_i}}$ and $P_{EP_{S_j.T_i}}$.

To handle the unique nature of the StairsChoice metric for S2.T4, its contribution is calculated as follows:

$$\hat{P}_{S_2.T_4} = w_{S_2.T_4}(w_{ST} \cdot P_{ST} + w_{RA} \cdot P_{RA}) \tag{2.9}$$

where $w_{S_2.T_4}$ represents the weight of task *S2.T4*, and w_{ST} and w_{RA} are special weights that are mutually exclusive, meaning they can only be 0-0, 0-1, or 1-0. P_{ST} and P_{RA} represent the points related to the metric, but with opposite directions (staircases or ramps).

In the same vein, the evaluation of task S2.T5 must be adjusted to take into account the impact of the SUD, the sole subjective metric associated with only one task. As a result, the extra score for this S2.T5 is determined by computing:

$$\hat{P}_{S_2.T_5} = w_{S_2.T_5}(w_{SUD} \cdot P_{SUD}) \tag{2.10}$$

that is essentially the weighted score for the extra metric combined with the weight $w_{S_2.T_5}$.

The calculation of the subjective, per-scenario component of the score for scenario P_j will be performed as follows:

$$P_{s_j} = w_{S_j} \sum_{m=1}^{\#_s} w_{s_m} \cdot P_{s_{m_j}} \tag{2.11}$$

where w_{S_j} represents the weight of the scenario, $\#_s$ is the count of subjective metrics related to the scenario, and w_{s_m} and $S_{s_{m_j}}$ respectively represent the weight of a subjective metric m and the relative score for the given scenario

The calculation of the overall score in terms of points for scenario S_j involves merging the per-scenario subjective contribution obtained in equation (2.11) with the per-task objective contribution obtained in equation (2.8), resulting in:

$$P_{S_j} = P_{s_j} + \sum_{i=1}^{\#_t} P_{S_j.T_i} + w_{PE} \cdot P_{PE_j} \tag{2.12}$$

in which $\#_t$ represents the number of tasks in the scenario, and the formulation also takes into account w_{PE} and P_{PE_j}, which are respectively the weight and points for the PE objective metric.

The points for the overall metrics is computed as:

$$P_o = \sum_{i=1}^{\#_o} w_{o_m} \cdot P_{o_m} \tag{2.13}$$

in which $\#_o$ represents the number of overall metrics, and w_{o_i} and P_{o_m} respectively represent the weight and points of the m-th overall metric.

Using the above formula, the overall points for the specified locomotion technique are determined as follows:

$$P = P_o + \sum_{j=1}^{5} P_{S_j} + \hat{P}_{S_2.T_4} + \hat{P}_{S_2.T_5} \tag{2.14}$$

Using the Testbed

After applying the scoring system to the data in the RDB, the results are stored in the *Weighted Database* (WDB). This database contains computed values for each metric, their M values and significance for a given set of experiments, as well as the overall scores obtained by applying weights.

The WDB has two parts: a *fixed* part and a *variable* part. The fixed part is related to the testing activity and includes the set of techniques and scenarios used in the experiments, as well as any demographic constraints that may have been applied.

The variable part of the WDB needs to be configured by the user to compute overall scores. It includes the weights assigned to the FR/NR metrics, the direction assigned to each metric, and the subset of techniques to be compared. This part can be adjusted depending on the specific study to be performed without the need for additional experimental activity.

The user can adjust the variable part of the WDB without needing to conduct further experimental activity to match the specific study. To find the most suitable locomotion technique for a particular scenario with a set of options, the user can use an existing WDB (if the fixed configuration matches the study requirements) or generate a new one by conducting experiments and applying statistical analysis. The user can adjust the weights of FR/NR metrics based on the specific application scenario, such as selecting the best technique for a VR application where users need to complete assigned tasks quickly, and assign higher weights to FRs like Sprinting and NRs like OS. The overall scores computed by the system would then rank the techniques based on the specific application. If certain techniques are not compatible with the given scenario, they should be removed from the RDB, and a new WDB can be generated by recomputing the weighted scores after conducting statistical analysis. In addition to overall scores, the user could also analyze the scores computed for individual metrics to identify effective aspects of a technique that could be used in designing a new technique.

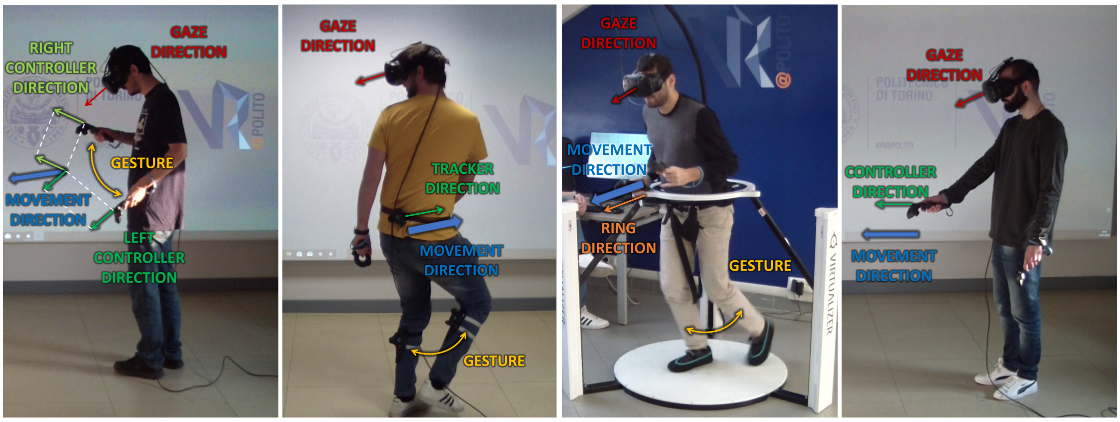
2.1.4 Use Case

Having outlined the features of the testbed, a use case that illustrates the workflow for comparing a range of locomotion techniques is hereafter presented. Initially, several techniques from academic and industrial research that could be used to address space limitations have been examined. Four techniques were ultimately selected for comparison, based on their availability (of implementation details, or ready-to-use packages).

First, the chosen techniques and their respective implementations will be described. Then, the user study that was conducted to gather experimental data will be introduced, and how the scoring system was employed will be explained. Specifically, it will be explained how to generate the WDB and how to use it to evaluate the selected techniques under specific usage conditions for a given VR application.

Evaluated Techniques

As previously mentioned, four locomotion techniques were chosen for the evaluation: AS, WIP, SlideMill (SM), and JS. The experiments were carried out using the HTC VIVE VR system (Figure 2.5).



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

Figure 2.5: Locomotion paradigms evaluated with LET-VR: a) Arm-Swinging (AS), b) Walking-In-Place (WIP), c) SlideMill (SM), and d) JoyStick (JS).

Among the techniques not involving the use of feet, AS was selected because it was found in the experiments conducted in Loup et al. [52] to use the most natural gesture. To generate movement, the user is required to hold a button on the hand controllers (specifically, the grip button in the implementation used), and then swing their arms back and forth to walk or run. The trigger buttons on the hand controllers are used to interact with objects, as is the case with the other techniques. In this study, the publicly available AS implementation by ElectricNightOwl [69] was utilized, in which the movement direction is determined by averaging the orientation of the two controllers (Figure 2.5a). The length of the stride is directly linked to the swinging of the arms. This approach enables the user to separate the direction of walking from their line of sight.

The second technique considered in the study is WIP, which is a widely used method for locomotion that has been proposed in various forms and implementations. To initiate movement, the user must perform a specific gesture with their legs while remaining stationary. The direction of movement can be determined from the user's head (by aligning it with the gaze direction) or from other devices (such as wearable sensors when the gesture is recognized). For this study, the low-latency, continuous-motion WIP variation described in Feasel et al. [70] was selected. According to this approach, movement is generated by directly mapping the space covered by two sensors attached to the user's legs to the user's speed in the VE, while direction is obtained from a third sensor positioned on the user's back. In the implementation used, three Vive Trackers were employed, with two attached to the user's calves via custom 3D-printed supports and one affixed to their back with a belt (Figure 2.5b). This allowed for the technique to be "aligned" with the other methods by adjusting the configuration parameters, and for the head/gaze orientation to be decoupled from the movement direction. The marching gesture was used, which is considered a standard for this technique [49]. A filter was applied to horizontal leg movements to mitigate unwanted motion when the user turns around.

The third technique included in the comparison is SM, which falls under the category of passive repositioning systems for locomotion in VR. The functioning of this technique is illustrated in Figure 2.5c. It employs a low-friction walking surface and a rotating containment ring, which prevent the user from physically displacing. The walking direction is determined by the orientation of the ring. The device is a commercially available solution, the Cyberith Virtualizer, presented in Cakmak and Hager [71]. Although previous research works have compared this technique with others, it has only been done in broad terms [29], [46].

The JS-based technique used in the study was based on the implementation described in Boletsis and Cedergren [72]. To activate movement, the user presses the pad button on either of the hand controllers and adjusts the speed by moving the thumb up or down the touch pad. The upper bound generates maximum speed, which linearly decreases as the thumb moves towards the lower bound (zero speed). The direction is determined by the controller whose touch pad was pressed to initiate motion. To switch from walking to running, the user must press the grip button on the same controller while performing the previously mentioned actions. This sets the speed to a fixed value, which is higher than the maximum speed achievable through modulation. This kind of locomotion achieved with this kind of JS-based techniques is commonly referred to as SM (SlideMill), in contrast to the different behaviour of the usual alternative included in commercial VR games (the *teleport*).

During the experiments, users were required to remain within a predefined working area, and room-scale movements were disabled to study specific locomotion techniques. The working area was indicated in the VE by a semi-transparent cyan cylinder with a 65cm radius around the user's initial position, which was displayed as a brighter circle at the user's feet level. A warning region ranging from 90% to 100% of the working area radius was used to signal when the user unintentionally moved out of the working area. Movement in the VE was disabled once the user left the working area, with the user only able to step back into the working area. This was done to prevent cybersickness and disorientation while allowing the user to crouch and make on-the-spot head and hip movements. To ensure consistency in the experience, the maximum speed for all four techniques was set to 7m/s [73]. For the JS technique, the maximum value was used for running, while the speed for walking could be adjusted between 0 and 3.5 m/s. For the other techniques, the user could adjust the speed between zero and the maximum value.

The method of interacting with objects was made uniform by using the same approach utilized in previous research studies [46], [47]. To manipulate an object, the user was required to touch it using the tip of the hand controller and hold the grab button to attach the object to their hand and move it around. After receiving visual and tactile confirmation of contact with the object, the user can bind the object to their hand by pressing and holding the grab button (trigger) while visual and tactile feedback confirms the contact. A blinking outline indicates the button's activation. The object can be dropped by releasing the trigger button or transferred to the other hand by touching it with the relative controller and pressing the corresponding grab button.

2.1.5 Results and Discussion

Acquisition of Experimental Data

To produce the RDB, an experimental study was conducted involving 48 volunteers from the student and academic community of Politecnico di Torino. The participants included 37 males and 11 females, ranging in age from 19 to 37. Each participant was assigned to one of the four locomotion techniques. Based on the pre-test section of the questionnaire, it was found that the participants had limited experience with VR technology, with only 12.5% being regular users, and 31.25% playing 3D videogames often or very often. Furthermore, 8.33% of the participants had used the assigned locomotion technique before the experiment, 20.83% had some knowledge of it, and 70.83% had never used it. Prior to the experiment, none of the participants reported high symptoms of motion sickness as done in Jaeger and Mourant [74]. The study conducted statistical tests, specifically Kruskal-Wallis and Dunn's post-hoc tests, to determine if there were any significant differences among the groups during the pre-test phase. However, no significant differences were observed. To ensure that all participants were equally prepared for the training, those who were unfamiliar with VR were given some time to become familiar with the technology before beginning the training. They then followed the experimental protocol, completing all scenarios in the default order, and all participants were able to complete the experiment.

Computation of Normalized Scores

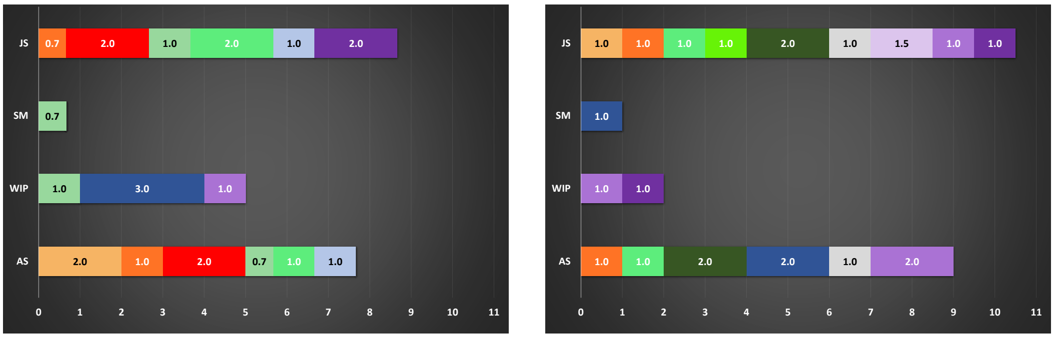
The RDB data underwent processing using the described scoring system. That involved removing any possible outliers using the Z-score, and checking the normality distribution of the data using the Shapiro-Wilk test for each metric. Since participants experienced all the scenarios using only one interface, the collected data was considered independent. Statistical significance of the differences between data sets was calculated using the Kruskal-Wallis test for non-normally distributed data and ANOVA for normally distributed data, followed by Dunn's post-hoc test and Tukey's HDS test for pairwise comparisons, respectively. An Excel spreadsheet was developed to implement the scoring system and generate the WDB for the four techniques. The spreadsheet contains both raw data and weighted scores, and can be accessed online[3]. The contribution of individual metrics for each technique is plotted in Figure 2.6, Figure 2.7 and Figure 2.8, with colors used to indicate the task/scenario where the differences were statistically significant. The ranking of the techniques was obtained assuming all the requirements have the same importance for the application they are used in, and are as follows: JS (54.5pts), AS (53.0pts), SM (23.6pts), and WIP (17.9pts). It is crucial to acknowledge that the scores obtained do not represent absolute values but rather indicate the degree to which a particular technique fulfills the weighted set of requirements in comparison to the other techniques evaluated. Moreover, the results of pairwise comparisons were not included in the figures, which makes it not always possible to determine the superior technique in a given comparison.

Weight Assignment and Final Ranking Generation

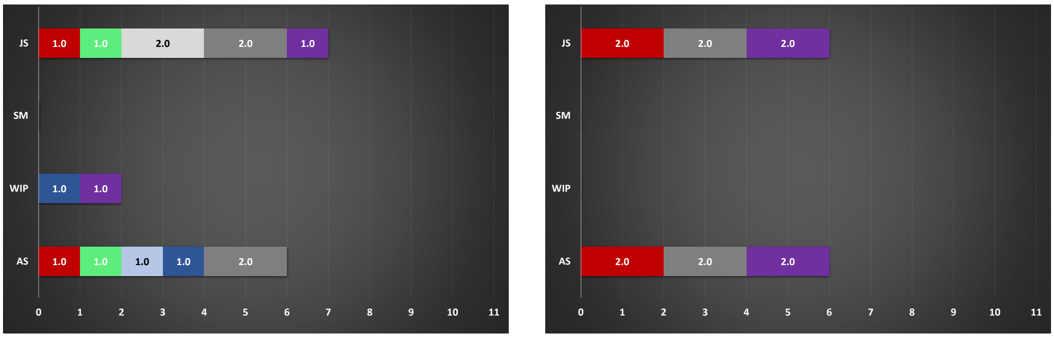
To demonstrate how to utilize the generated WDB to select the technique that maximizes the weighted requirements for a particular scenario, a very popular VR game, *Half-Life: Alyx* [75], which rapidly become one of the best VR games of all times [76] was also considered. The game was chosen due to its popularity, complexity (in terms of FRs and NRs), and because it represents a standard in

---

[3] LET-VR (R)WDB:
`https://github.com/VRatPolito/LocomotionVR/raw/master_public/Experimental%20Material/(R)WDB.xlsx`



(a) Accuracy (b) Operation Speed

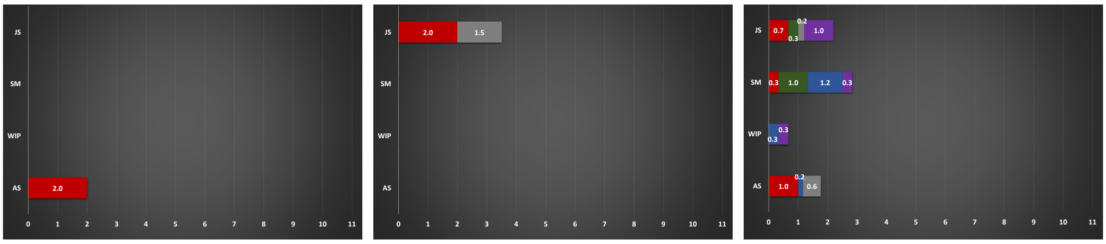


(c) Error-Proneness (d) PE

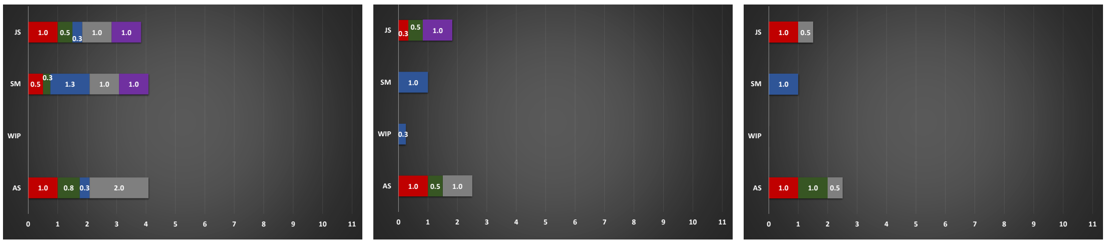
Figure 2.6: Objective metrics scores for the four evaluated techniques (with unitary weights). The reported values are the points obtained in pairwise comparisons, while lack of a value indicates that the technique did not perform better than the other three. The color coding used is the same as that adopted in Figure 2.4 and Table 2.2.

terms of locomotion and interaction in the field of entertainment in VR.

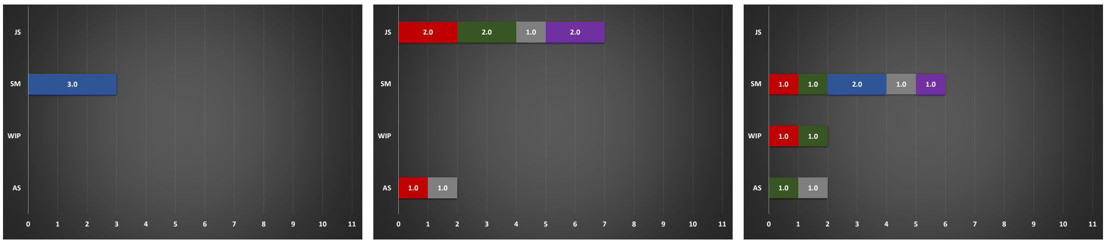
The weights for the FRs were set based on the importance of various requirements in the game. For instance, the game is organized as a set of consecutive levels, each of which is enriched with elements of interest intended to slow down the gameplay (e.g., collectibles, threats) [77]. In this context, the need to run is very limited, and thus the relative weight can be set to a relatively low value ($w_{S1.T3} = 0.3$). On the other hand, the presence of localized hazards (e.g., *barnacles*, *trip mines*) and impervious routes makes the other straight movement-related requirements very relevant ($w_{S1.T1-3} = 1$). For what it concerns direction control, the ability to accurately proceed in zigzag between various covers is fundamental ($w_{S2.T1} = 1$) for combat situations, where you have to dodge projectiles and grenades, or avoid slow-moving threats (e.g. *zombies*). Similarly, due to the various hazardous situations depicted (e.g., walking close to cliffs, climbing on ledges, claustrophobic situations), the impact of the fear related to the specific locomotion paradigm should be minimized ($w_{S2.T5} = 1$). For what it concerns curved walking, backward walking, and



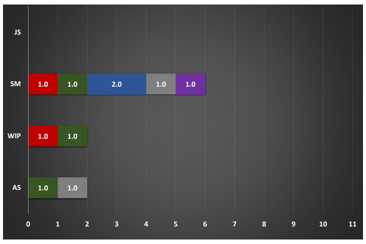
(a) Input sensitivity (b) Input responsiveness (c) Ease of use



(d) Perceived errors (e) Appropriateness (f) Satisfaction



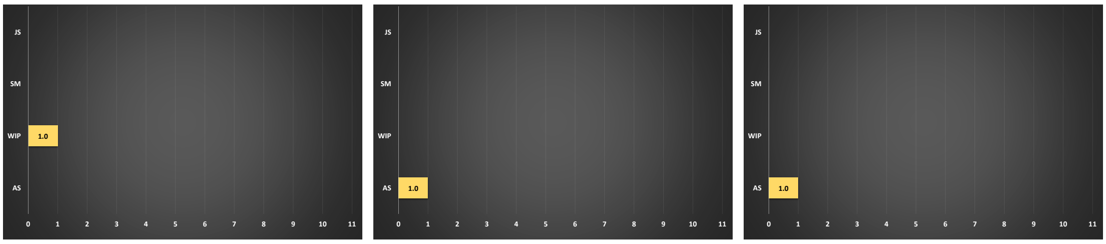
(g) Mental effort (h) Perceived PE (i) Naturalness



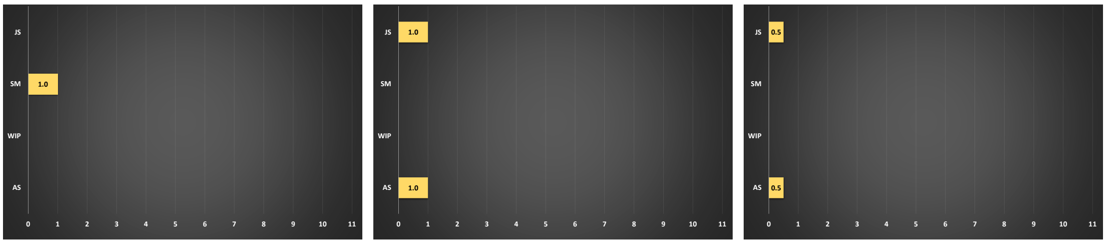
(j) V/R Phys. str. similarity

Figure 2.7: Subjective per-scenario metrics scores for the four evaluated techniques with unitary weights. The reported values are the points obtained in pairwise comparisons, while lack of a value indicates that the technique did not perform better than the other three. The color coding used is the same as that adopted in Figure 2.4 and Table 2.2.

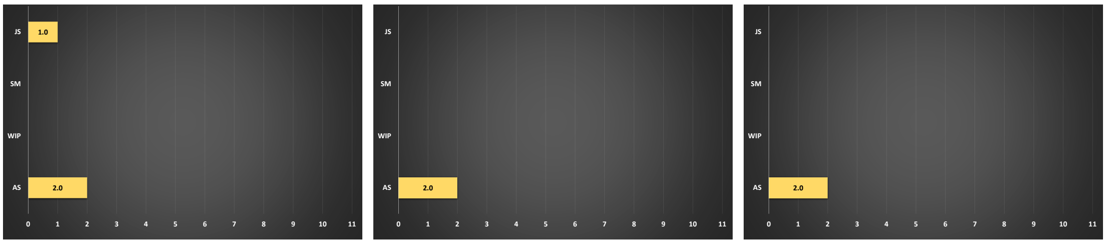
stairs/ramps, the game was specifically designed to avoid these situations, since the experience was tailored to fit with a teleport-based locomotion approach. Thus the relative weights can be set to zero ($w_{S2.T2-4} = 0$). For what it concerns decoupled movements, the possibility to decouple hands from movements is a core functionality ($w_{S3.T3} = 1$), especially considering the peculiar grabbing approach supported



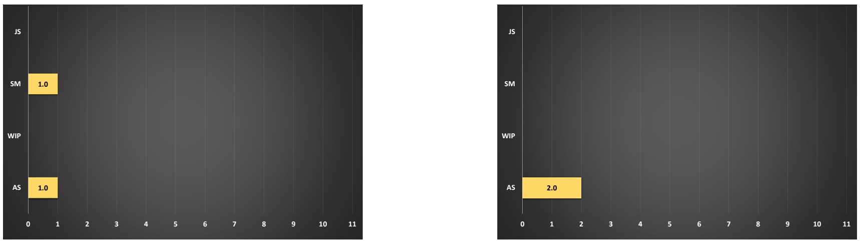
(a) Self-motion compellingness (b) Acclimatisation (c) Control



(d) Presence (e) Learnability (f) Intuitiveness



(g) Overall system usability (h) Nausea (SSQ) (i) Oculomotor (SSQ)



(j) Disorientation (SSQ) (k) Total (SSQ)

Figure 2.8: Subjective overall metrics scores for the four evaluated techniques with unitary weights. The reported values are the points obtained in pairwise comparisons, while lack of a value indicates that the technique did not perform better than the other three.

by the game (the "gravity gloves"), whereas there are no situations in which keeping the arms completely straight is useful or requested ($w_{S3.T2} = 0$). Decoupling gaze from movements is also very important in situations in which you have to aim and shoot at specific targets while moving ($w_{S3.T1} = 1$). Regarding agility, the ability to run away from fast approaching enemies (e.g., *antlions*) is fundamental, as well as to dodge very agile threats (e.g., *headcrabs*), so the relative weights should be maximized ($w_{S4.T1} = 1$ and $w_{S4.T3} = 1$). To a lesser extent ($w_{S4.T2} = 0.8$), also stationary agility can play an useful role when it comes to ducking behind covers or avoiding headcrabs. Finally, due to the particular focus put by developers on hand interaction in every form, all the weights for the fifth scenario should be set to 1 ($w_{S5.T1-3} = 1$).

Defining the set of NRs is more complicated, since they depends on what delelopers wanted to achieve in the first place. Considering the objective ones, OS can be considered as important ($w_{OS} = 0.8$), but less than AC, EP, and EP ($w_{AC,EP,PE} = 1$).

Input responsiveness, ease of use, mental effort, perceived EP and PE, appropriateness, satisfaction, control, acclimatisation, presence, learnability, intuitiveness, comfort, enjoyability, and overall system usability could be all considered very important and assigned a weight of 1. The same can be said for cybersickness. The input sensitivity can be deemed less relevant and assigned a weight of 0.25. Finally, weights for naturalness, V/R physical strain similarity, and self-motion compellingness were all set to 0, considering that the focus put by developers on designing an experience completely accessible via teleportation.

The rankings obtained using the generated WDB with the given FR weights were JS (40.97pts), AS (39pts), SM (10.6pts), and WIP (8.72pts). It is worth noting that the actual implementations of the JS technique in the game differ slightly from those considered in the WDB, which would require creating a new WDB or introducing matching implementations in the game. Furthermore, the considered implementation of the AS (with the grip button as trigger for the motion) would not be employable in Half-Life: Alyx, since the grip button is already used for grasping objects, and there are no other free buttons on a standard controller. An alternative method would be to recognize a specific swinging gesture avoiding the need for a button for triggering the movement, but then they new implementation should be evaluated with the testbed, and results may change (as well as the final ranking).

Limitations

A current limitation of the scoring system is its methodology for accounting for the contribution of different metrics and resolving the MCDA problem. The selected method may result in different rankings when techniques are added or removed from the analysis. Furthermore, the approach used to normalize the contributions of metrics only considers statistical significance, without taking into account the actual magnitude of differences between the M values of a given metric for different techniques. It only indicates whether one technique is better than another, but not by how much (in terms of a specific metric). Additionally, the overall results are influenced by the number of statistical differences found, making it necessary to expand the availability of experimental data to enhance the reliability of the obtained scores.

Future Developments

In the future, the aforementioned issues will be tackled by devising alternative normalization approaches, such as using different significance thresholds and expanding the range of points that can be assigned, among others. Additionally, there are plans to perform further experiments using real walking data in the normalization step. To extend the representativeness and significance of the publicly provided data-set, more data will be collected and tested with these and other techniques. Furthermore, although LET-VR already covers a substantial number of FRs, there are still some missing, such as those related to movement in the third dimension. These requirements, as well as action-specific tasks that focus on interactions typical of specific applications (e.g., first-person shooter games), could be added. Other scenarios supporting the study of jumping, swimming, climbing, and flying configurations may also be considered. Similarly, NRs, such as those relating to psychological and/or cognitive aspects (e.g., spatial awareness, sense of direction, navigation abilities, etc.), could be included. To this end, a study aimed at identifying possibly missing requirements and characterizing the testbed's discrimination capabilities could be conducted.

2.2 Passive Haptic (PH) Interfaces for Interactive Simulations

Over the past decade, VR has become increasingly popular in the fields of training and education, with particular applications in creating effective emergency training experiences [78]–[81]. For what it concerns procedural training, VR has been demonstrated to be more effective than traditional training methods such as printed materials [82] or video-based training [83] in aspects such as knowledge acquisition and retention, as well as usability, trainees' confidence, and self-efficacy.

In the context of procedural training involving the simulation of physical tools, the use of haptic interfaces to enable users to touch and feel virtual objects can enhance their sense of immersion and presence in the VR experience [84]. One of the most interesting fields where this type of approach can be applied is emergency training. Designing VR-based training experiences for first responders encounters a challenge in reproducing the frequent interactions of operators with specific equipment in the VE. One typical approach is to generate digital copies of the tools and allow users to manipulate them through hand controllers that come with commercially available VR systems. Nonetheless, although these virtual reconstructions can achieve a high level of visual realism, they do not possess the physical characteristics of the actual equipment, limiting their accuracy in simulating real-world counterparts [85]. In order to address this problem and improve the simulation of hands-on operations, PH interfaces (or simply, PHs) can be used [38]. By representing the weight, shape, and other physical characteristics of objects [39], PH interfaces, which are physical prototypes of low fidelity, can be merged with the visual data offered by the VE [38], thus improving users' sensory input.

This study reported herewith was aimed to explore the efficacy of using PH-based VR training within a formal training program for first responders. To acquire skills related to firefighting, live-fire training is a highly effective approach as it offers realistic conditions for trainees in a controlled environment [86]. However, this methodology has its drawbacks, such as safety hazards, and is not always feasible and is thus infrequently included in firefighting courses. On the other hand, traditional training approaches, which mainly rely on text and multimedia materials, may not be entirely effective as they may not provide adequate feedback on trainees' individual behavior [87]. Moreover, these methods may fail to engage trainees emotionally, leading to a less effective learning process [88]. These constraints are especially noteworthy when imparting procedural knowledge, which is critical in firefighting and other tasks carried out by first responders [78].

To achieve this objective, a VRTS was created and incorporated into the regular training course provided to volunteers of the forest firefighting unit in the Piedmont Region of Italy[4]. The VRTS can be used to instruct users on the safe use of three firefighting tools (the shovel, McLeod, and firefighting broom) through simulated experiences. To provide a more realistic training environment, physical replicas of the tools were constructed and used in place of VR controllers, enabling trainees to interact with the VE in a natural manner. Additionally, a real-time fire spreading algorithm was developed, which could be influenced by the use of the firefighting tools.

Next, a user study was undertaken to evaluate the effectiveness of the VR that had been enhanced with PH interfaces (which were designed to simulate a practical session to be conducted after a video-based lesson), in comparison with video-based training by itself. The study considers both subjective and objective measures to evaluate the two methods. The subjective measures were collected through standard questionnaires, evaluating trainees' motivation (including factors such as attention, relevance, confidence, and satisfaction) and the fidelity of their learning experience. In contrast, the goal of the objective measures was to measure trainees' conceptual and procedural learning outcomes. This was accomplished by administering a theoretical quiz both before and after the VR experience, as well as conducting the final outdoor practical exam of the standard course. Moreover,

---

[4] Corpo AIB Piemonte: `https://www.corpoaibpiemonte.it/`

a specific questionnaire was used to evaluate the usability of the VRTS.

The hypothesis regarding procedural learning in the study was that the PH-based VR training would introduce a “learning by doing” aspect that is not present in the traditional video-based lesson. This should enhance the trainees’ understanding and retention of the procedures for performing the tasks, compared to relying solely on the video-based training. However, the study did not expect to see a significant improvement in conceptual learning, as the video-based training already provided a detailed explanation of theoretical concepts, which were only partially reiterated in the VR experience.

2.2.1 Background

In recent years, there has been a growing interest in using VR as a tool for training. VR provides an immersive and interactive environment that allows users to experience realistic scenarios and learn through hands-on practice. However, to ensure effective and realistic training, it is essential to provide accurate and realistic tactile feedback. This is where PH devices come into play. By providing users with physical feedback when interacting with virtual objects, these devices bridge the gap between the virtual and physical worlds, allowing for a more immersive and engaging training experience.

The importance of VR training becomes particularly evident in critical and high-stakes scenarios, such as those encountered by first responders. First responders, including firefighters, paramedics, and law enforcement personnel, often face unpredictable and demanding situations where split-second decisions and precise actions are essential. The effectiveness of VR technology in training first responders operators has been extensively investigated in both past and current research [40], [89]–[93]. Many of these studies have focused on training firefighting personnel, which is why it has been chosen as the focus of this study.

As an example, in Engelbrecht et al. [86], the authors conducted a SWOT analysis (Strengths, Weaknesses, Opportunities, and Threats) on the use of immersive VR in the field of training for high-risk situations. They noted that VR has several strengths, including the ability to enhance safety during training and engage trainees, high ecological validity, and cost-effectiveness. Furthermore, VR offers additional advantages such as the ability to record data and create complex and diverse scenarios.

Nonetheless, there are several drawbacks to the technology, such as the restricted accuracy of multi-user interactions, the absence of validation for developed VRTSs from actual first responders, and the technology’s continuing immaturity and related technological barriers.

The authors of Engelbrecht et al. [86] identified several opportunities for the use of VRTSs in firefighter training. One opportunity is the progress in system engineering, which has improved simulation tools for other domains, such as fire propagation models that consider wind, flying embers, fire extinguishing tools, and smoke. These advancements could be applied in future VRTSs for firefighting. Another opportunity is the transfer of findings from other training contexts, such as military, medical, and industrial domains, which could provide valuable insights for the firefighting domain. Additionally, VR technology could benefit from continuous advancements in sensory stimulation fields, such as visual, haptic, and olfactory stimulation, to improve physical fidelity and replicate the sensory inputs necessary for firefighting skills. Furthermore, VR could enhance mental preparedness by providing realistic experiences that can be repeated several times, increasing resilience against adverse effects, as experiencing a real emergency scenario may be traumatic.

The authors of the study found some potential risks during their analysis. An example of such a risk is the uncertainty regarding the transfer of skills, as the heightened complexity of utilizing a VRTS could impede the efficacy of the training experience, potentially preventing it from reaching the requisite level of transfer needed to replace traditional learning methods. There are also potential concerns regarding habituation and engagement. Habituation could result in a gradual numbing of the trainee's response to stimuli from the VRTS, which could lead to suboptimal training outcomes and overconfidence in real-world scenarios. While engagement is generally a desirable quality, it can also be a source of concern. In the case of VRTSs, there may be a risk of overemphasizing engagement through the use of various elements such as rewards. However, the actual experience of firefighting may not be as engaging, leading trainees to prioritize these added elements over the essential firefighting tasks. Additionally, there may be a risk of diminishing returns from the training due to overuse of VR. While VRTSs cannot entirely replace traditional training methods, they can serve as a valuable supplement. However, there is a risk that trainers may be inclined to rely excessively on VRTSs due to their lower costs and easier management, which may lead to a reduction in the quality of overall training outcomes. Therefore, it is crucial to balance the use of VR with traditional training routines, such as live-fire exercises, to avoid over-reliance on VR training.

Numerous studies have investigated the use of VRTSs for firefighter training due to the importance of this field and the numerous issues and opportunities it presents. For instance, in Querrec et al. [94], the authors proposed a multi-agent-based VRTS called *SécuRéVi* for training firefighting officers. The authors presented a tool that enables officers to guide firefighting teams in handling intricate incidents that are challenging to simulate in real-world training drills, such as a factory gas leak or an explosion. Additionally, they presented a typical pedagogical scenario that outlines the roles and responsibilities of each participant, including the designer, instructor, and trainee.

In Cha et al. [95], the authors utilized a VRTS that incorporated a Fire Dynamics Simulation (FDS) to simulate evacuation and rescue operations in a firefighting scenario in a road tunnel. To develop a VRTS based on computational fluid dynamics data, they suggested several data conversion methods and a framework for real-time processing. Despite the framework ability to process FDS data in real-time, the simulation itself was still non-interactive as a result of the extensive processing time necessary. As a result, the firefighting tasks featured in the study did not include extinguishing fires or any other activities that might disrupt the simulation of physical phenomena.

The limitations observed in previous works were addressed to some extent in a previous study reported in Calandra et al. [96], where a VR-based training simulator for emergency operations was developed as multi-role, multi-user, and multi-technology application. The focus of the simulation was a road tunnel fire based on a real incident that occurred in the Frèjus tunnel. To maximize the effectiveness of training, the authors utilized a range of different technologies and techniques. While the simulator did leverage FDS data, it was limited to the realistic generation of smoke visuals, whereas the fire simulation relied on a plausible but non-physically precise spreading algorithm. This allowed for direct engagement with the fire during dynamic extinguishing operations.

In Çakiroğlu and Gökoğlu [97], a VRTS was presented to provide primary school students with basic fire safety training. The training consisted of multiple phases, beginning with a VR-based behavioral skill training phase, in which a virtual firefighter avatar in a VE taught the students fire safety concepts. In the subsequent in-situ training and assessment phases (in a VR-based fire safety training setting), students were transported to different locations in another VE and instructed to execute a range of fire safety-related activities. Initially, they had to apply the concepts they had learned, and then their performance was monitored and evaluated. In the final phase, an in-situ assessment was conducted in an actual fire scenario at a local fire department, where further evaluation was carried out in a controlled environment. The findings of the study indicated that VR significantly improved training effectiveness, and most students were able to apply their newly acquired behavioral skills in real-world situations.

In the study presented in Buttussi and Chittaro [82], a comparison was made between three different approaches for procedural safety training, specifically for door opening procedures in different aircrafts. The study examined three different approaches for training on safety procedures, namely immersive head-worn VR with a HMD, non-immersive handheld VR employing a smartphone, and conventional printed safety cards. The efficacy of each approach was evaluated based on several factors such as performance, knowledge acquisition and retention, confidence, presence, and engagement. The results showed that the immersive VR was significantly more effective than the printed material, and it provided a greater sense of presence compared to the smartphone-based system. The VR experience was also the most successful approach in terms of trainees' engagement and satisfaction.

The present study also focused on utilizing PH interfaces, specifically haptic interfaces, to enhance the trainees' experience and performance.

A relevant example is the work of Suhail et al. [85], who employed a physical prop to simulate interactions with firefighting equipment. They developed a VRTS using a consumer-grade VR headset, the HTC VIVE, and a PH interface in the form of a physical firetruck pump panel. This approach aimed to mitigate the risks associated with real-life training on this equipment while avoiding the expense and complexity of traditional pump simulators. The HTC VIVE headset was coupled with a HTC Tracker to allow real-time tracking of the PH interface in the VE.

In their research presented in Morélot et al. [98], the authors investigated how immersion and sense of presence can affect conceptual and procedural learning outcomes in VR-based fire safety training. The study utilized a CAVE-based VR environment with dynamic fire and smoke simulation and included the use of three tracked replicas of fire extinguishers as PHs for interaction with the VE. The decision to utilize PHs was suitable since the training did not necessitate physical interaction with the VE, as fire extinguishers are tools that can be operated from a distance. To facilitate comparison, a non-immersive VR environment was also assessed, which involved using a desktop PC with traditional controls (mouse & keyboard). The assessment involved administering pre- and post-tests on theoretical knowledge and a procedural post-test, which was evaluated through interviews with trainers and trainees, as well as observations made by the authors during the execution of the learned procedure. The findings revealed that immersion had a significant impact on improving procedural learning but did not influence conceptual learning.

The literature review described earlier served as the foundation for the creation, implementation, and assessment of the VRTS. The aim was to address some of the limitations of prior studies and capitalize on the potential benefits identified in the literature for this type of training technology [86].

To address the issue of insufficient validation in previous works, the development of the VRTS was carried out in collaboration with a forest firefighting unit. Additionally, since previous studies did not explore the impact of the training on real firefighting operators, the VRTS was evaluated within an existing course for novice volunteers in the mentioned first responders body.

To overcome the existing technology barriers [86] associated with immersive VR and minimize the discrepancies with real-world operations, the study adopts several design strategies. For instance, instead of conventional VR hand controllers, they used tracked replicas of firefighting tools as PH interfaces. Additionally, they opted for natural walking as the primary method of moving in the VE, which is the most instinctive VR locomotion technique (as discussed in Section 2.1), and employed a wireless setup for the HMD to potentially reduce the cognitive workload associated with the use of VR.

Using PHs helped to improve the physical accuracy of the VR simulation as consumer VR systems generally have low fidelity, as discussed by authors of Engelbrecht et al. [86]. This was particularly important for the case study since it involved the use of handheld firefighting equipment. Additionally, the floor of the physical space where the VR experience takes place could be seen as part of the User Interface (UI) since most of the interactions with the VE take place when the PHs make contact with the ground.

To enable interaction with the fire while operating the firefighting tools, the decision was made to forego offline fire simulations that are physically accurate. Instead, a real-time, two-dimensional spreading logic based on tiles was implemented. This model was inspired by the wildfire spreading method described in Rothermel [99], although it is less precise.

Finally, with regards to the concern about the uncertainty of skill transfer from virtual to real-world scenarios [86], the evaluation of the proposed VRTS was designed to address this issue. The experimental study aimed to assess the effectiveness of the VRTS in improving procedural skills for a specific firefighting task and compared it to traditional, video-based lessons from the standard course.

To integrate the use of the VRTS into the standard course, the methodology adopted was inspired by the training process proposed in Çakiroğlu and Gökoğlu [97], with some modifications necessary to accommodate the VR experience within the existing training schedule.

2.2.2 Materials and Methods

As previously mentioned, the objective of this study is to assess the effectiveness of a PH-based VRTS for firefighter training in the context of a formal training program. To achieve this objective, a partnership was established with a firefighting organization to develop a training program that could be seamlessly incorporated into one of their regular training courses. To reduce the potential impact of trainees' pre-existing knowledge in the field, the training program was designed specifically for novice volunteers starting their journey as forest firefighters.

The considered firefighting course organization typically provides a two-day theoretical course to novice volunteers, primarily consisting of in-person lectures that cover procedural and safety concepts essential for first-time operators. Upon completing the course, the trainees must pass a certification examination comprising both theoretical and practical components. The course accommodates a maximum of 30 learners in each round.

The training program covers a comprehensive range of subjects. Among them, the application of firefighting modules and individual equipment resulted to be the most interesting for the purpose. In fact, the topic encompasses both long-range implements, such as the backpack pump and blower, as well as handheld tools like the shovel, McLeod, and firefighting broom.

The selection of the use case for this study was based on the observation that the use of individual firefighting tools, particularly hand tools, could benefit greatly from the use of VR and PH interfaces. While the current course is effective in teaching theoretical concepts such as safety regulations, it may present challenges in teaching practical activities such as assembling compound equipment, executing first aid maneuvers, and using individual firefighting tools. This is due to the fact that trainees who have no prior knowledge of basic concepts regarding these subjects and the associated safety risks cannot participate in live-fire exercises. However, they still need to perform these activities correctly in the practical part of the examination to obtain certification.

It should be noted that, while the course is designed for beginners, the participants may already belong to a forest firefighting team and have some prior knowledge related to the topic, possibly gained through informal experiences such as common forestry activities. However, the fact that they are taking the course indicates that they have not yet obtained the necessary qualification to carry out firefighting operations.

Use Case

This study focuses on the use of three specific hand tools (shovel, McLeod, and broom) in the context of forest firefighting. These tools are utilized in close proximity to the fire, exposing the operators to high temperatures and flames. Therefore, their application is limited to slow-burning fires with low flames that affect grass, foliage, or shrubs. Each tool has unique features, and the selection of which tool to use depends on the intended purpose (putting out an existing fire or preventing its spread) and the characteristics of the terrain.

The McLeod is a multi-purpose tool that can be employed to impede the spread of fire by clearing away combustible material, such as foliage or shrubs. To protect the tines during transport, operators commonly use a case to cover them.

On the other hand, the broom is a stick with fire-resistant fabric strips attached to one end, and it is used to smother flames by striking them. It is crucial to use the broom in a consistent, rhythmic motion, hitting the fire every two or three seconds without exerting excessive force. Improper use of the broom may result in insufficient oxygen removal, which could intensify the nearby flames.

Finally, the shovel is a versatile tool that can serve a dual purpose of removing fuel, similar to a McLeod, or smothering flames, like a broom. Unlike the broom, whose fabric strips are more suitable for rocky soils, the shovel is effective in extinguishing fires on both regular and earthy terrains.

Because of their close proximity to the flames and high temperatures, the operators utilizing these tools are required to wear appropriate personal protective equipment. This includes a firefighting suit, firefighting gloves, a helmet with glasses or a visor, and boots. Additionally, the helmet, gloves, and boots offer protection against the sharp edges of the McLeod and shovel.

Given the weight and exposed cutting parts of the considered tools, operators are required to follow a set of guidelines while using them. Specifically, they must ensure to keep the tool in their field of view, maintain a safety distance of four meters from other operators, use the tool appropriately to extinguish or contain the fire, rather than fueling it, and maintain correct posture both during transport and use to avoid unnecessary fatigue.

Shovels, McLeods, and brooms are commonly utilized alongside backpack pumps and blowers. Presently, there is ongoing work to integrate these additional tools into the VRTS, and their potential for VR-based training is being examined as well [100].

Passive Haptic (PH) Interfaces and Simulated Scenario

This section will detail the VRTS and the proposed PH-based VR interfaces. An imaginary scenario was generated using the instructions provided by the Italian forest firefighting unit of the Piedmont Region, and it takes place in a forest clearing (Figure 2.9). The simulation is designed to deal only with grass, foliage, and shrubs fires where the flame height cannot exceed the waist height of the operators. This selection was made to focus the VR training on the use of low-flame tools and not on other tools that are more effective for higher flames.

For this scenario, a flat 10m×10m area was designated where trainees can move around and interact with virtual objects (as the tracked physical space). The designated zone was intentionally devoid of vegetation, but it has the capability of generating digitally reconstructed foliage, grass, and shrubs as game objects in Unity. To initiate the simulation, fuel can be randomly generated within this region or predetermined parameters can be established to specify the quantity, density, and type of fuel at the outset. Fires can also be spawned in this area, which will interact with the fuel. Optionally, Non-Player Characters (NPCs) outside the area will can be enabled to play the role of other operators fighting non-spreading fires to provide context for trainees' actions and offer continuous, visual examples of proper behavior. A bird's eye view of the scenario is reported in Figure 2.9.

The VRTS was developed as an additional component of an existing forest firefighting course, using a VR application created with the Unity 2019.4 game engine and the SteamVR (OpenVR) framework. The program is designed to be used with an immersive HMD that is integrated with PH interfaces. Specifically, the HTC VIVE Pro VR system was used along with several HTC VIVE Trackers (2018) to track the virtual firefighting tools in the VE. To obtain a tracking area of up to 10m×10m, four HTC SteamVR Base Stations 2.0 were placed at the corners of the room. To minimize encumbrance for the trainee, the standard HMD cables were removed, and an HTC VIVE Wireless Adapter Kit was utilized.

Figure 2.9: Wildland scenario depicted in the VR simulation. The invisible grid is composed by tiles of 25cm×25cm (foliage and grass/shrubs have been randomly spawned in correspondence of it).

To create PH interfaces, the physical characteristics of actual tools were replicated. In As depicted in Fig 2.10, the plastic blade of the snow shovel was reshaped to match the original firefighting tool blade shape. This modification ensured safety during training because plastic was used instead of metal. Similarly, for the McLeod, the tines were removed from a real McLeod to make it safer for use in VR. As for the broom, a real tool was used without any alterations, whereas the Obi Rope[5] Unity asset was employed to simulate the physics of the eight flexible strips of the digital replica.

After creating each PH interface, an HTC VIVE Tracker was attached to each of them. This sensor allows the physical object to be aligned in real-time with its virtual counterpart in the VE. This technique is similar to the one described by authors of Suhail et al. [85]. The weight of the tracker is negligible compared to the weight of the tool.

The standard hand controllers that come with the HTC VIVE kit were replaced with a custom configuration to allow trainees to manipulate the PH interfaces more naturally. To achieve this, two standard firefighting gloves were provided to the trainees to simulate the feeling of real personal protective equipment. Two additional HTC VIVE Trackers were attached to the trainees' wrists to track the

---

[5] Obi Rope: `https://assetstore.unity.com/packages/tools/physics/obi-rope-55579`

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 2.10: Considered firefighting equipment (shovel, McLeod, and broom): real usage (a–c), PH props (d–f), and virtual replicas (g–i).

movement of the gloves. However, finger tracking was not implemented, but this was not a significant issue as the focus of the trainees (and the assessment of their performance) was primarily on the handheld prop. The positioning of all the tracking devices was carefully chosen not to interfere with the trainees' actions.

The fire simulation is powered by a plausible cell-based spreading logic, which is not entirely based on physical accuracy. This logic was created in collaboration with experts from the Piedmont Region's Italian forest firefighting unit. A simplified versions of the Rothermel mathematical model [99] was employed to simulate the fire life-cycle and spreading.

The simulation begins with the creation of fuel on the terrain, which can take three different forms: foliage, grass/shrubs, or nothing. The simulation area will be filled with 3D meshes of the corresponding fuel type or have empty spots on the terrain, which quantity and density are chosen randomly. Once the spawning process is complete, the terrain is covered by these meshes, which are spread without any particular structure. To simulate the composition of a real forest terrain, the spawned meshes are allowed to overlap.

After creating the terrain, an invisible grid is placed over it, consisting of a variable number of cells, also called "tiles", with a default size of 25cm × 25cm. To analyze the fire behavior, five rays are projected for each tile, emanating from a point one meter above the tile. These rays are directed towards the corners and center of the tile. These ray-casts provide information about the fuel in the corresponding area, whether it is bare ground, a single mesh, or multiple overlapping meshes. The *maxFuel* parameter for each tile is then derived from this information. Initially set to zero, maxFuel is incremented by five if the fuel type is foliage, 10 if it is grass/shrubs, seven if it is a combination of both foliage and grass/shrubs, and zero if it is an empty spot [99]. A tile with a maxFuel value greater than zero is deemed *Flammable*, while tiles with a maxFuel value of zero are labeled as *Non-flammable*. Additionally, each tile is associated with a pseudo-random *humidity* parameter that depends on the humidity value of the surrounding cells.

Once the setup phase is complete, the simulation starts by generating a fire line on one edge of the grid consisting of multiple fire elements. In the tile-based spreading logic utilized, every fire element corresponds to a tile of the invisible grid. The fire simulation operates at two levels of logic: a low-level that manages the lifecycle of each individual fire element, and a high-level that manages the spread of all fires.

The life-cycle of a fire element comprises three stages, namely *Birth*, *Development*, and *Extinction*. In the Birth phase, a fire element is created on a tile and set to the *Burning* state. During the Development phase, the fire element consumes the fuel associated with its tile periodically by subtracting a value from the remaining fuel (starting from maxFuel) every 0.2 seconds. The subtracted value decreases as the remaining fuel decreases. These parameters also control the particle systems used in the game engine for visualizing the fire element. Once the fuel reaches zero, the fire stops and enters the Extinction state, and the tile is labeled as *Burned* and Non-flammable.

A higher-level logic governs the spread of fire and handles all fire elements collectively. It calculates the damage caused by each active fire element to the flammable tiles within its proximity. This calculation is performed at every simulation frame and takes into account various parameters such as the fire's speed, wind strength and direction (predefined before launching the simulation), the humidity level of each flammable tile, and the remaining fuel of the fire element tile. The resulting value is subtracted from the remaining fuel of the flammable tile, starting from maxFuel. Once the remaining fuel reaches zero, the tile catches fire and is set to the Burning state, and a new fire element is generated. The spread of fire stops when there are no more flammable tiles surrounding the fire elements, and new fire elements cannot be spawned.

The behavior of the fire simulation can be influenced by firefighting tools, each of which has a specific function that can affect the state of the tiles and the fire's behavior (as shown in Figure 2.11). For instance, the McLeod can decrease the maxFuel parameter, thereby reducing the amount of fuel associated with a tile.When the McLeod is applied to a non-burning tile and all remaining fuel is removed, the tile is marked as Non-flammable and cannot be damaged by the spreading logic. Conversely, if the McLeod is used on a burning tile, it causes the fire to spread to the surrounding flammable tiles. In contrast, the broom is used to extinguish the fire directly. Each fire element has an associated *oxygen* value that governs its interaction with the tools. The maximum value is 100, and it is depleted every time a tool interacts with the fire. Once all the oxygen is depleted, the fire element is extinguished, and the corresponding tile is marked as Flammable again. Using the broom excessively or forcefully does not affect the oxygen level, but it can speed up the fire spreading. Finally, the shovel is a versatile tool that combines the functionalities of both the McLeod and the broom. In fact, it can be used for both removing fuel and extinguishing fires.

The VRTS has two operating modes: *guided mode* and *evaluated mode*. The former is designed to provide trainees with practical training on firefighting tools in a step-by-step manner. It also serves as a refresher for concepts covered in the theoretical course that are important to the experience. The latter, on the other hand, is used as a testing ground to assess the trainee's ability to apply what they have learned in a spreading fire scenario.

Guided mode directs the trainee through various stages of tool usage, encompassing transportation, cover removal (with the exception of the shovel), estimation of safety distance, and operation. Each phase comprises two components: an initial segment that provides an explanation of procedural and safety considerations, followed by a performative section in which the trainee must accurately execute a sequence of actions to advance to the next phase. When fire is present, it does not spread, or it spreads in a controlled manner. The detection of errors is achieved by utilizing the 6-DOF positional tracking data of all the HTC VIVE tracked elements (HMD and Trackers), which are utilized to calculate various evaluation parameters such as tool orientation, tool roll, hand position, body posture, etc. This is done at each simulation frame. To make the trainee aware of the mistake, a series of visual cues in the form of on-screen icons are displayed on a panel in the center of the trainee's field of view as soon as an error occurs.

The evaluated mode of the VRTS enables trainees to autonomously apply the

(a) (b) (c)

Figure 2.11: Effect of the firefighting tools on the simulation: shovel (a) and the McLeod (b) can be used to remove the leaves and prevent the fire from spreading; the broom (c) can be used on the flames to removing the comburent and extinguish the fire.

skills learned in the guided mode by simulating a forest fire line. Unlike in guided mode, no auditory or visual prompts are provided, and the can evaluate their performance solely by observing the alterations in the scenario and the fire's behavior. All trainee actions are monitored, and a final report is generated to summarize their overall performance. The report includes scores for different aspects of using firefighting tools, such as transportation, protection removal (rubber case for the McLeod, rubber band for the broom, not considered for the shovel), safety distance estimation, and operation. Additionally, the system alerts the trainee whether they got burned during the simulation or not. The mode is intended to be used multiple times, allowing the trainee to gain confidence in their abilities based on their performance in previous runs as well as the assessment results.

At the start of the experience, the trainee must specify their physical characteristics, such as height and arm's length. The application allows for manual or semi-automatic calibration to be performed. The trainee's body posture is determined by comparing their height to that of the HMD during the simulation, while the extent of arm extension is determined by comparing the length of their arm to the distance between the HMD and wrist tracker.

2.2.3 Experiment

A study involving 30 users was conducted to evaluate the effectiveness of the proposed VRTS.

Experimental Setup

The 30 volunteers (28 males and 2 females) were between 19 and 56 years old (M = 30.70, SD = 10.49). These individuals were selected at random from trainees who were enrolled in the forest firefighting training course. Although most of the participants had little to no experience with VR, almost all of them had some prior experience with the firefighting tools used in the training (particularly the shovel) in non-firefighting contexts.

At the end of the first day of course, the trainees were asked to complete a demographic questionnaire to provide personal information such as their gender and age. After that, they were introduced to the experiment and provided with information about the procedure, the topics covered, and the technology used (including a brief overview of VR and the equipment used). Their previous experience with these topics was also noted.

Following the preparatory step, all trainees participated in a standard course lesson, which covered the use of the individual tool studied. The lesson consisted of three instructional videos that focused on the use of the firefighting shovel. In the first spot, the shovel was introduced, and the different materials used to make it were discussed, along with various ways to use it as a firefighting tool. The second spot highlighted the safety guidelines for transportation and operation, including how to properly carry the shovel and use it to estimate the safety distance for working on fires. Finally, the third spot demonstrated how to use the shovel to remove fuel and extinguish flames. The trainees were expected to acquire knowledge and skills related to the proper usage of the shovel on the fire front, based on the information and guidelines presented in the three instructional videos.

After the instructional videos, the trainees were divided into two groups. The control group (Video-only, V) watched only the videos, followed by a quiz session to assess their understanding of the material. An instructor provided feedback and comments on their answers in a debriefing session. An additional questionnaire was used to evaluate their motivation and overall experience. Notably, in the quiz sessions of the course, trainees were allowed to answer questions multiple times until they answered correctly. However, for this study, the first answer was recorded for later comparison between the two groups. The Video+VR group (V+VR), who experienced the VRTS after watching the videos, were exempt from the quiz session.

Following a brief break, the trainees in the V+VR group were directed to engage in a training session utilizing the VRTS in guided mode, during which they were provided with a detailed, step-by-step tutorial on utilizing the shovel.

Immediately following the previous session, the V+VR group trainees were instructed to utilize the VRTS once again, but this time in evaluated mode. They were required to apply what they had learned from the prior activities (video lesson and VR training) and were provided with an automatic evaluation report of their performance. The presence of NPCs in the VE was enabled to provide context for the trainees' actions.

After the initial Evaluated trial with the fire speed set to the minimum value, the trainees attempted it again at a slightly higher difficulty level in order to improve their previous performance.

After completing the second VRTS session, the trainees were given the same quiz and questionnaire that was used for the control group. An instructor provided feedback on the trainees' behavior and correct application of the learned procedures to ensure that both groups received the standard training required for certification. For the V+VR group, two additional sections were included in the questionnaire to gather their feedback on the usability of the VRTS.

After one week, both the V and V+VR groups were asked to participate in a practical session, which involved applying the concepts they had learned in the field. The practical exam covered all the topics taught in the forest firefighting course, including a section dedicated to individual firefighting tools. In the regular course exam, trainees are divided into squads of six members, and an instructor assesses their performance on a per-squad basis, considering factors such as the use of personal protective equipment, adherence to procedures, teamwork, and timing.

In order to conduct this study, an additional instructor was hired specifically for the purpose of assessing the trainees' use of the shovel during the individual tool exam. This assessment was done on a per-trainee basis and evaluated the same aspects as the VRTS in evaluated mode. The instructor was completely unaware of which group each trainee belonged to.

During the exam, the instructor positioned the squad of six trainees, all wearing their personal protective equipment, in a line with enough space between each trainee. A corresponding line of hand tools was placed a few meters in front of them on the ground. Each trainee was instructed by the instructor to walk towards the tool, pick it up, transport it to an area that roughly represented the fire front, and operate it for a few minutes (as shown in Figure 2.12). The instructor kept track of each trainee's correct and incorrect actions using an assessment sheet. It is important to note that all of the evaluated actions were mandatory, and any non-compliance was deemed unacceptable for the purposes of receiving the certificate.

Once the instructor completed the evaluation of the entire squad, he instructed the trainees to return to the starting point, leave the hand tools in their original position, and switch their positions with their squad members to ensure that each trainee would face a different tool. This process was repeated three times to guarantee that each trainee was evaluated on the use of the shovel and had operated each of the three tools at least once.

Figure 2.12: Objective evaluation phase of the experiment, related to the practical examination of individual firefighting tools taught in the Italian forest firefighting unit course.

Evaluation Criteria

To assess participants' performance and experience with the VRTS, both objective and subjective measures were used. The objective evaluation involved two metrics. The first one, called the *quiz score*, measured the number of correct answers out of 10 multiple-choice questions. The maximum achievable score for this metric was 10. The second metric, referred to as the *practice score*, was based on the evaluation provided by the instructor during the practice exam. The practice score considered the same three dimensions assessed by the VRTS experience in evaluated mode, namely transportation, safety distance estimation, and operation. To simplify the assessment process, each dimension was further divided into several atomic actions, resulting in a total of 12 items to be evaluated.

Four of the 12 items assessed the transportation of the tool, while two items were related to the estimation of safety distance, and the remaining six items focused on the operation of the firefighting tool.

In the practice exam, the instructor awarded one point for each correctly executed item and gave zero points if the item was performed incorrectly or neglected by the trainee. The maximum score achievable for this metric was 12, which was then normalized between 0 and 100.

While the proposed VRTS in evaluated mode has the capability to generate scores for the same three performance dimensions, it was determined that the results obtained during the VR training phase should not be used for comparison, similar to what was done in Çakiroğlu and Gökoğlu [97]. As in that study, the VRTS score was only utilized to provide feedback to trainees between the two trials and to encourage them to adopt appropriate behaviors.

Trainees' subjective evaluation was conducted through questionnaires that were distributed after they had completed video-based training (for the V group) or the VR training (for the V+VR group). The questionnaires were divided into two sections, each aimed at exploring different dimensions.

The first section was based on the Instructional Materials Motivation Survey (IMMS) [101] and evaluated trainees' motivations for learning the topics covered in the training, similarly to what was done by Strada et al. [102]. The questionnaire consisted of 36 statements that were scored on a 1-to-5 Likert scale (ranging from "not true" to "very true"). These statements were categorized into four sub-scales: attention, confidence, relevance, and satisfaction. By using a scoring strategy described in Keller [101], scores were computed for each sub-scale and an overall score.

The second section aimed to collect feedback on the learning experience and utilized the *AttrakDiff* questionnaire [103]. As suggested by Jost et al. [104], only the Attractiveness and Hedonic Quality Stimulation dimensions were analyzed, and the questionnaire included 14 pairs of opposing terms that evaluated the experience on a 1-to-7 scale (with 1 being the positive term and 7 being the negative one).

The trainees in both groups filled in the two sections mentioned earlier. However, the V+VR group had two additional sections in their questionnaire aimed at evaluating the usability of the VRTS. One of these sections requested the participants to rate the usability of the system according to the 10 statements of the System Usability Scale (SUS) [105]. The other section evaluated several usability factors such as functionality, user input, system output, user guidance and help, consistency, flexibility, simulation fidelity, error correction/handling and robustness, sense of immersion/presence, as well as overall system usability. The VRUSE questionnaire [67] was used to assess these factors, and both sections required participants to rate them on a 5-point Likert scale ranging from total disagreement to total agreement.

The complete questionnaire, the quiz given at the conclusion of the theoretical component of the course, and the assessment sheet utilized by the instructor to assess the trainees' performance in the practical exam can be obtained online [6]. In addition, some footage of the experimental activity can be viewed at [7].

---

[6] Experiment questionnaire:
`https://www.dropbox.com/s/7qr0jg1gary8bez/Questionnaire%20%5BEN%5D.pdf?dl=0`

[7] Experiment videos:
`https://www.dropbox.com/sh/2ma9pcuxavsl15b/AAASpROw0mqYtJ1hoi8ZHY-ra?dl=0`

2.2.4 Results and Discussion

The results obtained from the objective and subjective metrics discussed earlier were utilized to compare the performance of the V and V+VR groups, and subsequently, to compare the two training modalities.

To examine the statistical significance of the results, the normality of the data was initially checked using the Shapiro-Wilk test. As the data were found to have non-normal distributions, the non-parametric Mann-Whitney test for two independent samples was used with a 5% significance level ($p < 0.05$) to identify any significant differences.

There were no statistically significant differences observed between the two groups in terms of the individual quiz questions or the overall quiz score (81.33% vs. 85.33%, $p = .547$). This result was expected since both groups received the same video-based lessons on the topics and the amount of repeated information in the VR training was minimized. However, it is noteworthy that the V group's scores had a higher Standard Deviation (SD) compared to the V+VR group. This could be attributed to the fact that trainees in the V+VR group had the opportunity to apply the learned content through the VR experience, which might have equalized their acquired skills. Conversely, the larger variation in the V group's scores may have been influenced by external factors such as trainees' level of attention while watching the videos and their prior knowledge.

Looking at the results of the practical exam, it is clear that the additional VR experience had a positive impact. The instructor evaluated the participants and their scores are shown in Table 2.3 as percentages. It is evident that the V+VR group outperformed the V group in terms of total score (77.77% vs 91.66%, $p = .034$). It is important to note that the evaluation focused on mandatory proficiency aspects, and each evaluated action should have ideally received a score of 100%, except for action number 11, which was optional and concerned the use of a shovel as a McLeod for fuel removal.

The 12 items that make up the total score can be divided into the three phases of transportation, safe distance estimation, and operation, and analyzed separately. In terms of the transportation phase, there were no significant differences (81.66% vs 95.00%, $p = .290$) between the V+VR group and the V group, although the V+VR group showed higher adherence to safety guidelines for each item, reaching 100% adherence for items 3 and 4. It is important to note that the practice exam was structured with a short transportation distance of around 3-4 meters, which compressed the transportation phase for both groups compared to real-life usage and VR experience. A longer transportation phase may have better highlighted the potential advantage of the additional VR experience for the V+VR group.

There were no significant differences (73.33% vs 73.3%, $p = 1.00$) in safe distance estimation between the V+VR group and the V group, and both groups had particularly low scores. The trainees' failure to adhere to safe distance estimation

Transportation	**V**	**V+VR**
1. Hand held the shovel from the balance point / shovel parallel to the ground	86.67%	93.33%
2. Shovel blade oriented outwards	66.67%	86.67%
3. Correct height of the shovel from the ground / arm outstretched	86.67%	100%
4. Shovel blade always kept in the field of view	86.67%	100%
Score	$\mu = 81.66\%$ $\delta = 0.29$	$\mu = 95\%$ $\delta = 0.1$
p-value	$p = .29$	

Safe distance estimation	**V**	**V+VR**
5. Correct body pos. during estimation	80.00%	73.33%
6. Correct arm pos. during estimation	66.67%	73.33%
Score	$\mu = 73.33\%$ $\delta = 0.40$	$\mu = 73.33\%$ $\delta = 0.44$
p-value	$p = 1$	

Operation	**V**	**V+VR**
7. Correct freq. of fuel compression	73.33%	100%
8. Pressure on fuel maintained	73.33%	100%
9. Shovel blade always kept in the field of view	80.00%	100%
10. Safe distance maintained	93.33%	100%
11. Shovel used as a McLeod (fuel removal)	73.33%	80%
12. Correct positioning of hands on the tool	73.33%	93.33%
Score	$\mu = 77.77\%$ $\delta = 0.22$	$\mu = \mathbf{95.55}\%$ $\delta = 0.09$
p-value	$p = \mathbf{.019}$	

Total score	**V**	**V+VR**
Score	$\mu = 78.33\%$ $\delta = 2.57$	$\mu = \mathbf{91.66}\%$ $\delta = 1.41$
p-value	$p = \mathbf{.034}$	

Table 2.3: The practice score metric results include the percentage of trainees who performed a specific action correctly. The M valuess, SDs, and p-values have been provided for the total scores as well as for each of the three phases (transportation, safe distance estimation, operation). Significant p-values ($p < .05$) and better results are highlighted using bold font.

during the practice exam may have been due to the lack of an actual fire front and the absence of a perceived real threat. Although the correct sequence of actions was theoretically known, trainees may have forgotten to estimate a safe distance before beginning to operate on the fire. The VRTS was designed to increase awareness of this aspect, but this result was not surprising given that many trainees in the V+VR group had already demonstrated similar behavior in earlier training phases. While trainees were required to adopt the correct safe estimation pose to advance through the step-by-step VR training in guided mode of the VRTS, most of them overlooked this step in the evaluated mode of the VRTS, likely for the same reasons as in the practical exam.

In the operation phase, which accounted for the majority of the practice exam duration, the V+VR group significantly outperformed the V group (77.77% vs 95.55%, $p = .019$). Trainees in the V+VR group showed 100% adherence to almost all mandatory guidelines, resulting in higher scores. Furthermore, a greater proportion of trainees chose to utilize the shovel for fuel removal, which was a noteworthy aspect of the VR training, was observed.

The results indicate that the extra VR training experience assisted the V+VR group trainees in remembering how to correctly perform various operations, allowing them to avoid errors that were frequently made by the V group trainees. This outcome is consistent with expectations, as trainees in the V group used the shovel for firefighting purposes for the first time, while V+VR group trainees had already undergone the VR experience.

Regarding the subjective evaluation, the outcomes based on the IMMS and AttrakDiff questionnaire are presented in Figure 2.13 and Figure 2.14, respectively.

To simplify the comparison between the two groups, a score was calculated for each sub-scale of trainees' motivation investigated through the IMMS, as suggested in Keller [101].

The results for the four sub-scales and the overall score are displayed in Figure 2.13.

In terms of statistically significant findings, as indicated by the * symbol in Figure 2.13, it can be observed that the trainees in the V+VR group were able to maintain their attention level more effectively than the V group (3.16 vs 4.25, $p < .001$), and rated the experience as more satisfying (3.27 vs 4.07, $p = .009$). Additionally, the significant difference in the total score further indicates that the V+VR trainees were more motivated than the V trainees (3.56 vs 4.14, $p = .004$). These findings can be better understood by examining the specific responses of the trainees to the questions related to attention and satisfaction dimensions.

To delve into the reasons behind the difference in motivation between the V+VR and V trainees, it is useful to examine their responses to individual statements within the attention dimension. Specifically, trainees in the V+VR group found the information presented during the training to be of higher quality, which helped them maintain their attention (statement 11, 3.2 vs 4.13, $p = .040$). Conversely, the V trainees found the training content to be more abstract, which made it harder for them to stay focused (statement 12, 2.53 vs 1.53, $p = .031$). Additionally, the V trainees found the training contents to be more dry and unappealing compared to the V+VR trainees (statement 15, 2.53 vs 1.27, $p = .022$). Finally, the V trainees perceived the experience as having fewer characteristics that stimulated their curiosity than the V+VR trainees (statement 20, 2.8 vs 4.07, $p = .009$).

Further analyzing the attention dimension, the V+VR trainees found the learning experience more surprising and unexpected than the V trainees (statement 24, 2.00 vs 3.87, $p = .001$). The V+VR trainees also considered the variety of information provided as more effective in helping them keep their attention than the V trainees (statement 28, 3.00 vs 4.07, $p = .019$). Finally, the V trainees judged the style of the learning experience as more boring and irritating than the V+VR trainees (statement 29, 2.80 vs 1.60, $p = .015$, statement 31, 2.47 vs 1.33, $p = .018$).

Regarding satisfaction, the trainees who underwent the V+VR training showed a higher tendency than those who underwent the V training to indicate that they enjoyed the experience to the extent that they wanted to learn more about the relative topic (2.93 vs 4.00, $p = .012$). Additionally, V+VR trainees reported enjoying studying the subject matter more than V trainees (2.80 vs 4.13, $p = .001$) and described it as a pleasure to take part in such a well-executed experience (3.13 vs 4.20, $p = .01$).

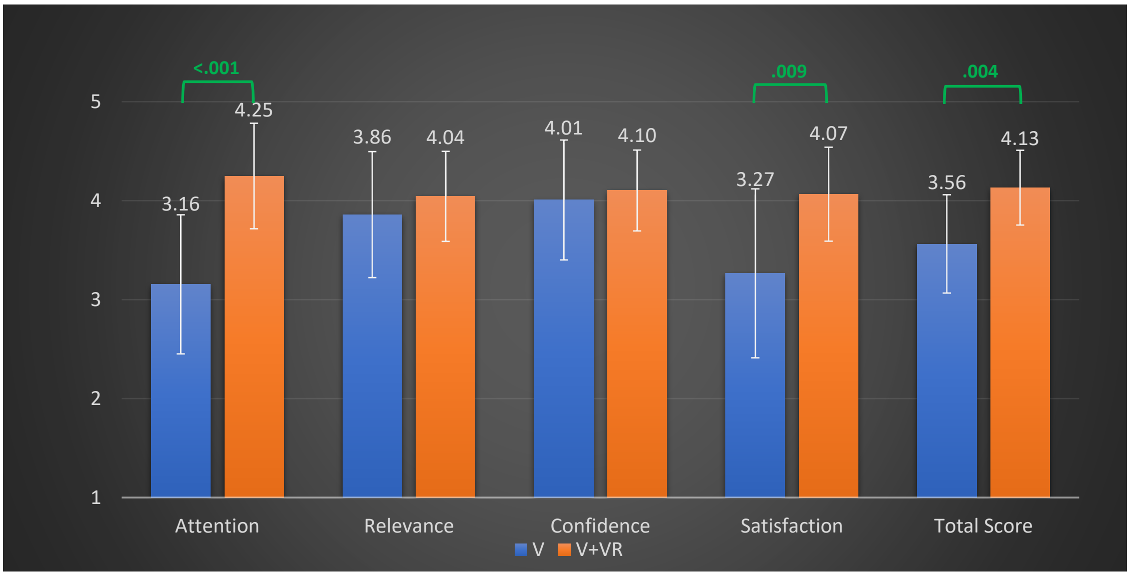


Figure 2.13: Subjective results have been obtained based on the IMMS [101]. The aggregate results concerning the four sub-scales are presented, and bar heights and error bars represent the M values and SDs, respectively.

Regarding the second part of the questionnaire, that assessed the appeal and the hedonic quality stimulation, it can be observed from Figure 2.14 that all assessed aspects exhibit noteworthy variances. The mean scores for the V+VR group on all attribute pairs were better than those for the V group, as indicated by the lower scores, indicating a superior outcome.

In particular, considering the ATT dimension (Figure 2.14), benefits in terms of motivation and satisfaction that were analyzed while discussing the first section of the questionnaire were confirmed, as V+VR trainees considered the experience as more motivating ($p = .038$), appealing ($p = .048$), good ($p = .015$), inviting ($p < .001$), likable ($p < .001$), attractive ($p < .001$), and pleasant ($p < .001$). New positive aspects in favor of the V+VR training that emerge from the analysis of the HQ-S dimension (Figure 2.14) regard the perceived innovativeness, originality, and appealingness of this experience with respect to the V experience. V+VR trainees found the experience more novel ($p < .001$), challenging ($p < .001$), captivating ($p < .001$), innovative ($p < .001$), bold ($p < .001$), creative ($p < .001$), and inventive ($p < .001$).

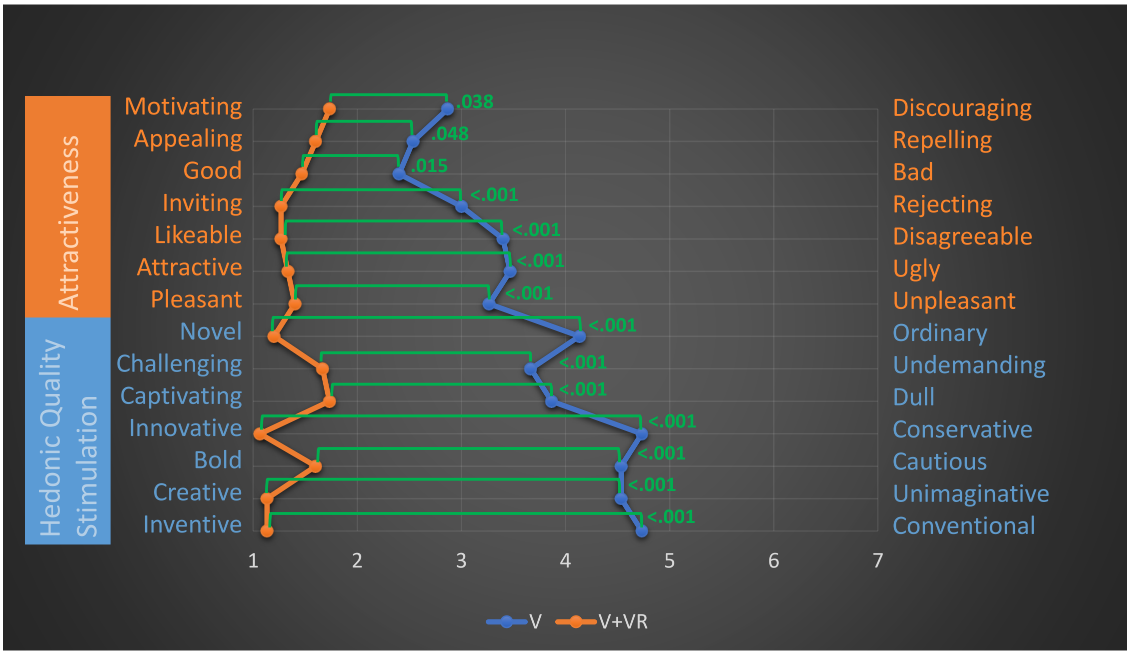


Figure 2.14: Subjective outcomes concerning the AttrakDiff questionnaire [103]: Attractiveness , and Hedonic Quality - Stimulation dimension. A significant difference ($p < .05$) was found for every pair of attributes.

The second part of the questionnaire concludes the comparison between V+VR and V trainees. However, in addition to this analysis, an in-depth assessment of the VRTS used in the experiments was carried out using the SUS and VRUSE questionnaires. The purpose of this analysis was to evaluate specific aspects of the VRTS. Regarding the SUS, the proposed system received a usability score of 78.33. This score corresponds to a B+ grade according to the categorization in Bangor et al. [106], which falls under the “Good” category in the adjective rating scale.

Ultimately, the trainees expressed positive feedback regarding the usability of the VRTS in almost all the aspects evaluated by the VRUSE questionnaire. The figure represented in Figure 2.15 displays the M scores for each dimension.

The trainees expressed a high level of appreciation for the VRTS in terms of its functionality, user input, system output, immersion/presence, and overall usability, as indicated by scores that were generally close to or greater than 4. This suggests that the trainees found the system controls, input device (a real shovel tracked in the immersive environment serving as PH), and output (the HMD and visual feedback) suitable. These factors likely contributed to the trainees’ perception of a high sense of presence and immersion and to their overall positive evaluation of the system’s usability.

Results suggest that error correction/handling and robustness may have limitations in the VRTS, as the trainees showed limited awareness of their errors and the system’s error detection and correction methods. Nonetheless, the other dimensions manifested satisfactory values, which suggest that the system is user-friendly (with user guidance and assistance), has logical system behavior and icon usage (consistency), responds suitably to diverse trainee behaviors (flexibility), and presents precise simulation accuracy of the environment and fire spread.

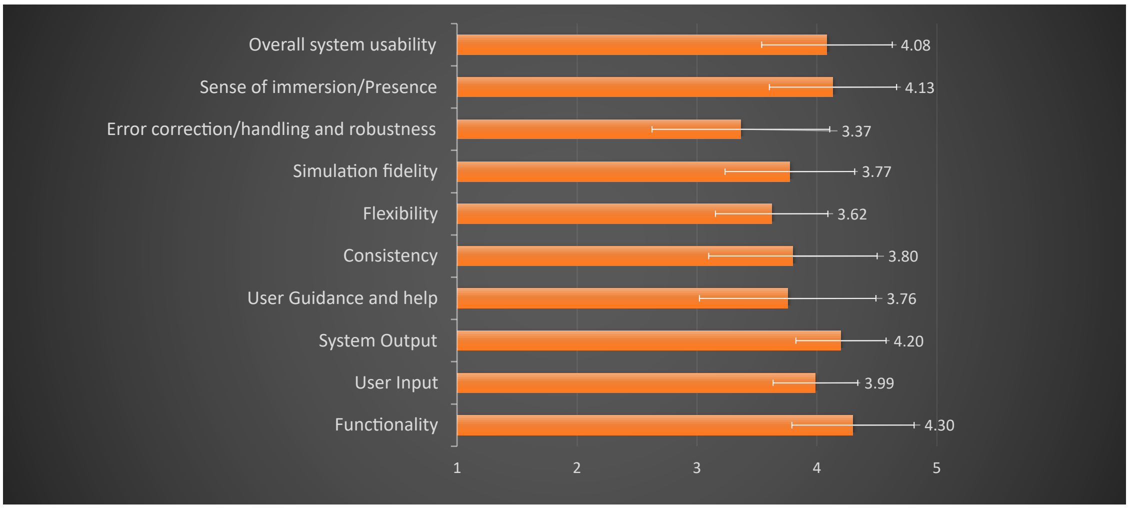


Figure 2.15: Subjective outcomes concerning the VRUSE questionnaire [67] (V+VR trainees only). Bar widths and error bars represent the M values and SDs, respectively.

Limitations

The experimental protocol and the VR training experience have revealed some limitations. Trainees have identified certain hardware limitations related to the tracking performance of the equipment used in the developed VRTS. During the training, trainees may cause occlusion for the trackers on the PH prop they are handling, which can result in unpredictable behaviors of the virtual tool. Similarly, the tracking of trainees' hands can also lead to errors and inaccuracies during the automatic assessment of their actions. Although these issues are infrequent, they could be addressed by repositioning the trackers associated with the PHs or by utilizing two trackers per tool.

Aside from tracking issues, some trainees also reported needing more physical space to perform their actions. In the guided mode of the VRTS, users were required to simulate transportation by moving around in circles for a few minute, which may be perceived as disorienting and tedious due to the limited physical size of the room. Additionally, the virtual space depicted in the evaluated mode is much larger than the actual space available in the real world. Although the playable area adjusts to the real room size, some trainees felt restricted and constrained due to the lack of complete freedom of movement. These issues could be addressed by expanding the tracking area, for example, by utilizing a higher number of base stations or by employing inside-out VR devices. However, in the latter case, the PH props may require different tracking technology as inside-out HMDs usually do not support additional tracked elements other than the hand controllers.

During the VR phase, an issue arose concerning the error icons' functionality. Some trainees found them confusing and ill-placed, leading to ambiguity when tracking issues occurred. Moreover, some found them annoying. It may not be feasible to replace them with audio feedback, as it could increase perceived annoyance. Therefore, alternative approaches should be explored to offer more intuitive and comfortable continuous feedback on the actions performed.

For what it concerns the training with the evaluated mode of the VRTS, a few trainees indicated that they wished they could have had more opportunities to practice, in order to enhance their proficiency even further. Due to logistical considerations, the experimental design only allowed for two attempts at this training mode. However, it seems reasonable to assume that allowing trainees to repeat the experience multiple times until they feel fully adept would yield even more favorable outcomes compared to the traditional course lesson alone.

An observed issue during the experiments was the trainees' failure to adhere to the prescribed safety distance estimation. To address this, one possible solution could be to modify the guided mode of the VRTS by prompting trainees to assume the safety distance estimation pose multiple times and emphasizing its importance in the voice-over explanation. Additionally, evaluated mode could be improved by adding more elements to immerse the trainees in the situation, such as extending the distance traveled during the transportation phase or adding a visual representation of the fire front, and even a controlled fire if possible.

Future Developments

In addition to addressing the aforementioned limitations, future developments could focus on expanding the analysis to include other hand tools supported by the VRTS, such as the McLeod and broom, by applying the devised experimental protocol in future course rounds.

Furthermore, it could be valuable to expand the evaluation of the VRTS beyond individual hand tools to encompass ranged tools like backpack pumps and blowers. Since these tools would require unique simulation approaches for both the VR scenario and PH interfaces, the findings could be particularly insightful.

Finally, one possible way to assess knowledge retention is to invite the trainees who participated in the experimental activity to a follow-up session, such as a planned refresher course, and assess their ability to apply what they learned and recall from their prior experience in the VRTS.

2.3 Consumer Devices for Haptic Simulation

Despite the growing availability of haptic devices in the consumer market [3], [9], the majority of end-users interacting with VR environments still rely on handheld or hand controllers. These commonly used devices, such as the HTC VIVE wand[8], Oculus Touch[9], and Valve Index Knuckles[10], are capable of delivering haptic feedback such as ATF through vibrations on the palm or back of the hand, as well as PTF and PKF due to their shape and material textures.

In order to encourage the widespread adoption of VR , the cost of standard hand controllers has been kept relatively low to ensure affordability [3]. As a result, these controllers typically offer simplified input functions based on buttons, as well as basic tactile feedback such as vibration and material texture [3]. However, one potential limitation of most hand controllers is that they require continuous gripping [9]. This could be problematic in scenarios that involve interacting with virtual objects using bare hands, or requiring the user to hold one or more virtual tools during the experience [9].

There are instances where using a prop to represent the virtual object being manipulated is considered a more natural way to interact compared to hand controllers [107]. This mode of interaction can be achieved by using real tools as interfaces or by utilizing PHs, which are physical props that substitute them [108]. In both cases, the props must be aligned with their visual counterparts in the VE [38]. For instance, simple elements such as levers, knobs, buttons, and manual tools like

---

[8] HTC VIVE controller: `https://www.vive.com/eu/accessory/controller/`

[9] Oculus Touch controller: `https://www.oculus.com/accessories/quest/`

[10] Valve Index Controller: `https://www.valvesoftware.com/en/index/controllers`

flashlights or probes can be easily approximated using lower-fidelity PH interfaces with 6-DOF tracking capabilities [39].

PH can also be combined with hand controllers to enhance the fidelity of PTF and PKF [109]. In this case, the VR system's tracking capability can be used. However, the issues that can affect hand controllers mentioned earlier may not be resolved.

To address the issues associated with hand controllers, there has been growing interest in hand tracking. Some consumer VR systems, like Meta Quest, have already integrated optical-based hand tracking, while other systems can use dedicated components like the Leap Motion Controller[11]. Hand tracking-based interfaces offer potential advantages over hand controllers, although the lack of PH feedback could reduce the perceived naturalness of the interaction[30], [110]. Furthermore, combining hand tracking with PH props is currently not feasible due to visual occlusions that would interfere with the optical-based hand tracking.

Haptic gloves are another type of VR interface based on hand tracking. These devices allow for free-hand, multi-finger interaction and can provide both ATF at the fingertips and, in some cases, kinesthetic feedback (AKF and/or PKF) [111]. Kinesthetic feedback is typically delivered through kinematic structures based on exoskeletons, but these structures can be limiting in terms of wearability and encumbrance [112]. To improve wearability, one approach is to move the base of the kinematic chain closer to the point of application of the stimulus, but this can result in the device being unable to provide kinesthetic feedback [112]. Additionally, the complex kinematic structure used by these gloves can take up a lot of space around the hand, making it difficult to handle generically shaped physical props or interact with virtual objects and other hands.

However, VR gloves that only provide finger tracking and ATF can offer more natural interactions with both virtual and real elements. These gloves do not require articulated structures and can fit like normal fabric gloves [39], [113].

Based on the previous analysis, it seems that while fully-featured haptic interfaces offer advantages, it may be worthwhile to also investigate combinations of devices such as gloves, hand controllers, or PH props to leverage the benefits of each technology while mitigating their limitations. For instance, in a previous study reported in Calandra et al. [39], gloves were utilized in conjunction with PHs to provide both ATF and PKF feedback during manipulation-based tasks.

Object manipulation is a frequently studied use case for exploring haptic interaction in VR. Many studies in the literature have focused on the use of haptic devices to simulate feedback during the operation of both passive tools such as screwdrivers, saws, and hammers [114], and active tools such as ES and drills [109], [115]. Simulating these tools, which are commonly used in fields such as carpentry,

---

[11] Leap Motion Controller:
`https://www.ultraleap.com/product/leap-motion-controller/`

can be challenging and may require the use of multiple (AH and PH) haptic stimuli to enable realistic interactions [116]. Providing both AKF/PKF and ATF stimulation can be crucial for both types of tools, but ATF becomes more important for active, electromechanical tools as it can simulate the vibrations generated by their motors [115].

Building on a previous research discussed in Pratticò et al. [40], a study to further evaluate haptic configurations for simulating the use of an active electromechanical tool in VR was conducted during the PhD period. Specifically, the previous study mentioned above evaluated two haptic setups: one based solely on gloves with haptic feedback for both ATF and PKF, and the other on gloves with ATF feedback only, combined with a user-made mockup. The evaluation focused on various aspects of UX, including usability, fidelity, presence, cybersickness, task load, and task performance, and simulated different phases of an ES task. This type of task is particularly relevant for haptic simulation, as the lack of haptic feedback in VR can make it difficult to use these types of tools effectively, especially in training scenarios involving assembly operations [108], [110].

Based on the evaluation conducted in Pratticò et al. [40], it was found that the combination of ATF-enabled gloves and a custom-made mockup of the tool (used as a handheld prop) outperformed the use of a single, more feature-rich device. However, the increased effectiveness of this setup came at the expense of greater complexity, and it was not determined how each component contributed to the various aspects of the users' experience. As a result, it is possible that alternative configurations derived from this combined setup could be applied to the use case in question while still maintaining an appropriate balance between performance and setup complexity.

The new study focuses on exploring the delivery of multi-component haptic feedback with minimal hardware requirements. To achieve this, the study conducts a breakdown analysis of the reference setup, which involves identifying simpler configurations based on its components and comparing them with the original design. The aim is to investigate how downgraded configurations perform in terms of the dimensions studied in the previous research.

2.3.1 Background

Numerous studies have investigated haptic devices for simulating interactions within VR environments, as documented in the literature [3], [9], [117]–[119].

With the increasing availability of commercial handheld and wearable haptic devices on the market [3], these first two categories have become more relevant to consumers than the others. Among the various scenarios that can benefit from haptic feedback in VR, it is necessary to mention applications that require users to interact with manual tools such as carpentry or surgical instruments. For example, in Strandholt et al. [114], a system is proposed in which physical props representing tools such as hammers, screwdrivers, and saws are used to provide PH feedback through the shape of the real objects. Additional sensors (HTC VIVE Trackers) are utilized to track the physical props, and the working table is also tracked and aligned with its virtual representation to better simulate the interaction between the table and the handheld physical tools. The results demonstrate that this approach can offer a higher level of realism compared to using traditional VR controllers or physical props (the table) alone.

Regarding active tools, in Choi et al. [115], an interactive drilling simulator designed for training purposes is presented. The simulator employs a mockup to provide various haptic feedback including contact response, thrust force, machinery vibration, torsional forces, and edge penetration forces associated with the task. The mockup also functions as a PH interface, enabling the user to feel the shape of a conventional hand drill in their hands.

In Jeong et al. [109], a method for simulating the use of an Electric Screwdriver (ES) in VR is proposed using a standard VR hand controller as a virtual representation of the ES. The tracking capabilities of the controller are utilized to easily incorporate the device in the VE, and vibrational feedback is provided to simulate the screwing action. To achieve PH stimulation, the work explores four different setups: realistic, which uses a controller with the same shape, center of mass, and weight as the real tool; grip-force, which mounts the components of the real tool on the hand controller to simulate the shape but not the weight of the tool, as the controller is lighter than the real tool; grip-only, which provides feedback only of the grip part of the tool through the hand controller; and virtual, which only uses the hand controller. The four setups were compared with the direct use of the real ES in both virtual and real settings, as the experiment required the users to perform a screwdriving task in wood.

While several studies have examined various dimensions related to VR interactions such as immersion, presence, usability, and task load, few have explored the combination of PHs, hand tracking techniques, and kinesthetic devices like haptic gloves. Furthermore, many of these studies tend to ignore consumer technologies. In contrast, authors of Pratticò et al. [40] assessed two configurations based on consumer haptic devices. The first configuration involved a exoskeleton-based VR haptic glove (SenseGlove DK2) that provided Active Tactile Feedback (ATF) and Passive Kinesthetic Feedback (PKF), while the second configuration combined a ATF-enabled glove (Manus Prime X Haptic VR) and a 3D-printed PH prop mounted on an HTC VIVE VR hand controller to provide the same kind of feedback. The study focused on applications in which users needed to operate an ES, and experimental results showed that the second configuration was perceived as more usable and less straining than the first one. Additionally, the glove-and-prop configuration made users perceive the tool shape and weight better, leading to better task accuracy.

Building upon the interesting results achieved in Pratticò et al. [40], the study reported herewith expands upon the above investigation by examining the setup that demonstrated the most promising outcomes - the integrated use of ATF-enabled gloves, the 3D printed PH prop, and a VR hand controller. The research specifically examines the contributions of various combinations of the different elements (consumer devices and prop) used in the selected setup towards providing a realistic feedback. To accomplish this goal, a user study was conducted to compare the initial setup (proposed in Pratticò et al. [40]) to two other configurations: the exclusive use of ATF-enabled VR gloves, and the combined use of VR gloves and VR hand controller.

2.3.2 Materials and Methods

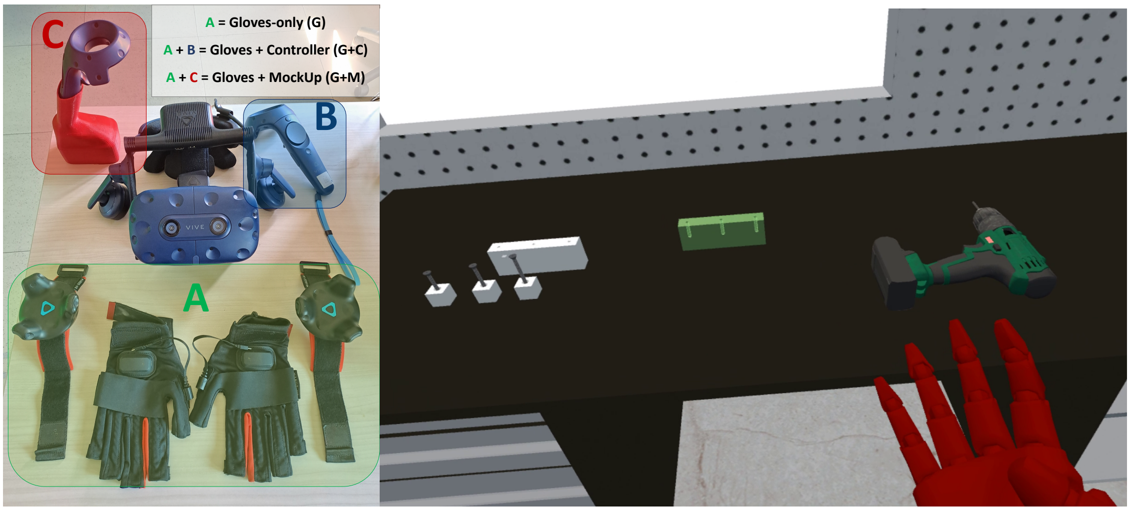
As mentioned before, this investigation focuses on exploring the aspects related to delivering multi-component haptic feedback in VR simulations of active tools. The study aims to investigate the impact of different combinations of haptic devices and props on the realism of feedback provided to the user. To achieve this, a breakdown analysis is conducted, to compare the original configuration, which showed the best results in the previous study [40], with two downgraded versions that isolate the contribution of individual components.

To simplify the setup, subsets of the assembly components were identified to be reasonably applied to the original use case. The reference setup, referred to as *Gloves+Mockup* (G+M), utilized a pair of ATF-enabled gloves and a custom-made mockup of the simulated ES as a prop to interact with objects in the VE. The user's wrist's 6-DOF tracking was obtained through two additional HTC VIVE Trackers. To track the mockup, a common VR hand controller could be used instead of another HTC VIVE Tracker, along with a motor to deliver vibrations. It was chosen to use an HTC VIVE wand inserted in a shell as PHs, similar to what was done in Joyce and Robinson [38]. Consequently, the controller was not employed as a direct means of interaction between the user and the VE, but rather as a component of the ES mockup.

Haptic Configurations

For this study, the setup used in the reference work [40] was replicated using similar technologies (Figure 2.16a). Specifically, the Manus VR Development Kit (DK) gloves [120] were used, in addition to a 3D-printed PH prop with a mounted HTC VIVE Pro wand. Two HTC VIVE Trackers (2018) were also utilized for tracking the user's wrists.

A thorough analysis of the possible alternatives was conducted to identify other configurations that could be evaluated for obtaining simpler setups. The results of this analysis are presented in Table 2.4.



(a) (b)

Figure 2.16: Hardware (a) and scenario (b) used for the experiment.

The first configuration that underwent downgrading involved removing the 3D-printed shell from the mockup used in the reference setup, resulting in what is referred to as *Gloves+Controller* (G+C). In this configuration, the gloves still serve as the primary interface between the user and the VE, while the hand controller acts as a prop for the virtual ES and provides 6-DOF tracking capabilities. As a result, ATF is delivered by both the gloves and the controller, while PTF and PKF are delivered only by the controller, albeit with lower fidelity compared to the 3D-printed shell. Similar to the G+M configuration, the trigger button of the controller is used to simulate the activation button of the virtual ES. The rationale for this configuration is to simplify the setup while still maintaining some of the key components.

The second downgraded configuration is referred to as *Gloves-only* (G), which solely relies on the gloves to provide ATF. With this setup, the physical prop is eliminated, and users are required to touch the virtual ES with their hand and partially close their fist to grip the tool. They can then manipulate the ES, use their index finger to activate its button, or release it onto the table by fully opening their hand. Additionally, ATF is used to indicate significant events, such as contact with a virtual object or tool grab/release.

In Pratticò et al. [40], the use of VR gloves with finger tracking capabilities was deemed necessary for the activity being performed. Therefore, for fairness purposes, these gloves were also included in all configurations evaluated in the new study. Finger tracking was utilized to create a virtual representation of the users' hands in the VE and to enable object interaction through grabbing and manipulation.

ES MOCKUP			DESCRIPTION	CONFIGURATION
6-DOF POSITIONAL TRACKING	ACTIVE TACTILE FEEDBACK	PASSIVE TACTILE AND KINESTHETIC FEEDBACK		
X	X	X	Original configuration	G+M
X	X	-*	Removal of the 3D-printed shell	**G+C**
X	-	X	Vive controller does not provide vibration	Discarded (no reasons to discard vibrational feedback)
-	X	X	3D-printed shell without 6-DOF tracking	Discarded (impossibility to interface the mockup with the VE)
X	-	-*	Removal of the 3D-printed shell, Vive controller does not provide vibrations	Discarded (no reasons to discard vibrational feedback)
-	X	-*	Completely virtual ES, vibrations provided through VR gloves	G
-	-	X	3D-printed shell only, no tracking and no vibrations provided	Discarded (impossibility to interface the mockup with the VE)
-	-	-	Completely virtual ES, no vibrations provided	Discarded (no reasons to discard vibrational feedback)

Table 2.4: Breakdown analysis of the compound configuration presented in Pratticò et al. [40]. The symbol -* denotes the indirect provision of low-fidelity feedback through the hand controller.

Use Case

The task that was simulated remained mostly the same as that of Pratticò et al. [40], which was in turn based on the screwing activity described in Jeong et al. [109]. Like the previous study, the task was created using Unity (2020.3) and the SteamVR framework.

The user is placed in a VE that simulates a garage workshop[12], where a working table is located. In the G+C and G+M configurations, where interaction with physical objects is required, a real table is aligned with the virtual one across all configurations. The real surface enables the user to pick up the handheld device from the working table at the start of the experience and to place it back when necessary (if present).

In Jeong et al. [109], the task involved tightening four screws that were already

---

[12] Simple Garage:
`https://assetstore.unity.com/packages/3d/props/interior/simple-garage-197251`

placed in a vertical position and ready to be screwed into a wooden table. To better simulate real-world screwing conditions, the present work includes a manipulation part in the task, similar to [40]. Specifically, the user is required to take a pre-drilled bar (0.26m ×0.05m× 0.04m) and place it in a specified location on the working table before tightening three blocks of the same material (0.04m × 0.02m×0.04m) using three screws of varying lengths (30mm, 50mm, and 80mm). Both visual highlighting and ATF are used to aid interaction with these virtual objects. Figure 2.16b displays a screenshot of the VE, whereas Figure 2.17 shows the various steps of the simulated procedure. It is noteworthy that this manipulation task is more realistic than the one in Jeong et al. [109].

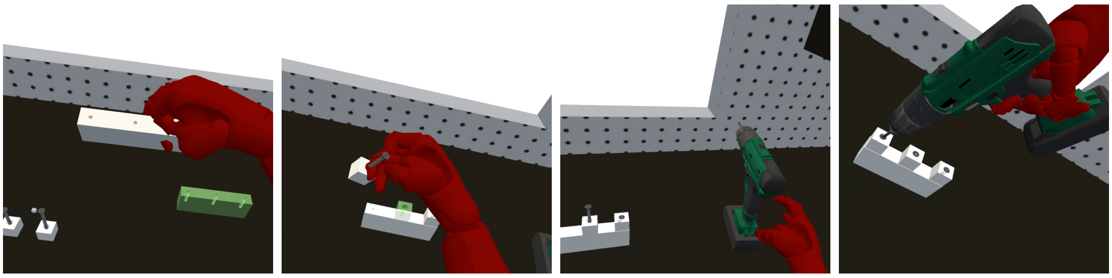

(a) (b) (c) (d)

Figure 2.17: Steps of the simulated procedure: a) positioning of the pre-drilled bar, b) positioning of one of the blocks, c) grabbing of the simulated ES, and d) ES operation.

Before starting the main task, an initial training phase is provided to allow the users to become familiar with the VE and related equipment. This preliminary experience, called the “sandbox” training, enables the users to interact with various virtual objects of different shapes and sizes placed on the working table.

The simulation allows for haptic feedback to be reproduced during different phases of the screwing activity, including screw head touch, screw head slip, loose screwing, screwing, and screw tightened. In accordance with [109], it is necessary to keep the ES perpendicular to the screw head’s surface to initiate and correctly perform screwing. Failure to maintain this position will cause the ES tip to slip from the screw head.

In the previous study, screwing was investigated on wood and aluminum, which are materials commonly studied in haptic research due to their varying stiffness [121]. Additionally, these materials are frequently used in real-world screwing tasks [122]. However, since no significant differences were found between the two materials in the previous study, only wood was used for the bar and gussets in the new study.

To improve the fidelity of the ATF associated with the simulated ES, vibrations generated by a real ES during the described task were collected using an accelerometer (ADXL345) mounted on the grip of the actual ES. Furthermore, signals related to each screwing phase in the selected material were filtered with a low-pass filter set at 450Hz, which represents the upper limit of tactile perception for the human sensory system [3], [30].

2.3.3 Experiment

To minimize possible confounding factors related to perceptual variability among participants and to better identify potential differences among the configurations, the experiment followed a within-subjects design, which was the same as the one used in Pratticò et al. [40]. In this design, participants were first instructed to perform the task described in the "Task" section by using a real ES, three wooden gussets, and screws (as shown in Figure 2.18a). This first phase allowed participants to obtain a common reference for the haptic sensations perceived during the task, which they could later compare with the simulated scenario. Additionally, this approach accounted for participants who may not have had prior experience with a real ES.

Before being exposed to the simulated ES activity, the participants were given time to interact with the VE using the Manus VR gloves in a sandbox scenario. Afterward, they were asked to perform the task with the three considered configurations (as shown in Figures 2.18b, 2.18c, and 2.18d). The Latin square order of exposition was adopted to counterbalance potential learning effects and minimize possible biases.

A total of 28 volunteers were recruited among the staff and students at Politecnico di Torino, none of whom had participated in the study of [40]. The participants ranged in age from 21 to 32 years old (M = 26.11 years, SD = 2.78 years) and consisted of 75% males and 25% females.

In a similar way to the reference work, several objective measures were automatically recorded within the VR experience, including the accuracy of centering the screw head with the tip of the ES, the duration of manipulating virtual objects, the time spent on each phase, and the number of times the tip of the ES slipped out of the screw head.

Likewise, the same questionnaire as in Pratticò et al. [40] was employed for subjective evaluation. The questionnaire consisted of the following sections:

1. A before experience questionnaire, which was administered before the screwing activity with the real ES, consisting of questions about demographics, prior knowledge of the operation of a real ES, and expertise with the used technologies.

2. A post-real experience (with the ES) questionnaire, which was administered after the screwing activity with the real ES but before exposure to the sandbox

(a) Real ES

(b) Gloves+Mockup (G+M)

(c) Gloves+Controller (G+C)

(d) Gloves-only (G)

Figure 2.18: Real ES procedure and evaluated haptic configurations.

VE. This section evaluated the participants' ability to discern among the various haptic sensations related to the use of the real tool, and it also included a pre-experience Simulator Sickness Questionnaire [63].

3. After experience questionnaire, which was administered after the three configurations were experienced, designed to assess the remaining dimensions of interest for the activity. This section included four standard tools: the SUS [105] for usability, Simulation Task Load Index (SIM-TLX) [123] for task load, User Experience Questionnaire (UEQ) [124] for UX, and two sections of the VRUSE [67] for fidelity and presence. This was followed by a second post-experience iteration of the SSQ and a custom section to directly compare the three configurations by asking the participants to rank them along several relevant dimensions.

In addition, open feedback from the participants was collected.

2.3.4 Results and Discussion

To analyze the statistical significance of the results, the Friedman test was performed with a significance level of $\alpha \leq 0.05$. In case of significant results, the Wilcoxon signed-ranks test was used as a post-hoc analysis.

There was no statistical difference found between the aspects considered in the before experience questionnaire of the present study and that of the reference work [40].

Table 2.5: Objective results related to the screwdriving task.

Metric	[unit]	**M (SD)**			***p*-value**			
		G	*G+C*	*G+M*	*Friedman*	*G /G+C*	*G /G+M*	*G+C /G+M*
ES tip centering accuracy	[mm]	8.87 (4.17)	7.23 (1.88)	7.31 (1.96)	.114	–	–	–
Number of slips	[#]	2.43 (2.06)	1.18 (1.44)	1.32 (1.23)	.054	–	–	–
Grabbing time before first ES interaction	[s]	2.64 (1.22)	2.43 (0.8)	2.33 (0.63)	.381	–	–	–
Grabbing time after first ES interaction	[s]	2.22 (0.73)	2.20 (0.53)	2.33 (0.53)	.156	–	–	–
Time elapsed at screw tightened	[s]	1.13 (0.88)	0.54 (0.11)	0.52 (0.09)	**<.001**	**<.001**	**<.001**	.350

Out of the participants, 57.0% had little to no experience with a HMD, while 43.0% were quite familiar with immersive VR technologies. Additionally, 35.0% of participants had limited experience with a real ES, while 65.0% were moderately to very familiar with it. The only haptic sensation that was not easily discernible in the post-real experience questionnaire was the ability to detect screw length (Figure 2.18a).

Regarding the objective data collected during the experiments (Table 2.5), most metrics showed no significant differences. However, it took longer to recognize that the screws were fully tightened when using the G configuration compared to the other configurations.

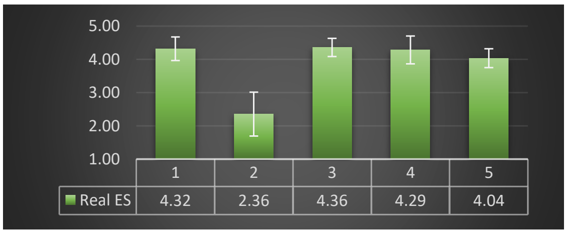
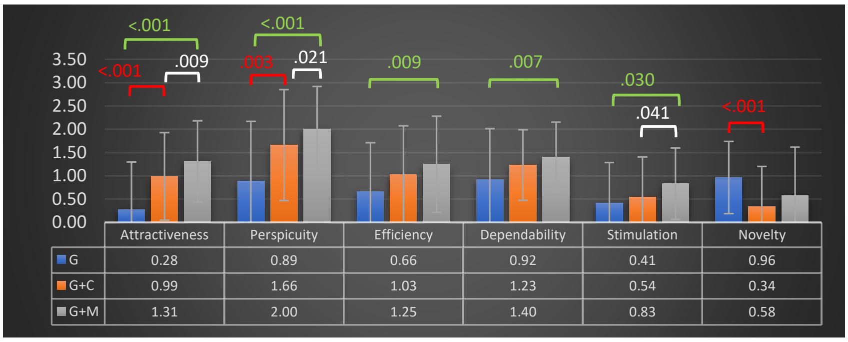


Figure 2.19: Subjective outcomes concerning the post-real experience (with the ES) questionnaire.
Considering the haptic feedback perceived: #1. I felt the contact with the screw head; #2. It felt different screwing screws of different lengths; #3. I felt when screw was tightened; #4. I felt when screw head slipped; #5. I was able to distinguish the different phases of the screwing activity in the wood;

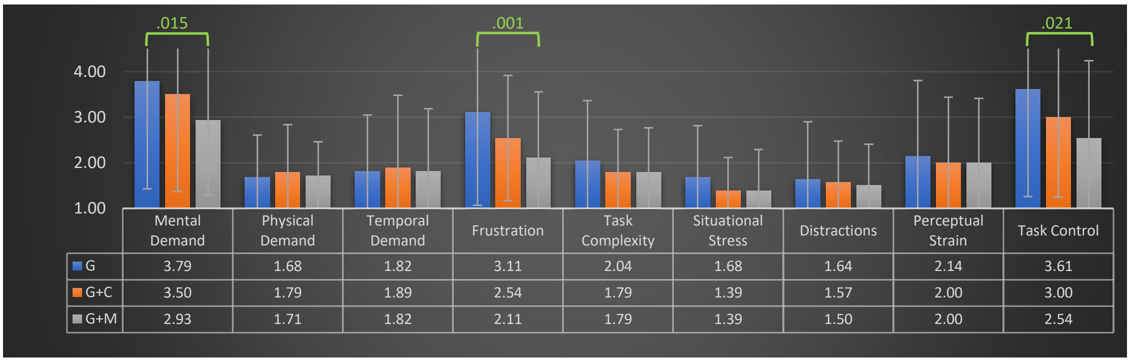
There were no significant differences in pre-/post-SSQ indicators for cybersickness. Based on the results of the SUS questionnaire ($M_G = 62.23$, $M_{G+C} = 74.11$, $M_{G+M} = 76.70$, $p < .001$), the G configuration was found to be significantly less usable than both the G+C ($p = .003$) and G+M configurations ($p < .001$).

The results of the UEQ regarding the UX are presented in Figure 2.20. Significant differences were found among the three configurations, with participants perceiving the G+M as the most attractive and perspicuous, followed by G+C and G. The G configuration was perceived a less efficient than G+M, while the latter was considered the best configuration in terms of stimulation. Additionally, G+M was rated significantly more dependable than G. The G+C configuration, which



	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
G	0.28	0.89	0.66	0.92	0.41	0.96
G+C	0.99	1.66	1.03	1.23	0.54	0.34
G+M	1.31	2.00	1.25	1.40	0.83	0.58

Figure 2.20: Subjective outcomes concerning the UEQ. Brackets are used to report significant differences ($p < .05$) and error bars are used to indicate the SDs.



	Mental Demand	Physical Demand	Temporal Demand	Frustration	Task Complexity	Situational Stress	Distractions	Perceptual Strain	Task Control
G	3.79	1.68	1.82	3.11	2.04	1.68	1.64	2.14	3.61
G+C	3.50	1.79	1.89	2.54	1.79	1.39	1.57	2.00	3.00
G+M	2.93	1.71	1.82	2.11	1.79	1.39	1.50	2.00	2.54

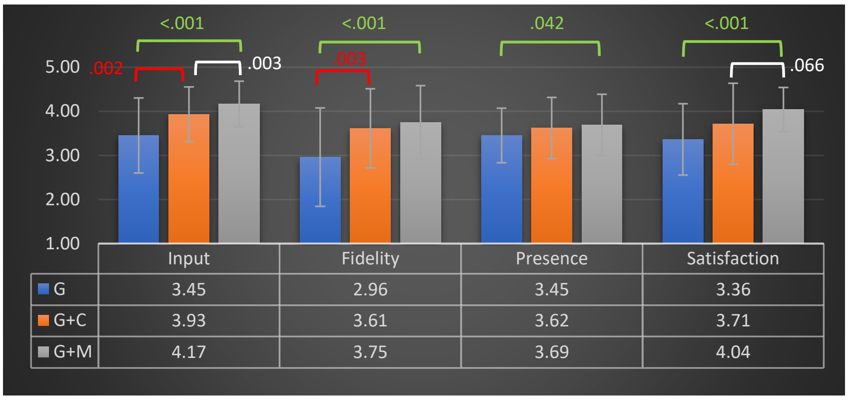
Figure 2.21: Subjective outcomes concerning the SIM-TLX [123]. Brackets are used to report significant differences ($p < .05$) and error bars are used to indicate the SDs.

relies on a more traditional use of the hand controller, was considered less innovative than G, which utilizes only hand tracking and ATF. However, no significant differences were found in terms of novelty when comparing G+M with the other two configurations.

The reasons underlying the observed differences in UX can be deduced from the results of the remaining sections of the questionnaire, which are discussed below. Specifically, in the SIM-TLX section (Figure 2.21), statistically significant differences indicate that the G+M configuration was perceived as better than G in terms of mental demand, frustration, and task control.

The findings may be partially explained by the grabbing logic used in the G configuration of the system. Unlike the G+C and G+M configurations, where the tool was aligned with a physical object that could be handled like a real tool, the G configuration required participants to perform multiple actions (such as grabbing the virtual tool and operating its trigger while holding it) with an entirely intangible object. The absence of PH feedback could have made this task particularly challenging and tedious. However, the lack of significant differences between the G and G+C configurations suggests that the main reason for the observed differences can be attributed to the presence of the mockup.

In terms of evaluating usability factors with the VRUSE questionnaire (Figure 2.22), participants expressed the greatest satisfaction with the input method when using the G+M configuration, while the lowest satisfaction was reported for the G configuration. In terms of simulation fidelity, taking into account visual, aural, and haptic feedback, participants generally had higher satisfaction with the simulation when using both the G+C and G+M configurations than with the G configuration. Additionally, significant differences were observed between the G+M and G configurations in terms of presence.



	Input	Fidelity	Presence	Satisfaction
🟦 G	3.45	2.96	3.45	3.36
🟧 G+C	3.93	3.61	3.62	3.71
⬜ G+M	4.17	3.75	3.69	4.04

Figure 2.22: Subjective outcomes concerning the VRUSE [67]. Brackets are used to report significant differences ($p < .05$) and error bars are used to indicate the SDs.

The results of the custom section (Table 2.6, dimensions analyzed in the caption) reveal that the G configuration was perceived as less comfortable, despite not requiring any physical object to be held. This may be due to the clunky interaction paradigm required to grab and use the non-physical ES. The naturalness of the manipulation of the wooden equipment (bar and blocks) could not be distinguished between the three configurations as interaction relied on gloves for all configurations. The G+M configuration was considered the best configuration for interacting with the virtual ES and for fidelity of the ES handle's shape, whereas the G configuration was perceived as the worst. No significant differences were observed regarding the provision of haptic feedback similar to that of a real ES, which suggests that the three configurations are equivalent in this regard. However, the G+M configuration allowed participants to perceive screw tightening status and screw head slip more accurately than the G configuration. The G+C and G+M configurations provided participants with a higher sense of control and efficiency compared to the G configuration. Overall, the G+M configuration was preferred,

Table 2.6: Subjective ranking of configurations (custom). Which configuration: #1. Is more comfortable to a sustained use; #2. Lets you feel more natural manipulating bars and blocks; #3. Allows you to better interact with the virtual ES; #4. Gives you a more faithful feedback of the real ES handle's shape; #5. Gives you a vibrational feedback closer to the real ES; #6. Allows you to better perceive your hands' and fingers' position/tracking; #7. Gives you a more faithful feedback of the ES's activation button; #8. Lets you more faithfully perceive screwing in the wood; #9. Lets you more faithfully perceive the contact with the screw head; #10. Lets you more faithfully perceive differences in screwing screws with different lengths; #11. Lets you more faithfully perceive the screw tightening status; #12. Lets you more faithfully perceive the screw head slip; #13. Lets you better distinguish the different phases of screwing in the wood; #14. Allows you to experience a higher sense of control over the task; #15. Allows you perform the task with higher efficiency; #16. Do you prefer (overall).

Item	**Rank Median (SD)**			***p*-value**			
	G	*G+C*	*G+M*	*Friedman*	*G/G+C*	*G/G+M*	*G+C/G+M*
#1	3 (0.82)	2 (0.75)	1 (0.68)	**0.001**	**0.010**	<**0.001**	0.204
#2	1 (0.89)	1 (0.77)	1 (0.56)	0.373	-	-	-
#3	3 (0.57)	2 (0.60)	1 (0.43)	<**0.001**	<**0.001**	<**0.001**	**0.003**
#4	3 (0.19)	2 (0.35)	1 (0.26)	<**0.001**	<**0.001**	<**0.001**	<**0.001**
#5	3 (0.92)	2 (0.49)	1 (0.84)	0.121	-	-	-
#6	3 (0.82)	2 (0.60)	1 (0.68)	**0.001**	**0.003**	**0.003**	0.579
#7	3 (0.57)	2 (0.66)	1 (0.49)	<**0.001**	<**0.001**	<**0.001**	**0.013**
#8	3 (0.88)	2 (0.62)	1 (0.68)	**0.004**	**0.004**	**0.011**	0.686
#9	3 (0.87)	2 (0.64)	1 (0.81)	0.129	-	-	-
#10	2 (0.91)	2 (0.68)	1 (0.56)	**0.008**	0.259	**0.037**	**0.049**
#11	3 (0.92)	2 (0.65)	1 (0.67)	**0.069**	0.340	**0.026**	0.068
#12	3 (0.96)	1 (0.68)	1 (0.55)	**0.042**	0.059	**0.001**	0.281
#13	2 (0.93)	2 (0.71)	1 (0.68)	0.113	-	-	-
#14	3 (0.67)	2 (0.63))	1 (0.56)	<**0.001**	<**0.001**	<**0.001**	0.123
#15	3 (0.67)	2 (0.66)	1 (0.49)	<**0.001**	**0.002**	<**0.001**	0.142
#16	3 (0.73)	2 (0.69)	1 (0.49)	<**0.001**	**0.005**	<**0.001**	**0.024**

followed by the G+C and G.

Limitations

The poor outcomes of the G configuration may be attributed to a hardware limitation of the VR gloves employed in the study. The Manus VR DK used for the experiment, as in the reference work, allowed for 6-DOF tracking through additional tracking elements on the user's wrists. However, the gloves' 3-DOF Inertial

Measurement Unit (IMU) was subject to significant drift over time, resulting in rotational misalignments between virtual and real hands [125]. This issue may have been mitigated in G+C and G+M by the hand controller, which ensured proper alignment of the virtual ES despite hand misalignments.

The questionnaire's open feedback section further highlighted the limited tracking performance of the hand's orientation, with some participants mentioning a slight but visible misalignment of their hands. Additionally, a few participants noted that the ATF was less intense in G+M than in G+C, suggesting that the mockup may have reduced the vibrations. Furthermore, one participant reported relying more on ATF in the G configuration than in the other two configurations, as it was the only form of haptic feedback available during the simulation.

Another potential limitation of the findings is the similarity between the shape of the simulated tool (ES) handle and the hand controller. This limitation was not an issue in the study by [40], where the controller was used only for 6-DOF tracking, vibrations, and a physical trigger for the ES, while the perceived shape was replicated through the 3D-printed shell.

Future Developments

As a future development, it may be worthwhile to expand the analysis to other active tools in order to address the previously mentioned limitation, i.e., that if the simulated tools have significantly different shapes and weights from the controller, the G+C configuration may not have an advantage over the other two.

There is another potential limitation that should be considered, which is related to the choice of VR gloves. The selected product was found to have poor tracking performance in terms of hand orientation, and the developers of Manus may have been aware of this issue, as evidenced by the latest version of the device where the mounting for the HTC VIVE Tracker has been moved to the back of the hand to override the IMU-based orientation. Therefore, it may be beneficial to investigate other gloves as well.

In addition, future studies could expand the current evaluation to other technologies that support VR interactions (possibly through the combination of different consumer devices) and include other relevant scenarios, such as simulating the use of different physical props for active tools or incorporating the ES operation as part of more complex carpentry tasks.

2.4 Considerations and Remarks

The goal of this chapter was to explore challenges related to locomotion and accurate haptic simulation in VR experiences.

In the first study, a testbed which aims to facilitate the comparison of locomotion techniques in large-scale VEs from multiple perspectives is proposed. In contrast to previous works that present ad-hoc evaluations, such as [46], [49], [50], [126], LET-VR integrates and complements various analysis methods and tools reported in the literature. It offers a comprehensive set of objective and subjective locomotion-related metrics to evaluate in the execution of representative tasks. Additionally, unlike previous works that defined alternative evaluation approaches, such as [6], [7], [43], [127], this work introduces a methodology accompanied by a scoring system. This system enables potential users to identify locomotion techniques that best align with a set of user-weighted requirements. A use case is also presented, illustrating how LET-VR can be utilized to choose the best locomotion techniques among four known alternatives for a specific application scenario. Finally, detailed procedure for collecting and processing experimental data are provided, with the aim of promoting reproducibility of results.

In the second study, the combined use of VR and Passive Haptic (PH) interfaces as supporting tools in a formal first responder training course is investigated. A VR Training Scenario (VRTS) was developed in collaboration with a forest firefighting unit in the Piedmont Region, Italy, to aid beginner trainees in learning the use of three firefighting hand tools: the shovel, the McLeod, and the broom. The VRTS was evaluated as a complementary addition to the standard course. In the VR experience, trainees utilized realistic replicas of the hand tools as PH in a safe and repeatable VE enriched with real-time fire simulation, allowing them to practice previously learned concepts. A user study involving 30 trainees was conducted during the course, focusing on the use of one of the hand tools (specifically, the shovel). The findings of the study revealed that the incorporation of the VRTS had a notable advantage in terms of procedural learning compared to the conventional course lessons alone. This allowed trainees to have a better recollection of the safety concepts related to the tool usage. The VR experience assisted the trainees in rectifying their mistakes before the actual examination, which resulted in better performance in the practice exam compared to the control group. However, there was no significant improvement in conceptual learning. Furthermore, the VRTS significantly enhanced the overall learning experience's perceived appearance, in terms of attractiveness and stimulation of hedonic quality.

The third study focused on the simulation of an active carpentry tools by combining a consumer haptic device, specifically VR gloves, with a custom-made mockup of the simulated tool. The result of a previous study showed that this compound solution outperformed a single, fully-fledged haptic product [40] (an hexoskeleton-based haptic glove). The new investigation performs a breakdown study of different configurations of the VR gloves-based setup, aiming to evaluate their usability, fidelity, presence, cybersickness, task load, and task performance in terms of UX. The evaluated configurations included the original setup with VR gloves and a PH-based mockup, a gloves-only configuration, and a configuration with VR gloves and a hand controller used as a prop, which were downgraded from the original setup to assess the impact of reduced setup sophistication, following the methodology used in the previous study.

In the considered scenario, the experimental results indicated that the prop-based configurations were superior to the glove-only configuration in several relevant dimensions such as attractiveness, perspicuity, efficiency, input, comfort, interaction, overall fidelity, PH fidelity (Passive Tactile Feedback, PTF, and Passive Kinesthetic Feedback, PKF), control, and overall preference. Among the three configurations, the original compound setup stood out for its dependability, stimulation, presence, satisfaction, mental demand, frustration, and task control, as well as for certain ad-hoc indicators related to the perception of different phases of the screwing action and the position of the fingers and hand. Furthermore, the original mockup-based configuration outperformed the controller-based one in aspects like attractiveness, perspicuity, stimulation, input, satisfaction, mental demand, frustration, interaction, shape, PH fidelity, and overall preference, as well as on the mentioned indicators related to the screwing action. The only strength of the glove-only configuration, compared to the controller-based one, was the possibility to interact in VR using only the hands, as the use of the hand controller as a prop was perceived as similar to the conventional use of the device as a pure hand interface.

Chapter 3

Voice and Body Language

The work described in this chapter has been formerly published in [128]–[130].

The most instinctive form of interaction for humans is through natural language [131], as it is used unconsciously in almost all aspects of daily life to communicate goals, exchange ideas, and express oneself [132]. In addition to verbal communication, non-verbal communication (e.g., emoticons, eye contacts, hand gestures, and gaze awareness) serves as an essential supplementary means to express emotions, feelings, and complex information in a natural manner [133].

Recently, a increasing number of literature works started to investigate voice as mean of input in VR [134]. This situation is probably related to the significant progress in speech recognition technologies, and to the fact that today's users are more accustomed to speech interfaces [135]. Speech-based interfaces can be either designed to detect simple voice commands, or to perform Natural Language Processing (NLP) to understand the user's intention or desired action, and it can support a wide range of interaction tasks (e.g., selection, manipulation, and navigation, among others) [134].

In this context, it may be interesting to study the possibility to combine the functionalities of different implementations, in order to find new HCI paradigms by mitigating the limitations of the various alternatives. To this purpose, a first investigation of the performance of three speech-based techniques (two from the literature, and one obtained by merging the other two) was performed. The chosen context for this investigation is that of navigation in VR scenarios, a problem closely related to locomotion discussed in Section 2.1. In fact, there is a variety of teleport-based techniques, belonging to the "magical" category, that can be built upon the use of hands-free voice input (Figure3.1). In this study, three different hands-free speech-based navigation techniques for VR were compared. These techniques included a speech-only approach, a speech with gaze variant (where gaze is used to point to the destination and speech is used as a trigger), and the combination of them. These techniques were deployed in a large VR that represented a common training setting (specifically a hangar), and evaluated in terms of usability, performance and general preference.

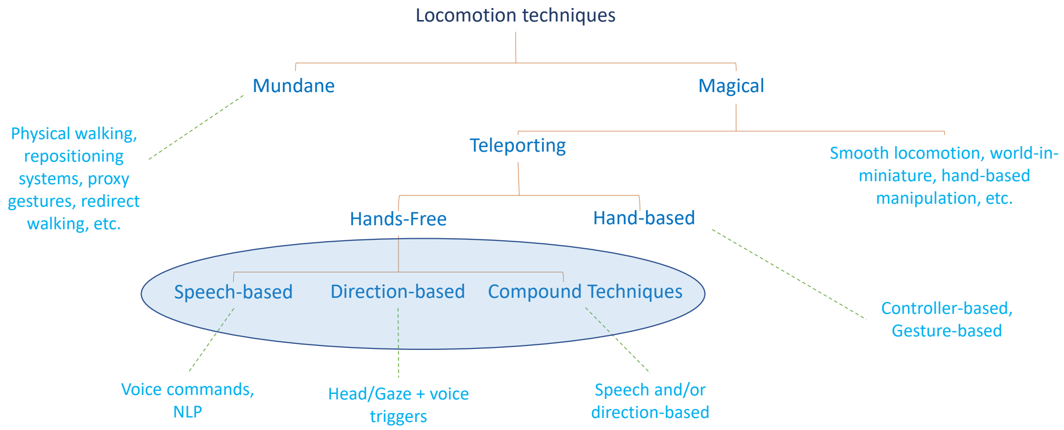


Figure 3.1: Voice-based hands-free navigation techniques in the locomotion taxonomy [21].

A second use of users' voice input is as a mean of communication among different users in a multi-user scenario. In these contexts, the voice usually complements the so-called avatar representation, which represent a common form of non-verbal communication for interactive VEs [133]. Avatars, which can be defined as one's interface to other human users that requires a process of constantly reading and interpreting [136], can vary in terms of visibility, complexity, and style. However, similarly to locomotion (Section 2.1), selecting the most appropriate representation for a given use case can be complex and challenging. In this perspective, two studies investigating two different aspects of avatar representation were carried out.

In the first study, two avatar representation techniques in VR are compared against a relevant training use case. Specifically, a head & hands implementation and a realistic whole-body reconstruction (a blend of inverse kinematics and animations), are employed within a multi-user VR simulation of road tunnel fire scenario. During the experimental activity, participants were paired with a second user (controlled by an experimenter) with which they can establish a communication (by combining voice and body gestures) to collaborate in responding to the emergency. The evaluation covers relevant aspects such as perceived embodiment, immersion, and social presence.

Finally, the third study moves the focus to the expressiveness of the avatar representation, by comparing two real-time modalities for conveying expressions in VR. The realistic, whole-body avatars, which happened to result the best technique in the previous analysis, were used as a basis for the new investigation. The first investigated modality utilized dedicated hardware such as eye and facial trackers to enable mapping between the user's facial expressions/eye movements and the 3D model of the avatar. The second modality relied on an algorithm that approximated facial motion from an audio clip, generating plausible lip and eye movements. In both cases, the voice of the user still plays a fundamental role for the communication. Study participants observed avatars of an actor performing scenes involving basic emotions for both modalities, and the evaluation focused on aspects related to social presence and emotion conveyance.

3.1 Voice-based Interfaces for Navigation

VR is increasingly being adopted across various industries, with training being one of the areas where it is proving to be highly advantageous. With the use of VR, it is possible to create VEs of any size, which can be utilized to simulate diverse scenarios. However, in cases where the VE is larger than the available physical space around the user, it becomes necessary to support additional navigation techniques [41] to supplement natural room-scale locomotion.

A fist solution, already investigated in Section 2.1, is to rely on stationary locomotion paradigms, which tend to take advantage of arm or leg movements or gestures to reduce cybersickness and preserve naturalness. Although these techniques can more or less effective serve their purpose, they are characterized by two critical limitations. Firstly, their nature may dramatically hinder the accessibility of the medium, because they may require to perform gestures assuming a fully able-bodied user (e.g., gait, pointing). Secondly, in case of hand-based techniques, they may impair hand interaction, especially if performed in combination with the locomotion (e.g., swinging arms while manipulating a virtual object).

To address these limitations, several studies have investigated the potential use of *Speech-based* interfaces for teleport-based navigation in VR scenarios [137]–[139]. The combined utilization of speech and gestures as inputs is not a new concept [140], [141]. In the context of immersive VR navigation, voice commands can be employed solely to specify the Point Of Interest (POI) to which the user wants to move (e.g., by stating its name) [137]. Alternatively, phrases like “take me there” can be combined with directional pointing gestures [138]. To enable a completely *hands-free* approach, the requirement for a pointing gesture or handheld device can be circumvented by utilizing the head-gaze orientation [139]. In some cases, the system may consider voice commands ambiguous. In such situations, additional techniques may be necessary to disambiguate the interaction [138], such as providing the user with extra information to clarify their intent.

The objective of the study reported herewith is to compare the performance of three different approaches for hands-free speech-based teleportation in VR. The analysis is conducted on a large indoor scenario that represents a typical industrial VR Training Scenario (VRTS), specifically a hangar [142]. The first method utilizes speech-only commands, where the user utters phrases consisting of the movement action and the name of the desired destination (e.g., "Take me to the yellow bin") [138]. The second technique is a multi-modal approach that combines voice commands to activate the teleportation action with head-gaze direction to indicate the intended destination [139]. The third implementation is a hybrid technique that merges the functionalities of the previous two methods. Each technique is supplemented with a dedicated disambiguation technique to address any potential ambiguities.

3.1.1 Background

The navigation of large VEs is a significant unresolved matter, as evidenced by the numerous recent studies on the subject [93].

Numerous locomotion paradigms have been proposed and widely studied for VR, including *Teleporting* which is considered one of the most popular techniques due to its high intuitiveness, low cognitive demand, and limited impact on cybersickness [22]. However, Teleporting relies on hand controllers commonly bundled with commercial VR systems, which may not be suitable for scenarios that heavily rely on hand interactions or when the user cannot hold a hand controller device. This assumption may not hold true in situations involving hand-tracking techniques [138] or wearable haptic devices without buttons [40].

As stated by authors of Monteiro et al. [134], a significant number of approaches to hands-free interaction in VR have been proposed and researched over the years. The most commonly studied techniques include the use of voice, ranging from simple commands to NLP, as well as eye/head gaze. Less common methods, such as brain activity, facial expressions, foot movement, body position, and muscle contraction, have also been explored.

Navigation is a type of interaction that can be easily facilitated using speech-based interaction. For instance, in Ferracani et al. [137], the authors proposed the use of voice commands to interact with a large VE (a museum). One of the functionalities of this approach was to allow users to move from one room to another, effectively creating a navigation system based on voice commands. The purpose of this approach was to ensure accessibility for users with motor disabilities. However, this approach may encounter deadlocks when users are unable to find the appropriate voice command to convey their intentions.

The author of Sin and Munteanu [138] investigated a multi-modal interaction experience for immersive VR by combining hand-tracking with voice input processed through automatic speech recognition. The proposed system includes four main functionalities. The first is *positioning*, which enables the user to trigger teleporting by saying phrases such as “I’ll go to X,” where X is a particular object or POI. The second is *object identification*, which allows the user to identify a specific POI by pointing at it without specifying its name (“I’ll go there”). The third is *information mapping*, which uses joint input of pointing and voice to add custom labels to objects within the VE. Finally, the fourth is *disambiguation*, which occurs when two or more objects fit the same physical verbal description. In this case, pointing can be used to identify the correct one.

Finally, in Mehra et al. [139], a technique for teleporting based on gaze direction and voice commands was used in the development of IDEs in VR. The trigger command “teleport” was used, while the direction of the head was used as the pointing action, achieving a completely hands-free alternative to the pointing gesture.

Based on these findings, two main categories of speech-based hands-free teleporting techniques can be identified: *speech-only*, which involves using voice commands or NLP to identify the POI, and *direction-based*, which uses voice commands as triggers. Additionally, *compound techniques* combine the functionalities of the previous two. As far as the authors are aware, a comparison of hands-free techniques belonging to the three categories for navigation of large VEs has not yet been conducted.

3.1.2 Materials and Methods

This section outlines the implementation of the navigation system based on voice commands, as well as the training scenario employed for the experimental activity.

Speech Recognition Application

The development of the speech-based navigation system utilized Microsoft Speech Platform Software Development Kit (SDK) (Version 11)[1] for speech recognition. The engine was programmed to recognize and output two elements: an *Intent*, indicating the type of action the user desires, and an *Entity*, representing the object of the action. Intents and entities were defined in a GRXML file, which contained all the possible statements expressing the considered intents. The VR application was connected to the speech engine using a client-server approach via HTTP Request/Response model. For the sake of this evaluation, only the movement intent was considered.

---

[1] Microsoft Speech Platform SDK: `https://learn.microsoft.com/en-us/windows/apps/develop/speech`

Evaluated Techniques

The evaluation focused on three techniques: Speech-only (S), Speech with Gaze (SG), and and Speech with Gaze & Descriptions (SGD), each with a specific disambiguation approach. With S (Figure 3.2), users solely rely on voice input to navigate the VE. If the name of the POI is unambiguous (e.g., "Take me to the green table" with only one green table in the scene), the user is teleported to the desired location. If the POI is specified generically (e.g., "Take me to the table" with multiple tables in the scene), the disambiguation logic manages the ambiguity. If the speech recognition system cannot recognize the requested action or the POI specification, an error is signaled to the user. S uses a disambiguation approach that involves opening a UI element, shown in Figure 3.5a, that displays various possibilities, such as the image of every table, and prompts the user to specify the correct POI with voice commands ("Which table?").

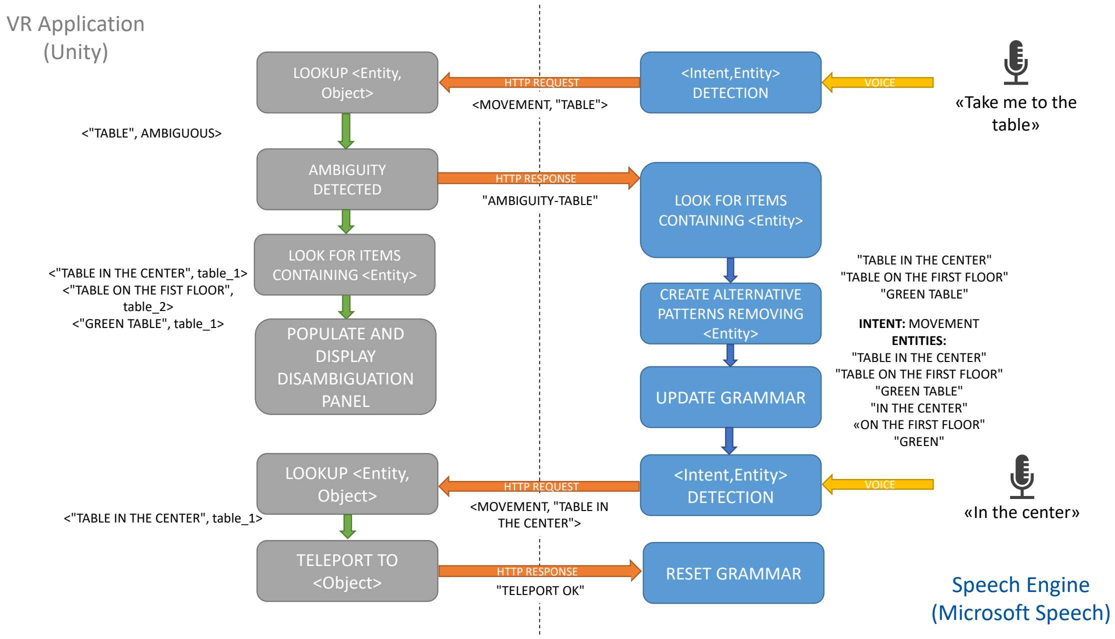


Figure 3.2: Example of workflow of the Speech-only (S) technique.

To handle disambiguation, the VR application processes the received *Intent-Entity* pair using a dictionary that maps entities to objects within the VE. The dictionary allows a single VE element to be associated with multiple entities, which correspond to the different synonyms that can be used to refer to the same element. In addition to specific entities, the dictionary includes several generic entries that trigger the *ambiguity* state. When ambiguity is detected, the UI panel is populated with the possible alternatives and displayed to the user. At the same time, the speech engine is notified of the ambiguity state and the generic entity, to prepare for the following voice recognition during the disambiguation process.

The technique known as SG (Figure 3.3) uses the user's head direction to create a line of sight across the VE that can be used to indicate a specific POI. By keeping the pointer on a particular object, the user can say phrases like "Take me there." If there is ambiguity in the POI selection, such as if the line of sight intersects more than one POI, the disambiguation process is invoked. In this scenario, the user can interact with the UI panel in the same way as with the navigation technique. This means pointing to the correct POI with the head and using expressions like "There." To make selection with the gaze easier, additional graphical elements were added. For example, occluding elements that are not among the possible POIs are automatically made transparent when the ray-cast hits them, and a specific highlight is used to indicate the selected POI, along with an arrow indicator positioned over the object itself.

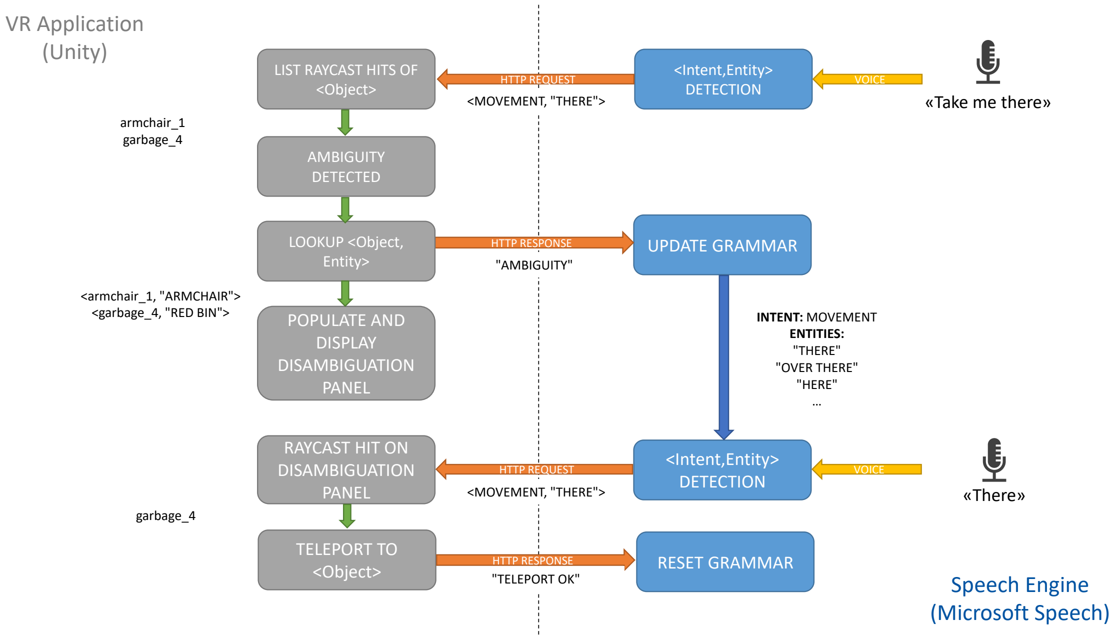


Figure 3.3: Example of workflow of the Speech with Gaze (SG) technique.

The final technique, SGD (Figure 3.4), combines the two previous methods into a single approach. The user has the freedom to use either natural language or gaze direction to identify the desired POI, and the system attempts to resolve any ambiguity by combining the two sources of information before resorting to the disambiguation technique. For example, if the user points to the green table and says "Take me to the table," the system assumes that the intended POI is the green table. If there are further ambiguities, the system presents a disambiguation panel, and the user can interact with it using either the S or SG approach.

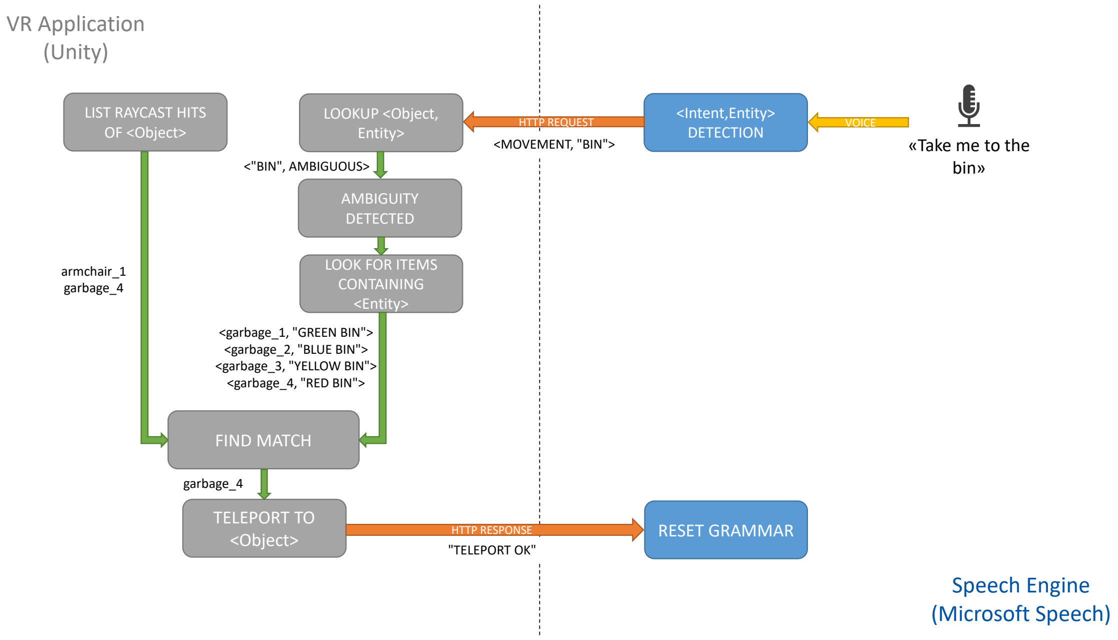


Figure 3.4: Example of workflow of the and Speech with Gaze & Descriptions (SGD) technique.

Test Scenario

The main goals of the test scenario were twofold: firstly, to ensure a level playing field for all techniques by avoiding any situations where one may be disadvantaged and to test the unique features of each technique. As previously stated, the VE was designed to replicate an industrial hangar (based on a simplified version of the Unity asset *FPS Hangar*[2]) and was developed using Unity 2020.2 and the SteamVR framework. Within VE, various virtual objects were placed, some of which were designated as possible POIs and could be accessed through teleportation, while others were treated as contextual elements (such as obstacles).

The experience was structured as a series of individual tasks, each with a randomized order of presentation and preceded by a preparation phase. During the preparation phase, the user was transported to a privileged POV, typically from an overhead perspective, where they could view the starting position (represented by a blue circle) and the POI that they needed to reach in order to complete the given task.

Afterwards, the user was taken to the starting position, and the teleporting technique was activated, allowing them to perform the task. Once the task was

---

[2] FPS Hangar: `https://assetstore.unity.com/packages/3d/environments/industrial/fps-hangar-20040`

completed, the user was returned to the privileged POV to begin preparing for the next task. It is worth noting that users were not permitted to teleport to incorrect POIs or to custom locations. In the event of selecting the wrong POI, an error would be signaled and logged, and after three consecutive errors, the system would provide a suggestion to help the user resolve the situation. A list of the tasks, along with their descriptions, is depicted in Figure 3.5 and detailed in Table 3.1.

Task #	Task	Condition	Target POI	Potential Wrong POI
1	Partially overlapped POIs	Medium Distance	Green bin at the corner	Yellow bin at the same corner
2	Partially overlapped POIs	Long Distance	Yellow bin at the corner	Green bin at the same corner
3	Similar POIs	Short Distance	Chair (closer, on the right, with the pot plant)	Chair (farther, on the left)
4	Similar POIs	Medium Distance	Sofa (on the left, with the ball)	Sofa (on the right)
5	Similar POIs	Long Distance	Sofa (on the right)	Sofa (on the left, with the ball)
6	Similar POIs	Different height	Table at the center	Table at the first floor
7	Visually occluded POI	Short Distance	Red bin	Armchair or any other bin
8	Visually occluded POI	Different height	Blue bin	Any other bin

Table 3.1: Details of the navigation tasks considered in the testing scenario.

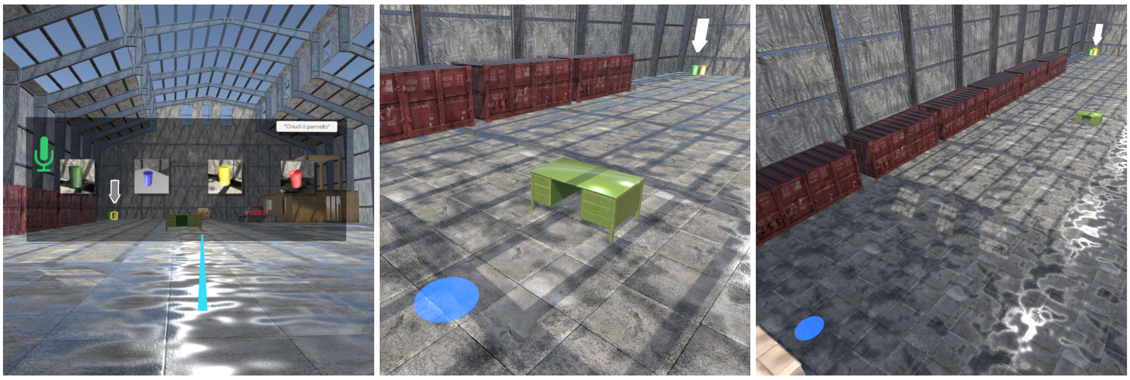
3.1.3 Experiment

A within-subjects user study was conducted, involving 15 participants between the ages of 20 and 30. A Meta Quest 2 headset was employed as VR system, which was connected to a VR -ready laptop via the Oculus Air-Link capability to create a wireless tethered OpenVR system.

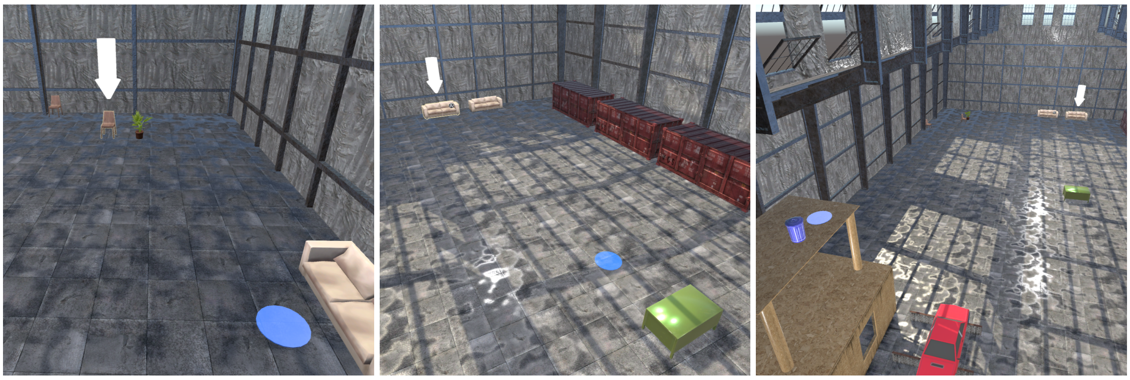
At the start of the experiment, each participant completed a demographic questionnaire that also recorded their previous experience with VR applications and speech interfaces (such as voice assistants). After being briefed on the purpose of the evaluation, each participant was exposed to the three modalities in a randomized Latin-square order. Before each run, the specific technique to be used and the operation of the disambiguation logic were described in detail.

As previously mentioned, the VR experience was divided into various randomized tasks. After completing the VR experience with one technique, the participants were asked to fill out a questionnaire with multiple sections. The first section, corresponding to the Subjective Assessment of Speech System Interfaces (SASSI) [143], was used to evaluate the usability of speech-based interfaces.

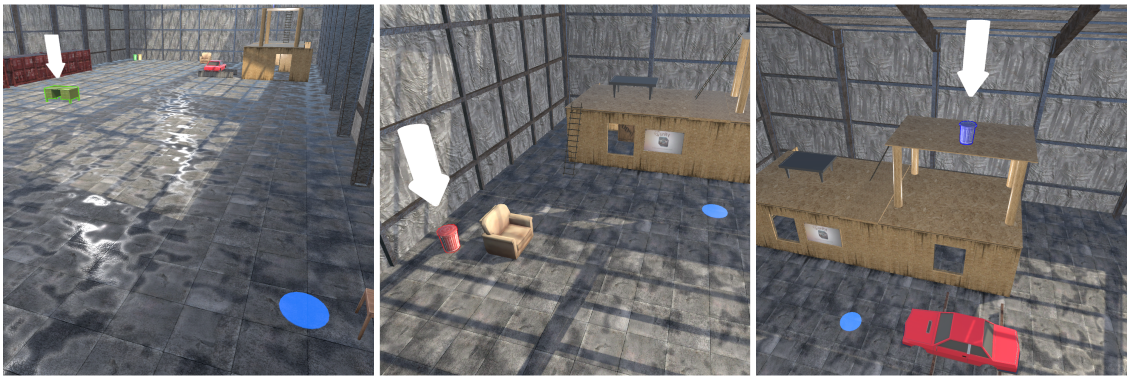
The second section of the questionnaire utilized the SUS questionnaire [105] to evaluate the overall usability of the VR system. The third section consisted of custom questions that addressed aspects not covered by the previous questionnaires, such as the difficulty in reaching POIs under certain conditions. The post-test section required participants to rank the three techniques in order of preference and provide any additional comments. The complete questionnaire in available for



(a) Disambiguation UI (b) Task 1. (c) Task 2.



(d) Task 3. (e) Task 4. (f) Task 5.



(g) Task 6. (h) Task 7. (i) Task 8.

Figure 3.5: (a) Disambiguation panel used to solve ambiguities; (b-h) Navigation tasks implemented for the experiment.

download[3].

---

[3] Experiment questionnaire: https://www.dropbox.com/scl/fi/13fr4zz59h42ah0cporpm/MELECON2022_Questionnaire.docx?dl=0

Objective Metric	**S**(SD)	**SG**(SD)	**SGD**(SD)	***p*-value**	S-SG	S-SGD	SG-SGD
Avg time per destination (s)	8.51(7.06)	7.47(5.16)	8.31(7.52)	.439	-	-	-
Avg errors (wrong POI)	1.62(1.80)	0.06(0.24)	1.31(1.45)	**.001**	**.002**	.595	**.003**
Avg errors (command not understood)	0.68(0.84)	1.00(1.27)	0.93(0.89)	.389	-	-	-
Avg errors (total)	2.31(2.19)	1.06(1.24)	2.25(1.64)	.082	-	-	-

Table 3.2: Outcomes concerning the objective metrics. Statistically significant results are indicated by bold font.

In addition to the subjective measures, the VE provided several objective performance indicators. For each run, the application recorded the average time taken to reach each destination (from the start to the teleport), the number of errors in selecting the requested POI (wrong destinations), and the number of commands that were not recognized by the speech-recognition algorithm (as well as the total number of these two types of errors).

3.1.4 Results and Discussion

The following section discusses the results obtained for the subjective and objective metrics mentioned earlier. To investigate statistical differences, One-Way repeated measures

The objective metrics results are presented in Table 3.2. Out of all the measures considered, only the average number of errors related to selecting the wrong POI showed a significant difference. Specifically, the number of errors was significantly lower in SG than both S (1.62 vs 0.06, $p = .002$) and SGD (0.06 vs 1.31, $p = .003$).

Specifically, the SG technique showed a significant advantage over both S and SGD in terms of the average number of errors related to selecting the wrong POI, with only 1 error recorded in the entire group of 15 participants compared to tens of errors for the other two techniques. This result can be attributed to the fact that SG allows users to point to the teleportation target more clearly and unambiguously with their head, making them less likely to select the wrong POI. In contrast, the use of descriptions (i.e., with S and SGD) means that users may not be fully aware of the outcome of their expression, leading to a higher potential for error.

Figure 3.6 shows the results for the six sub-scales of the SASSI [143], which were rated on a 7-point Likert scale ranging from strongly disagree to strongly agree. The statistical analysis indicated significant differences for three out of the six indicators.

Although the system response accuracy sub-scale did not show significant differences, the efficiency of interaction with the system using S was found to be significantly better than with SG (6.6 vs 5.73, $p = .006$).

In terms of likeability and annoyance, SG was rated significantly worse than S and SGD. The reason for these outcomes could be attributed to the fact that it employed a repetitive interaction scheme, which may have become increasingly

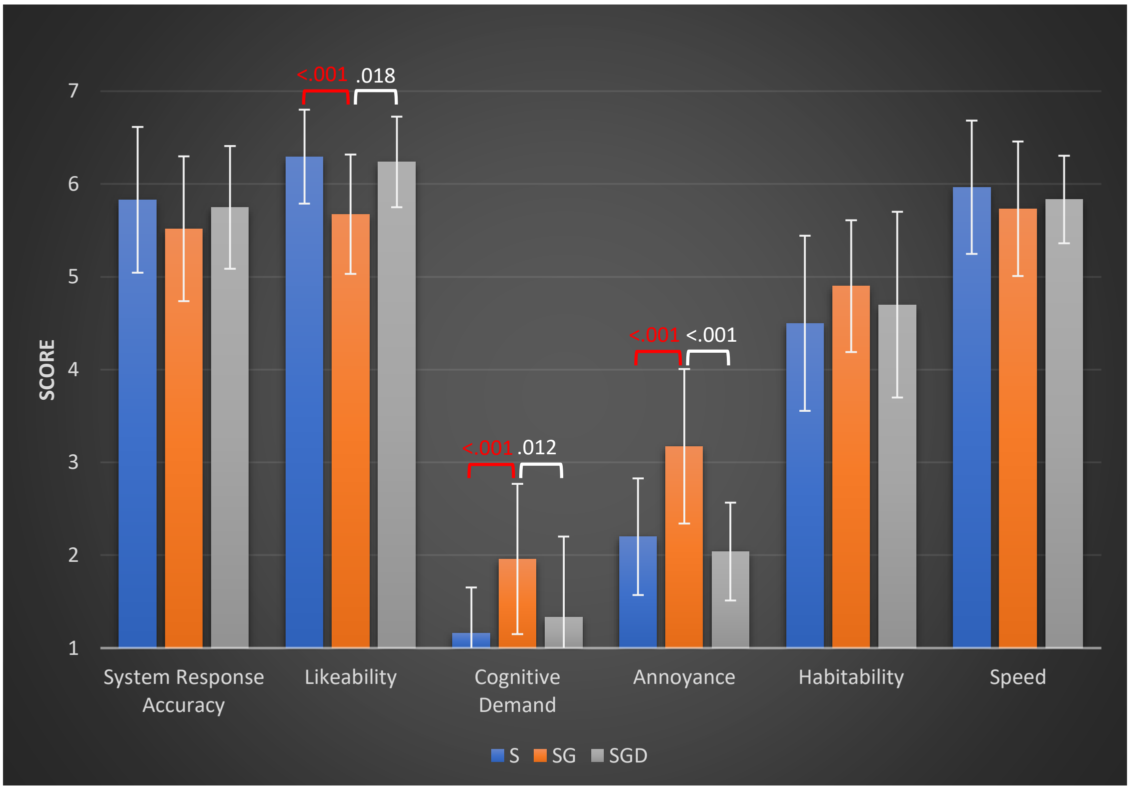


Figure 3.6: Subjective outcomes concerning the SASSI [143]. Statistically significant results are marked with brackets, SD is expressed through bars.

tedious over time despite its higher accuracy.

Surprisingly, SG was found to be more cognitively demanding compared to S and SGD. This result may seem counterintuitive, but it can be explained by examining the questions from this sub-scale. Specifically, participants reported feeling calmer while using S than SG (6.33 vs 5.45, $p = .003$) and that SG required a higher level of concentration than S (1.86 vs 3.46, $p < .001$) and SGD (3.46 vs 2.33, $p = .001$). Although these statements were originally linked to cognitive load, participants also interpreted them in terms of physical demand. In particular, maintaining gaze over the desired POV was probably perceived as requiring more concentration, especially for distant POIs, and this increase in difficulty may have also led to feelings of agitation.

As this sub-scale concerns aspects such as boredom, frustration and inflexibility, this result could explain the the poor likeability of SG too.

Finally, no significant differences were observed for the habitability and speed sub-scales, although participants stated that they sometimes wondered if they were using the right word less frequently with SG than with S (4.46 vs 2.8, $p < .001$) and SGD (2.8 vs 3.13, $p < .001$). This outcome is not unexpected, as the vocabulary in case of SG was much more uniform and contained.

In terms of the SUS [105] section, there were no significant differences observed among the three versions, as all of them received fairly high total score values (between 80 and 90), surpassing the threshold of 71.1 for a *Good* rating.

However, the custom section of the questionnaire, which was rated on a 5-point Likert scale from total disagreement to total agreement and displayed in Figure 3.7, provided more detailed information about the previous results, as significant differences were found.

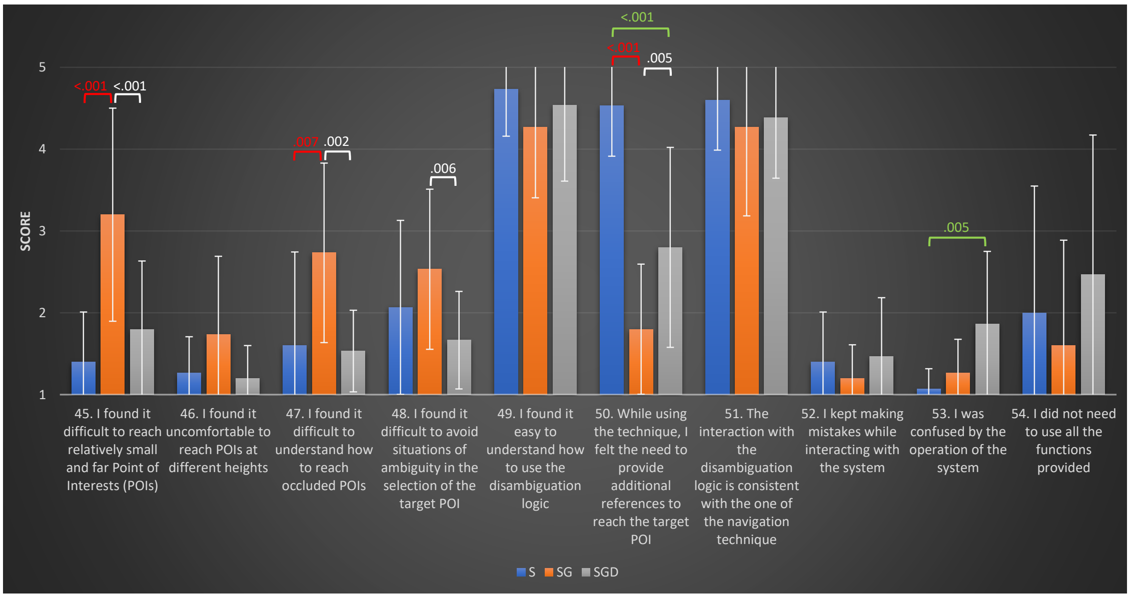


Figure 3.7: Subjective outcomes concerning the custom section of the questionnaire. Statistically significant results are marked with brackets, SD is expressed through bars.

Participants using SG found it more challenging to reach small and distant POIs compared to those using S and SGD (Item #45). This was likely due to the difficulty of maintaining focus on tiny objects before triggering the teleportation by speaking the required phrase. Similarly, SG made it more difficult to understand how to reach obstructed POIs compared to S and SGD (Item #47). As expected, the combination of S and SG features enabled SGD to mitigate the problem of avoiding ambiguities compared to SG (Item #48). SG and SGD allow for the resolution of ambiguity related to multiple POIs hit by the gaze's ray by simply describing the target POI. Participants felt less need to provide additional references to the desired POI using SG compared to both S and SGD (Item #50). SG reduces cognitive load related to the need for descriptions. SGD, by providing an alternative way to address ambiguous situations, mitigates this issue. However, participants also found the operation of SGD more confusing than that of S (Item #53) due to the less uniform UX resulting from the combination of S and SG functionalities.

In relation to the preference ranking shown in Table 3.3, there were notable differences observed between S and SG ($p < .001$) as well as between SG and SGD ($p = .001$). It was observed that S and SGD were frequently selected as either the first or second choice.

Rank	**S**	**SG**	**SGD**
1st	40%	0%	60%
2nd	60%	20%	20%
3rd	0%	80%	20%

Table 3.3: Ranking by preference of the three interfaces: $p = .002$, S-SG ($p < .001$), S-SGD ($p = .812$), SG-SGD ($p < .001$)

The findings indicate that Speech-only (S) and and Speech with Gaze & Descriptions (SGD) are notably superior to Speech with Gaze (SG) in various subjective aspects, such as likability, cognitive demand, annoyance, and preference. However, SG was found to minimize errors, including selecting the wrong POI and misunderstanding commands by the speech engine. This slight improvement in task performance, however, came at the expense of lower efficiency, pleasantness, enjoyability, control, calmness, and higher repetitiveness, boredom, frustration, and inflexibility. Interestingly, the combined technique (SGD) did not offer a significant advantage over S, except for a slightly reduced need to provide additional references for the desired POI. However, this approach also carried a risk of causing more confusion due to the less uniform interaction scheme.

Limitations

Some limitations of the current study emerged from the comments section of the questionnaire. In particular, participants pointed out that the considered use case may have been too simplistic to properly stress the S and SGD techniques, as real-world training scenarios may be characterized by a wider number of similar elements, which would not be as easily distinguishable as the ones considered in the study (e.g., a hangar may contain tens of aircraft of the same model). At the same time, the SG could have been put in higher disadvantage too by considering a wider VE (e.g., a building with three or more floors), or characterized by an higher prevalence of occlusions.

Future Developments

To address the limitations of this study, future developments should focus on broadening the evaluation to include other representative training use cases, such as those with a larger number of less distinguishable elements. Additionally, other forms of navigation, such as teleporting to arbitrary 3D coordinates, should be included in the experience to provide a better challenge to the performance of the considered techniques and relative disambiguation logic.

Additionally, the comparison could be extended to include other speech-based techniques, such as hands-busy implementations, which replace gaze with hand pointing. These variants could be based on hand tracking or the use of a VR hand controller. Furthermore, it would be interesting to broaden the investigation to include other relevant tasks beyond navigation, such as using speech to interact with objects (e.g., selecting tools, displaying and concealing UI elements), performing manual tasks with hands, and any other elements that may impact the performance of the considered techniques.

3.2 Avatar Representation for Multi-User Training Simulations

In multi-user VR scenarios, the representation of avatars plays a pivotal role in shaping the overall UX. Avatars serve as the visual embodiment of users within the virtual world, allowing them to interact and communicate with other participants. The importance of avatar representation becomes even more pronounced in the field of multi-user training, where realistic and meaningful interactions between trainees are crucial for effective learning and skill development. In this context, VR technology has the capability to create highly complex and expansive VEs, making it possible to develop training scenarios that may be impractical or costly to implement in real-life situations [86]. One area that has extensively utilized VR is emergency training, as evidenced by the significant amount of literature exploring the effectiveness of VR in creating realistic scenarios for emergency management [78]–[81], [91]). Among the various applications studied, fire emergencies have been a prominent and extensively researched area [93], [97], [98], [144].

As mentioned in Engelbrecht et al. [86], one of the most significant potential advancements in emergency training using VR technology in the near future is the application of findings from other domains that involve VR experiences with multiple users. Typically, VR kits come with a HMD and two hand controllers, which provide synchronized visual and motor feedback to the user [145]. However, these setups only offer sensory information related to the position and orientation of the user's head and hands, which is insufficient for capturing whole-body motion [146]. As a result, most commercial VR experiences for single users only display virtual representations of the hand controllers, such as in SteamVR Home [147], or floating hands/gloves that align with the user's real hands [145], as seen in Oculus First Steps [148]. Previous research has explored different visibility levels for the user's own avatar in single-user experiences, and found no significant differences in perceived embodiment between fully showing, partially hiding, or not showing a virtual body along with the hand controllers [145], which confirms the choices made in the aforementioned VR experiences.

One effective approach to represent and manage the complexity of situations, such as emergency scenarios, is to opt for a multi-user experience by introducing virtual characters in the simulation [86]. Alternatively, multi-user networked VEs offer a wide range of training and operational support possibilities, where multiple human users can visualize and interact with a shared scenario, practicing teamwork at a large scale [89]. However, a drawback of such multi-user VR experiences is the lack of fidelity compared to solitary VR simulations [86]. For instance, in situations where there is low visibility (e.g., due to smoke in a fire scenario), a first responder would heavily rely on the sense of touch, which cannot be easily replicated with commonly available haptic devices. Additionally, if the visual representation of other human users in the VE does not reach a proper level of credibility, factors such as immersion and perceived realism may be severely affected. Nevertheless, the constant evolution of commercial VR technology introduces new systems, devices, and functionalities that, if properly integrated, may help mitigate the aforementioned issues.

While these simplified avatar representations may seem adequate from a first-person perspective, they may not be suitable for representing the avatars of other users in shared experiences. This is especially relevant in emergency training scenarios, where such techniques could negatively impact the perceived realism of the simulated scenario and, as a result, reduce the efficacy of the training.

An alternative approach to represent a user's avatar in VR involves applying Inverse Kinematics (IK) to reconstruct the body based on head and hand sensor data [146]. From a first-person perspective, these techniques may enhance the user's sense of embodiment through visuomotor correlation [149]. However, they may also degrade embodiment when the estimated pose is inaccurate [150]. In multi-user scenarios, IK techniques require proper management of the users' legs to account for their motion in the VE. This can be achieved through procedural gait generation (e.g., in Dead and Buried [151]) or blending the IK outcome with animations (e.g., in VRChat [152]). By allowing for a full representation of the user's body, these techniques can effectively enhance immersion and embodiment, particularly in multi-user emergency simulations for training purposes with a focus on realism.

The objective of the research reported herewith is thus to investigate the impact of two different avatar representation techniques, namely VR Kit only and Full-Body (FB) reconstruction obtained through blending IK and animations, when used to represent users in a multi-user emergency training experience. The scenario, known as "FréjusVR", was presented in a previous work (Calandra et al. [96]) and already mentioned in Section 2.2. It deptics a virtual road tunnel fire event, and it features multi-user, multi-role, and multi-technology capabilities. The scenario incorporates a serious game that functions both as a training tool for firefighters and as a way to convey correct procedures to the general public.

3.2.1 Background

Currently, there is a significant body of research on the use of VR technology to simulate emergency situations, including fire simulations [40], [83], [89], [153]. Tunnel fires, in particular, have been widely explored in the past, as reported in Lovreglio [83]. However, most of these simulations were not designed for multi-user experiences. To address this gap, in a recent study (Calandra et al. [96]) a multi-user road tunnel fire simulator for training purposes was presented. The simulator has the capability to accommodate different roles, including civilian and emergency operators, and can be used with various VR technologies like consumer VR systems, locomotion TMs/SMs, and motion capture suits. Moreover, it allows for different configurations and incorporates real-time fire spreading logic. Moreover, the simulator offers the option to incorporate FDS data to visualize smoke.

As previously mentioned, the level of realism of avatars used in multi-user collaborative VEs is a crucial aspect, as evidenced by numerous studies investigating the effects of different avatar visualization techniques [154]. For example, Roth et al. [155] presented an experimental method to examine the effects of reduced social information and behavioral channels in immersive VEs using non-realistic FB avatars or mannequins. The study involved comparing physical and verbal interactions executed in both VR and real-life, and evaluating social presence, presence, attentional focus, and task performance. According to the findings, the suboptimal realism of humanoid avatars had a negative impact on social interactions and could potentially decrease performance. However, the authors acknowledged that compensating for the lack of behavioral cues like eye contact and facial expressions might be possible.

The main objective of the study conducted by Kasapakis et al. [156] was to assess the utility of implementing high-fidelity avatars in a multi-user VR -based learning environment. In this study, both educators and students were provided with the ability to access a shared VE through avatars. The educator's avatar was designed with high-fidelity representation, including facial cues and eye motion, and was motion-controlled in real-time by the educator. On the other hand, the student's avatar was implemented as non-anthropomorphic, in order to avoid drawing attention away from the educator. The findings of the study indicated that the use of high-fidelity avatars to represent subjects with significant roles in the simulation could enhance the overall UX for all participants.

A similar approach was taken in a study conducted by Benrachou et al. [157] that focused on the use of avatars in rehabilitation scenarios, albeit employing a different set of technologies. In this study, a Microsoft Kinect sensor was used in conjunction with virtual scenarios to monitor avatars and reproduce human posture failures, with the goal of improving the posture of individuals in various stages of rehabilitation. The results demonstrated that the framework developed had sufficient flexibility and precision, underscoring the potential of avatar representation to play a crucial role in diverse scenarios.

As mentioned in Schafer et al. [158], several studies have highlighted that FB and Head & Hands avatar representations are commonly utilized in current literature. In a broader context, Molina et al. [159] explored the notion of avatar realism that would be satisfactory to users. This line of research has led to the development of a standardized Embodiment Questionnaire [160] and has spurred studies investigating avatar usage in multi-user environments, with a particular focus on factors such as social presence and social interactions [161].

In the context of first-person POV avatar visualization in VR, Lugrin et al. [145] conducted a study to investigate the impact of visual feedback of different body parts on UX and performance in an action-based game. The study examined three different levels of visual feedback on the user's experience: a fully concealed body (excluding the virtual keyboard), a partially visible body (consisting of hands and forearms), and a moderately visible body (comprising head, neck, trunk, forearms, hands, and lower limbs). In contrast to some earlier studies, no significant differences were observed in terms of perceived embodiment among the three alternatives.

Finally, in the realm of avatar movements, research studies such as Parger et al. [146] and Caserman et al. [162] have demonstrated the promising capabilities of IK techniques for estimating the pose of humanoid avatars. Similarly, Gu et al. [163] achieved accurate results by utilizing IK methods on different body parts, specifically the head and hands, while blending together various animations for lower body movements.

However, despite the plethora of works on avatar representation techniques for single and multi-user VR experiences, there is a dearth of investigations into the impact of these techniques in the domain of VR simulations for emergency training.

3.2.2 Materials and Methods

In this subsection, a description of the fire training scenario used in the experimental activity is provided, including its configuration and the customization made specifically for the purpose of the evaluation.

Scenario

The application was created using Unity 2018.4 and the SteamVR SDK, enabling deployment on any OpenVR compatible VR system. Given the expansive nature of the scenario depicted, which involved a road tunnel, additional stationary locomotion techniques (Section 2.1) were incorporated to address the limitations of room-scale movements. Among these techniques, the AS technique was chosen for the experimental activity at hand, as it did not require additional hardware and had been found to outperform other techniques in previous investigations [46], [47].

In terms of multi-user capabilities, a client-server architecture was supported using the legacy Unity High-level Network API (U-NET). The host can be either one of the users or a dedicated non-VR machine, which helps lower the computational load of the two VR clients. To facilitate body-to-body communication, a VOIP (Voice Over Internet Protocol) channel is established between users using the Dissonance VOIP asset for Unity [164], with the addition of two additional U-NET channels, one reliable and one unreliable. Position and rotation updates for network objects occur at a frequency of 60Hz, with interpolation employed to ensure smooth transitions between consecutive updates. The scenario is designed to support different roles, such as civilians, firefighters, and truck drivers, which can be played by real users, NPCs, or deactivated as needed.

The training scenario offered a choice between two different avatars (male or female) for players to select as either civilians or truck drivers, while a generic firefighter avatar was provided for users assuming the role of firefighting operators. Regardless of the avatar chosen in the main menu, the VR application was modified to allow for configuration of one of the avatar representation techniques considered in this study (as shown in Figure 3.8), which are further described below.

The first considered technique, referred to as VR Kit (VK), did not require any specific modifications to the application. For the local user's avatar, the SteamVR *CameraRig* Unity prefab automatically handled the visualization of the VR hand controllers and their real-time synchronization with the physical controllers. As for remote users, 3D meshes representing a generic VR HMD and the two HTC VIVE hand controllers were displayed to show the synchronized position and orientation of the other user's head and hands (as shown in Figure 3.8a and 3.8b).

To implement the other considered technique (FB), the Unity asset FinalIK [165] was integrated into the scenario. This asset provides a VR -oriented FB IK solution called VRIK, which supports both procedural locomotion steps, suitable for micro-movements in place, as well as animated locomotion. VRIK is designed for room-scale movements or faster techniques. However, since the navigation requirements of the tunnel scenario involve relatively fast virtual movements, the procedural locomotion of VRIK was not used, and the animated locomotion was preferred (as shown in Figure 3.8c and 3.8d).

For the local user, VRIK was configured to apply the IK algorithm to the upper limbs, spine, and head at all times. Animated locomotion was automatically triggered when the HMD is moved horizontally on the X and Z planes, and the blending between animations was adjusted based on the movement speed. The Unity Mecanim animation system was used by default to manage the blended animations, which included standing animations such as idle, directional walking, and running. However, animations for movements in a crouched position were not provided by the assets. To address this, a second blend-tree was added to implement crouched movements, and the blending between the two blend-trees was determined based on the user's HMD position, normalized with respect to their height. Since the asset used does not apply IK to fingers, hand gestures such as open hand, fist, and pointing were managed using an overriding layer in the Unity animator. These hand gestures were activated based on the context, for example, showing the pointing animation when the user presses a button, or displaying the avatar with tight fists when the user is walking around using the AS technique.

To achieve a more natural representation of the other user, the standard behavior of VRIK had to be adjusted. Specifically, to eliminate the unnatural AS gesture that occurred during walking, the upper body IK was temporarily disabled by blending it with the full movement animation. This means that when a user triggers an additional locomotion technique, their avatar on the other user's machine will perform a FB walking animation. Once the user stops triggering the AS, the upper body IK is restored, and the animation is again applied to the lower limbs only, maintaining a more natural representation of the other user.

In order to ensure a natural representation of the other user, the standard VRIK behavior had to be modified. Specifically, to eliminate the unnatural AS gesture during walking, the upper body IK was temporarily disabled by blending it with the full movement animation. This means that when the additional locomotion technique is triggered by a user, their avatar on the other user's machine will perform a FB walking animation. Once the user stops triggering the AS, the upper body IK is restored, and the animation is again applied to the lower limbs only, maintaining a more natural representation.

To ensure that grabbed objects align properly with FB animations, the grabbing position is always adjusted to either match the actual controllers (when IK is enabled) or the hands of the other avatar's rig (when AS is triggered and FB animation is displayed).

VRIK facilitates the synchronization of FB IK over the network by synchronizing the position and orientation of the user's CameraRig, including the head and controllers. In order to synchronize the additional functionalities mentioned above, a custom U-NET network component was utilized.

Considered Procedure

As mentioned, the training scenario being considered involves different roles such as civilian, firefighter, and truck driver, each with its own procedures and interactions (some depicted in Figure 3.9). For the purpose of the current evaluation, it was decided to focus on the civilian role for both users. This decision was based on several reasons. Firstly, the civilian role does not require prior knowledge, unlike the firefighter and truck driver roles. Secondly, it maximizes the time during which the two users can see each other and interact together, as they can start the experience inside the same car. Lastly, the civilian role provides the highest level

Figure 3.8: Own mirrored avatars shown to participants prior to the VR experience (a, c) and other user avatars during the VR experience (b, d) of the two evaluated modalities, FB (a, b) and VK (c, d).

of flexibility for executing the procedure.

To ensure optimal visibility, the CFD-based smoke simulation was turned off. Additionally, to mitigate potential confounding factors related to the visualization of the other peer, certain NPC roles, such as firefighters, were disabled, or reduced to aesthetic features visible only at the end of the experience (e.g., the truck driver in the security shelter). Taking into account these modifications from the complete experience outlined in Calandra et al. [96], the summarized procedure can be described as follows:

1. Both civilians start their journey in the same car, traveling from Italy to France, with the car radio broadcasting usual messages for tunnel users.
2. After a certain travel time, during which the car occupants can interact with each other and the passenger can read the security brochure of the tunnel, a burning truck in the opposite lane is sighted. At this point, the driver can choose to apply the brakes and bring the car to a stop at any distance away from the burning truck.
3. Once the car comes to a halt, the two users can interact with the car interior, such as turning off the engine, enabling hazard lights, or exiting the car using the doors' handles.
4. After getting out of the car, the two users can press one of the many SOS buttons placed inside the tunnel to signal the accident. They then decide whether to head to the closer SOS shelter, which is beyond the truck, or turn back and reach a farther shelter. It is worth noting that another car accident involving two civilian cars occurs behind the users' vehicle after they exit, resulting in a second, more contained fire. Therefore, both routes to the shelters will be partially blocked by damaged vehicles on fire.
5. If the users choose the first option, they can take advantage of two SOS niches on both sides of the tunnel, just before the truck, which are equipped with SOS telephones and extinguishers that can be used freely. The users can either attempt to extinguish the main fire or directly run towards the selected shelter. To pass behind the burning truck, the users will need to crouch under a fallen wooden plank, along with other debris from the truck load that has blocked the way to the shelter.
6. If the users decide to head back to the other shelter, they will need to crouch again to pass under a metal rod located near the second accident. In this case, the closest extinguisher and SOS telephone will be inside the shelter. The users will be allowed to return to the tunnel to attempt to extinguish the smaller fire. In this shelter, the users will encounter an NPC character sitting on a bench near a locker containing a first aid kit and some water bottles.
7. In both cases, after opening the door of one of the shelters, getting inside, and making a call for help using the SOS telephone, the simulation ends. If the call was already made from an SOS niche in the tunnel, this step is not needed, and the simulation is terminated promptly.

(a) Driver as seen by the passenger.

(b) Other user pressing a SOS button.

(c) Users opening a SOS niche.

(d) Both users employing a fire extinguisher.

(e) Users reach the nearer shelter by crawling underneath the wooden board.

(f) Users making their way to the distant shelter by crawling beneath the metal rod.

(g) Users opening the shelter door.

(h) Users going into an SOS shelter.

Figure 3.9: Screenshots of the VR simulation taken during the experiments.

The option to extinguish both fires was intentionally disabled, but the users were not informed about this limitation and were given the freedom to attempt using the extinguishers, fail, and then proceed with the evacuation.

3.2.3 Experiment

The study involved 15 participants, consisting of 14 males and 1 female, ranging in age from 24 to 67. The majority of participants reported having moderate to high experience with video games, VR, and multi-player applications, but almost all of them had limited or no experience with serious games designed for emergency training.

During the experiment, two HTC VIVE Pro were used, with one worn by the participant and the other by the experimenter who played the role of the second user. The VR scenario was run on two Intel i9-9820X machines, each equipped with 32GB of RAM and a NVIDIA GeForce RTX 2080 Ti video card.

The experiment followed a within-subjects design, with the avatar representation technique being the independent variable. Participants were first asked to fill in a demographic questionnaire to assess their prior experience with the technologies involved. They were then introduced to the experiment and provided with instructions for the simulation. A sample footage of the experiments with both modalities is available for download[4].

Each participant experienced the two representation techniques in a randomized order. Prior to starting each simulation run, they were asked to select one of the available avatars for the civilians within the VR application. Following that, a mirrored version of the avatar with the currently used technique was displayed to the user, providing a preview of how they would be seen from the perspective of the other user. In the case of VK, the choice of avatar was not relevant as it would not be displayed as a virtual body, but rather as 3D meshes representing the VR equipment.

Subsequently, for both techniques, participants were asked to select the role of the driver. At this stage, a second user, controlled by one of the experimenters, connected to the multi-user session and appeared as a second civilian sitting beside the driver, automatically initiating the experience.

In order to ensure consistent experiences among participants and minimize learning effects, the experimenter followed predefined scripts during the simulation. For the first run, the experimenter:

1. Demonstrated interactions while being observed by the participant.
2. Suggested heading to the closer shelter beyond the truck.
3. Ensured that the participant noticed and interacted with the SOS niches.
4. Showed the participant how to navigate past the main fire and obstacles that required crouching.
5. If the participant attempted to extinguish the fire, encouraged them to give up after a while and proceed to the shelter.

During the second run, the experimenter:

1. Suggested turning back and heading to the farther shelter beyond the car accident.
2. Showed the participant how to navigate past the main fire and obstacles that required crouching.
3. After reaching the shelter, suggested equipping the extinguisher and attempting to deal with the smaller fire encountered on the way.

---

[4] Sample footage of the VR experience used for the experiment: `https://www.dropbox.com/sh/yeibnqtx34iatex/AAA_EftwHvkshSEB0LKWZn0Ka?dl=0`

4. After a brief attempt, suggested heading back to the shelter and waiting there for the rescue team to arrive.

It is worth mentioning that these guidelines were not strictly followed, as some interactions could be repeated, omitted, or executed in a different order based on the participant's level of collaboration during the simulation. After each run, the participant was asked to complete the evaluation questionnaire, which can be accessed online[5].

The questionnaire used in the study had four sections. The first section asked participants to complete the Embodiment Questionnaire [160], which evaluated their level of embodiment based on factors such as body ownership, agency and motor control, tactile sensations, location of the body, external appearance, and response to external stimuli (similar to Lugrin et al. [145]). The second section involved the Networked Minds Social Presence (NMPS) questionnaire [161], which aimed to assess the virtual representation of the other user's avatar in terms of mutual awareness, attentional allocation, mutual understanding, behavioral interdependence, mutual assistance, and dependent actions (with the exception of the empathy category, which was not relevant to the study's use case). The third section included the immersion and presence sections of the VRUSE questionnaire [67]. Finally, the last section, completed after both simulation runs, asked participants to express their preference between the two representation techniques in terms of usability, aesthetics, multi-player interactions, and overall satisfaction.

3.2.4 Results and Discussion

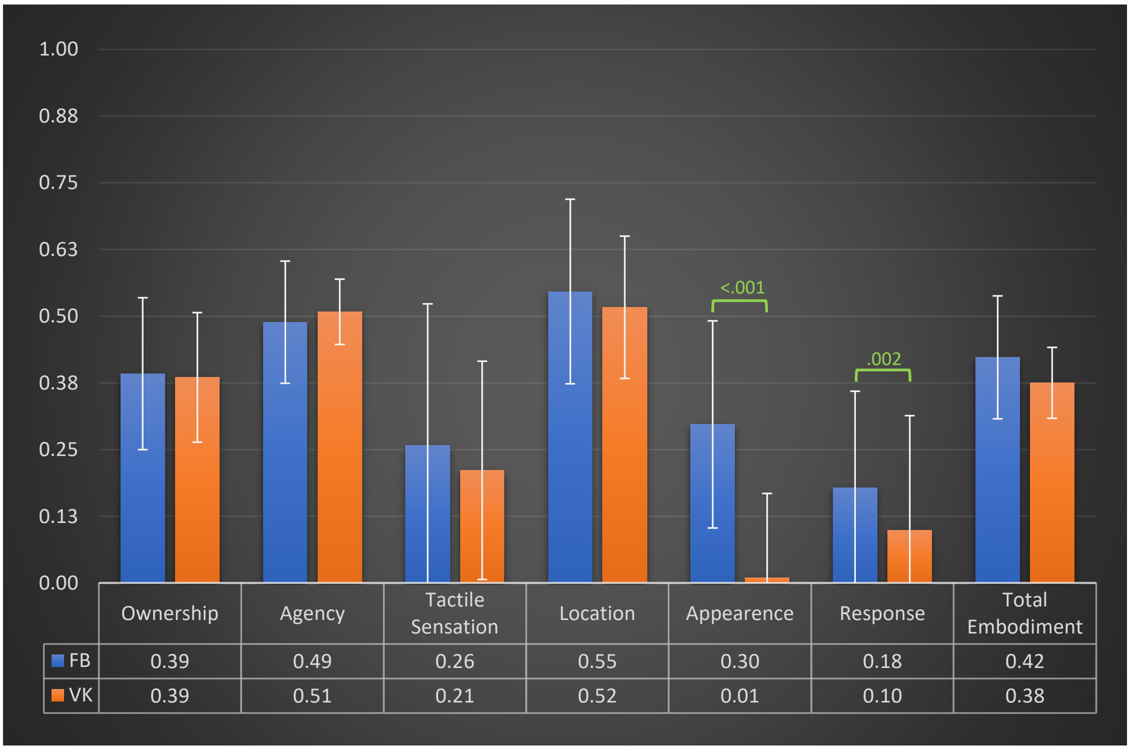
The subjective metrics results from the previous section were utilized to compare the VR Kit (VK) and Full-Body (FB) techniques. The normality of the data was analyzed using the Shapiro-Wilk test, which revealed that the data were not normally distributed. As a result, the non-parametric Wilcoxon signed-rank test with a significance level of 5% ($p < .05$) was employed to examine statistical differences.

3.2.5 Embodiment

Regarding the perceived embodiment, as shown in Figure 3.10, there were no significant differences observed in most of the sub-scales and the total embodiment, which were normalized in a range between 0 and 1 as suggested in Gonzalez-Franco and Peck [160]. It should be noted that each sub-scale originally had different minimum and maximum values. However, there were two exceptions: the appearance and response to external stimuli sub-scales, where the FB technique was perceived

---

[5] Experiment questionnaire:
`https://www.dropbox.com/s/8ljbzh2ckfpj226/Questionnaire.pdf?dl=0`



	Ownership	Agency	Tactile Sensation	Location	Appearence	Response	Total Embodiment
🟦 FB	0.39	0.49	0.26	0.55	0.30	0.18	0.42
🟧 VK	0.39	0.51	0.21	0.52	0.01	0.10	0.38

Figure 3.10: Subjective outcomes concerning the Embodiment Questionnaire [160] sub-scales (values scaled to a range between 0 and 1). Statistically significant differences ($p < .05$) are marked with brackets, and error bars are used to indicate the SDs.

as better than the VK technique. This finding is consistent with previous literature that did not find significant differences between displaying or not displaying the user's avatar body [145], as mentioned earlier.

Delving into the details of the scores assigned to the individual items, which were expressed on a 7-point Likert scale ranging from -3 to 3 (from strongly disagree to strongly agree), participants perceived the FB technique as a better representation of their body compared to the VK technique (1.4 vs 0.27, $p = .042$). However, they also reported a higher sense of having more than one body with the FB (0.73 vs -0.6, $p = .001$), indicating that the accuracy of the virtual representation could have either positive or negative effects on embodiment, in line with previous studies [149], [150]. This finding is supported by the results from another item, which revealed that participants felt that movements of the virtual representation influenced their real movements more with the FB compared to the VK (-0.13 vs -1.13, $p = .048$). Additionally, during the initial part of the experience where a mirrored version of the user's avatar is displayed, the FB technique was perceived as the participant's own body more than the VK technique (0.73 vs -0.6, $p = .041$).

Regarding external appearance, participants felt that their real body was becoming more like an "avatar" body with the FB technique compared to the VK technique (0.8 vs -0.6, $p = .003$), and that their real body was taking on the posture of the avatar body more (0.2 vs -1.13, $p = .009$). Furthermore, they reported that at some point the virtual representation resembled their real body more in terms of shape and other visual features with the FB technique compared to the VK technique (0.13 vs -2.0, $p = .003$), and that they felt more like they were wearing different clothes compared to when they came to the laboratory (0 vs -2.26, $p = .005$).

In terms of response to external stimuli, participants felt that virtual elements such as fire and objects could affect them more with the FB technique compared to the VK technique (1.2 vs 0.33, $p = .015$), and they had a higher feeling of being harmed by the fire with the FB technique compared to the VK technique (-0.07 vs -0.53, $p = .015$).

Regarding social presence, the metrics of which are illustrated in Figure 3.11, the disparity between the two techniques was more prominent. Participants provided responses on a 7-point Likert scale, ranging from 1 to 7 (from strongly disagree to strongly agree).

When considering the various sub-scales, participants perceived the VK as superior to the FB in terms of mutual awareness, mutual attention , and mutual understanding. These findings suggest that utilizing a body to represent the other user's avatar significantly enhances cooperation among multiple users. Specifically, with the FB, participants were more attentive to the presence of the other peer (1.33 vs 3.0, $p = .001$), had a higher sense of self-awareness within the VE (6.4 vs 4.93, $p = .002$), and perceived the other peer as being more aware of them (6.33 vs 5.4, $p = .023$) compared to the VK. Furthermore, with the FB, participants felt less lonely (1.2 vs 1.93, $p = .046$), and perceived the other peer as less lonely (1.4 vs 2.2, $p = .039$). The FB also facilitated higher attention towards the other peer compared to the VK (6.33 vs 5.47, $p = .031$). Additionally, with the VK, participants tended to ignore the other individual more than with the FB (1.67 vs 2.73, $p = .011$). Finally, the FB allowed participants to express their opinions more effectively than the VK (6.00 vs 5.27, $p = .002$), and facilitated a better understanding of the opinions of other peers (5.93 vs 5.13, $p = .048$).

The findings for the immersion and presence section, as adapted from [67], are presented in Figure 2.22. The scores, which were rated on a 5-point Likert scale ranging from 1 to 5 (from strongly agree to strongly disagree), revealed that the FB was perceived as superior to the VK only in terms of immersion (4.86 vs 4.26, $p = 0.031$), with no significant difference observed in terms of presence. This result may be attributed to the training experience, which emphasized realism and benefited from a more realistic-looking avatar representation provided by the FB.

The last section of the questionnaire, with the findings presented in Table 3.4, asked participants to express their preference between the VK and the FB for

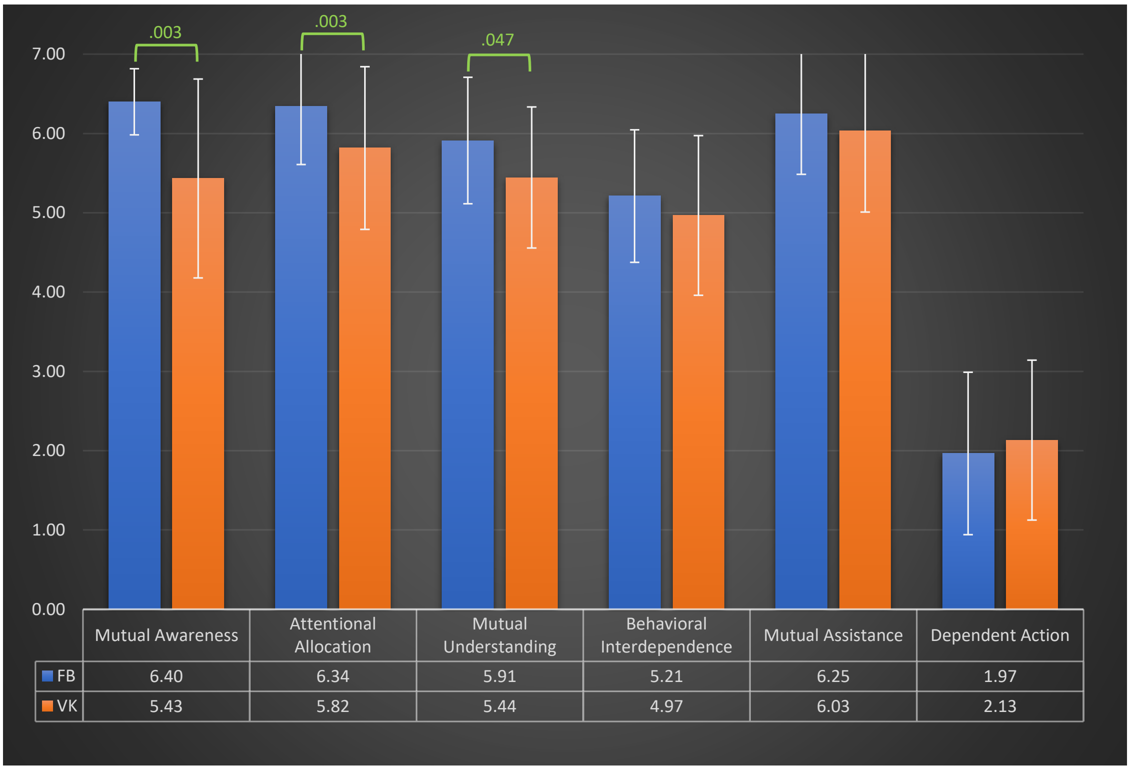


Figure 3.11: Subjective outcomes concerning the NMPS questionnaire [161] subscales (7-point Likert scale from strong disagreement to strong agreement). Statistically significant differences ($p < .05$) are marked with brackets, and error bars are used to indicate the SDs.

different aspects.

Table 3.4: Subjective outcomes concerning the direct comparison section of the questionnaire. Statistically significant differences ($p < .05$) and the best configuration are highlighted with a bold font.

Item	**Question**	**Preference**		p-value (SD)
		VK	*FB*	
#1	Regarding the usability (own avatar), I preferred the version...	53%	47%	.846 (0.5)
#2	Regarding the aesthetics (own avatar), I preferred the version...	40%	60%	.524 (0.49)
#3	Regarding the aesthetics (other avatar), I preferred the version...	7%	**93%**	**.001** (0.25)
#4	Regarding the multiplayer interactions, I preferred the version...	13%	**87%**	**.01** (0.34)
#5	Regarding my own avatar, I preferred the version...	53%	47%	.847 (0.50)
#6	Regarding the other avatar, I preferred the version...	13%	**87%**	**.01** (0.34)
#7	Regarding the overall experience, I preferred the version...	33%	57%	.276 (0.47)

Interestingly, the FB was perceived as significantly better than the VK for aesthetics, multi-player interactions, and overall preference. However, neither of the two techniques prevailed when it came to the representation of the own avatar. These mixed results for the own avatar are consistent with the findings related to embodiment, suggesting that the impact of the two avatar representations may be similar in multi-user experiences as well.

Future Developments

In the future, there are several potential developments that could be explored to further investigate avatar representation in multi-user training simulations. This could include expanding the investigation to consider other avatar representation techniques beyond the ones already studied, as well as incorporating additional technologies such as motion capture suits and leg sensors, as well as different locomotion modalities like locomotion TMs/SMs or WIP paradigms, which are already supported by the training tool under consideration.

Furthermore, more complex scenarios could be developed, such as tasks that require collaboration between users, such as handling objects that require coordinated effort, or simulations involving more than two users. To support these scenarios, the set of animations used in the FB implementation may need to be expanded to include additional movements like prone crawling, jumping, climbing, or other collaborative actions that may be relevant in an emergency scenario.

Additionally, machine learning-based avatar representation techniques could be included in the comparison, to evaluate their impact in terms of computational load compared to the IK algorithms.

3.3 Emotion Conveyance in Social VR Avatars

In recent times, the emergence of VR technologies has spurred numerous studies examining their effects in various domains [166]. One of the most rapidly adopted areas of application are social VR and collaboration, which has gained even more traction due to the ongoing global pandemic [167].

In the context of the emerging concept of metaverse, researchers are actively exploring methods to represent users' avatars in shared VEs. Avatars serve as the interface through which users' movements and expressions are transferred into the VE [168]. Avatar representations can vary in several aspects, such as their realism [169] or cartoon-like appearance [170]. They can also differ in terms of complexity [158], ranging from minimalist visualizations that only show the user's head and hands for ergonomic purposes [171], to more realistic configurations that depict the entire body from head to toe [169], which may also include facial expressions [167]. Research on avatars and their representation is still in its early stages, with limited studies focusing on measuring the psycho-sociological and perceptual impact of avatar representation, especially in VR applications [10]. However, some studies have shown that users tend to prefer avatars in VEs to be realistic, Full-Body (FB), and capable of facial expressions [167], [172]. Nonetheless, capturing facial expressions can be challenging, particularly when using a HMD for VR, although there are emerging devices in the market that support this functionality due to the importance of social cues for conveying emotions [173]. Another alternative approach for reproducing facial expressions in such situations is through lip synchronization (or simply lip-sync), which generates synchronized animations on a speaking avatar [174].

Based on these premises, the study discussed herewith compares two modalities for conveying emotions using realistic FB avatars in a multi-user VR -based scenario. The two modalities were implemented using currently available consumer-level hardware (HTC VIVE Facial Tracker for the first modality) and software (SALSA Lip Sync Suite v2 [175] for the second modality). The study involved 28 participants who wore an HMD and observed an avatar in a VE while playing six scenes, each representing one of Ekman's basic emotions [176]. Each scene was played twice, once per modality, with a pseudo-randomized order of exposure, aiming to create a context as similar as possible to real social VR scenarios. After spectating a scene in both modalities, participants blindly evaluated them in terms of relevant social presence aspects such as comfort, expressiveness, realism, naturalness, pleasantness, and emotion conveyance through direct comparisons.

3.3.1 Background

Numerous studies in the literature have investigated the impact of avatar appearance in multi-user XR scenarios. Researchers have explored various characteristics of avatars, including aesthetic traits, skills, movements, and behavior, and their influence on relevant aspects such as interpersonal communication, non-verbal communication, advertising, and more [154], [177]–[184].

Roth et al. [155] conducted a study on immersive VEs with FB humanoid avatars to investigate the effects of reducing social information and behavior channels. Their findings revealed that the lack of realism in avatars hindered and limited social interactions. However, the absence of facial expressions and gaze analysis in their study limited its generalizability. Similarly, Dobre et al. [172] examined the realism of humanoid avatars in a MR telepresence scenario (i.e., online meeting). They compared realistic and cartoonish FB avatars and found that the nonverbal behavior of the realistic avatar was perceived as more appropriate for interaction and more useful for understanding others compared to the cartoonish avatar.

Yoon et al. [169] conducted a study on the impact of avatar appearance on social presence in AR, examining three levels of body parts visibility (head & hands, upper-body, and whole-body). Their findings revealed that a realistic whole-body avatar was perceived as the most effective setup for remote collaboration, in accordance with the investigation detailed in the previous section (Section 3.2).

Kasapakis and Dzardanova [156] explored the inclusion of facial and eye tracking in a multi-user VR learning environment, where a high-fidelity avatar of an educator was used. The avatar was designed with facial cues, eye and body motion recorded in real-time using various equipment such as VIVE Pro HMD, VIVE Trackers for pelvis, hand, and feet tracking, ManusVR Gloves, Pupil Labs Eye Tracking system, and BinaryVR for mapping facial expressions to blendshapes. The reported results indicated that the combination of body and eye movement, along with facial cues of the educator's avatar, helped 90% of the students to maintain their attention during the lecture and enhance their understanding of the concepts presented.

Hube et al. [174] expanded on their previous work [185] to investigate the impact of facial visual parameters on virtual avatars in VR. They used a predetermined set of audio files and extracted the emotions represented in them, then applied real-time lip sync approximation using SALSA Lip Sync Suite to display this information on a virtual avatar in a VE. The experiment involved participants being placed in an immersive VE where they observed a half-body realistic-looking avatar exhibiting behavior synchronized with the associated audio, and this visualization was compared with an audio-only variant where the avatar remained static. The results showed that the inclusion of additional visual parameters on the avatars' face was beneficial in determining the emotions conveyed in the audio clip. However, the study did not consider the use of body language to enhance the expression of emotions, and only a limited number of non-verbal cues were evaluated, despite the importance of such cues in identifying specific emotions in real-world scenarios.

Based on the literature review, it appears that using FB, realistic-looking avatars is the most effective representation technique in various scenarios. Additionally, the presence of facial cues, such as facial expressions, tends to enhance the User eXperience (UX) compared to expressionless avatars. Different approaches have been explored to implement facial cues, including facial and eye tracking, and lip sync approximation. However, a comparison of these two approaches applied to a FB animated avatar in the context of a multi-user social VR scenario was not been conducted yet.

Based on this, the following hypothesis was formulated for the investigation. Overall, the use of facial tracking technology is expected to better convey the emotional content of users' actions compared to lip sync approximation, which only focuses on the lower part of the face. However, in real-life social VR usage, facial expressions cannot be separated from voice and body movements, and in some situations, facial expressions may not be the primary means of conveying emotions. Therefore, for certain emotions, a less pronounced difference between the two modalities may be anticipated.

3.3.2 Materials and Methods

The VE was developed in Unity 2021.3 as an OpenXR application and was experienced using an HTC VIVE Pro Eye[6] VR kit. The basis for both representations studied was a realistic-looking, FB avatar implementation described in Section 3.2. The VRIK module of FinalIK [165] was used to generate plausible FB motion based on the position and orientation of the HMD and hand controllers, using IK for the upper body and a blending of animations for walking. The user's voice was captured through the microphone on the HMD.

Evaluated Configurations

For the first modality being studied, referred to as Facial+eye Tracking (FT) hereafter, the selected HMD, which was already equipped with eye tracking, was supplemented with an HTC VIVE Facial Tracker[7] to enable facial tracking (Figure 3.12a).

(a) (b)

Figure 3.12: VR setup selected for the experimental activity (HTC VIVE Pro Eye + Facial Tracker) (a). Close-up of the user's avatar (b).

The reasons for selecting this configuration are multifaceted. Firstly, the HTC VIVE ecosystem is highly regarded by social VR users as it allows for the use of additional tracking devices to achieve FB reconstruction [186]. Secondly, the HTC

---

[6] HTC VIVE Pro Eye: `https://www.vive.com/us/product/vive-pro-eye/overview/`

[7] HTC VIVE Facial Tracker: `https://www.vive.com/us/accessory/facial-tracker/`

VIVE Pro Eye comes with built-in eye tracking capabilities, and it can be seamlessly integrated with the HTC VIVE Facial Tracker to add the desired functionalities, making it a convenient and readily available consumer solution for this purpose.

This configuration enabled real-time mapping between the user's facial expressions and eye movements, and a given mesh with eye and facial rig (such as blendshapes). In this case, the mesh used was the FB avatar, which was created using Autodesk Character Generator[8].

The character models generated using Autodesk Character Generator come with an automatic full rig for both the body and face, making them easy to integrate with FinalIK. They are also largely compatible with the Vive Tracker facial rig, with the exception of the "frown" blendshape, which had to be manually added in Autodesk Maya by merging other blendshapes.

The second modality, referred to as Lip-sync Estimation (LE), utilized real-time algorithms instead of additional hardware modules to generate facial motion from an audio clip. To achieve this, the SALSA LipSync Suite v2 [175] asset was employed, following a similar approach as in Hube et al. [174]. The *One-Click* automatic configurator for Adobe Character Generator was also used to ensure smooth integration with the humanoid mesh, along with the *Eyes* module for generating random eye motion.

The choice of SALSA as the lip syncing approximation was based on its common usage in providing real-time facial animation for talking avatars in Unity-based VR experiences. SALSA strikes a good balance between cost for the end-user (no additional hardware required) and performance (such as the option to animate the upper face randomly or leave it static). Furthermore, SALSA has been previously investigated in literature on this topic, adding to its suitability for the study.

Test Scenario

To assess the contribution of the studied modalities in terms of conveying emotions, a per-emotion analysis was conducted, following a similar approach as in Hube et al. [174]. In the previous study, the focus was on three main emotions (happiness, sadness, and anger), while the current study aimed to expand the set to cover all six of Ekman's basic emotions [176] (anger, happiness, sadness, fear, disgust, and surprise).

For each emotion, participants were shown an actor performing a short scripted scene designed to evoke and maintain the specific emotion being portrayed. The scripts were written with consideration of the scales of positivity and value commonly used in studies on perceived emotions [187], and efforts were made to make the scenes as "social" as possible, while keeping them under one minute in duration. An excerpt from the scenes that participants viewed and rated can be seen

---

[8] Autodesk Character Generator: charactergenerator.autodesk.com/

in Figure 3.12b. The original full scripts in Italian, their English translations, and recordings of the six scenes are available for download[9].

To mitigate potential bias arising from repeated viewing of the same scenes, a decision was made to avoid using live performances of an actor in a real multi-user scenario. Instead, a simulated scenario was chosen. The actor's performance was pre-recorded, and an avatar controlled by these recordings was presented to participants, ensuring a consistent experience across all conditions. In this study, a single amateur actor performed all six scenes to maintain consistency in the stimuli presented to participants.

The recorded inputs for driving the representation of each user's avatar in a real multi-user scenario included voice (microphone), movements (HMD and controllers), and facial expressions (blendshapes). FinalIK was used to drive body motion based on head and hand movements for both modalities being evaluated. In the FT modality, all recorded data were synchronized and played back simultaneously, creating a result akin to an actor performing in real-time in a multi-user scenario. For the LE modality, the same recordings were used, but facial blendshapes and eye motion data were discarded. Instead, SALSA was used to approximate lip sync while retaining the same body and voice input. This approach ensured that body movements were kept identical between the two conditions being evaluated in this study.

3.3.3 Experiment

The experimental activity was conducted as a within-subjects user study, involving a sample of 28 participants (16 males, 12 females) aged between 21 and 68 years. Participants were recruited from the students and staff at Politecnico di Torino. The same hardware used for recording the six scenes was also used for the experiment.

Initially, the participants were asked to complete a brief demographic questionnaire that included questions about their age, gender, general experience with VR, and experience with multi-user social VR /metaverse applications. A significant portion of the sample reported limited experience with VR technologies, and the majority indicated no prior experience with social VR or metaverse applications. Specifically, approximately 61% of the sample rated their experience with VR technology as 2 or below on a scale of 1 to 5, where 1 represented almost no experience with VR and 5 indicated daily use of the technology. Similarly, results showed that approximately 78% of users reported almost never using social VR /metaverse experiences.

---

[9] Scene scripts and recordings: `https://www.dropbox.com/sh/l03llheyvmavvwg/AABufS_cAjSveWUP4DV7ttJma?dl=0`

The participants were informed that the study aimed to investigate the expressiveness and communication ability of avatars, particularly focusing on the upper body. The upper body is where the three main communication factors that play a role in real-world conversations, namely tone of voice, arm and hand gestures, and facial expressions, are more prominent.

Following the introduction, participants were given the HMD and were invited to enter a VE that depicted a theater[10]. In this VE, they were able to view the recorded avatar of the actor. Throughout each scene, participants had the freedom to move around the stage area, using natural walking, to observe the gestures and expressions of the avatar from their preferred position (as shown in Figure 3.13).

Figure 3.13: The stage area within the VE where participants (the white HMD) can walk to spectate the avatar during the acting.

To recreate a realistic social VR environment, freedom of movement was provided to participants, allowing them to experience the scenes as they would in such an environment. A Latin square order was used to balance the exposure of the six scenes representing different emotions. Each scene was played once per modality, in a randomized order, so participants were not aware of which modality they were experiencing. After watching a pair of scenes depicting the same emotion, participants were asked to answer a set of questions to measure various aspects such as comfort, expressiveness, realism, naturalness, likelihood, pleasantness, emotional

---

[10] Madame Walker Theatre:
https://sketchfab.com/3d-models/madame-walker-theatre-98ba4154bbb644bb9cb4d9c68d7dd87b

conveyance, and overall preference. This process was repeated for all six emotions. To avoid ambiguity in questions related to comfort and pleasantness for negative emotions (e.g., disgust, anger, and sadness), participants were instructed to approach the scenes with a neutral and detached POV towards the expressed emotion. Finally, participants were given the opportunity to provide open feedback in the form of comments. The full questionnaire is available for download[11].

3.3.4 Results and Discussion

The findings from the experiment are presented in Figure 3.14.

The normality of the data was assessed using the Shapiro-Wilk test. As the data were found to be non-normally distributed, the non-parametric Wilcoxon signed-rank test was used with a significance threshold of 5% ($p < .05$).

Regarding the *Sadness* scene, Figure 3.14a presents the results in terms of M scores and SDs. In this case, the FT modality was consistently rated as significantly superior to the LE modality across all eight dimensions that were assessed through the questions posed to the participants. Specifically, the FT modality was perceived as more expressive in terms of avatar's general expressiveness and conveyance of emotions. Similar trends were observed for overall pleasantness, feeling of being at ease, realism, and overall preference. Although the differences were less pronounced, they were still statistically significant for naturalness of expressions and likelihood. These findings are in line with expectations, as sadness is primarily conveyed through eye and mouth movements according to Ekman's research [176], and the FT modality was observed to provide more accurate output compared to the approximation offered by the LE modality during the development stages.

In the *Disgust* scene (Figure 3.14b), participants showed a significant preference for the FT modality over the LE modality, although not as pronounced as in the previous scene. Participants rated the FT modality as more realistic, more natural, more likely overall, more pleasant, more comfortable, and overall preferable compared to the LE modality. However, there were no significant differences in expressiveness and emotion conveyance between the two modalities. This could be due to the impulsive nature of disgust and surprise, which are often expressed quickly compared to other emotions like anger, sadness, and happiness that may last longer. As a result, differences between the modalities may be harder to notice, especially if participants are not fully focused on the other person's face when the emotion is impulsively expressed, as reported by some participants in the open feedback section.

In the *Fear* scene (Figure 3.14c), the FT modality was perceived as superior to the LE modality in four dimensions. Specifically, participants rated the FT

---

[11] Experiment questionnaire:
`https://www.dropbox.com/s/ndin47lxezst3rr/Questionnaire.pdf?dl=0`

modality as more natural in terms of expressiveness, more plausible, more capable of conveying the emotional content of the scene, and preferred it overall. However, no significant differences were found for the other dimensions. This outcome could be explained by the importance of the upper face in expressing fear, such as the motion of cheekbones and eyebrows, which is not effectively conveyed by the LE modality.

In the *Happiness* scene (Figure 3.14d), there was no clear winner between the FT and LE modalities. However, FT was perceived as significantly better than LE in terms of realism, general likelihood, pleasantness, and general preference. It is worth noting that the M scores for expressiveness were very similar for both modalities. This result may be attributed to issues with the HTC VIVE Facial Tracker, which had difficulty detecting the user's smile or triggering the appropriate blendshape associated with smiling. This explanation aligns with comments from participants who noted that they did not see the avatar smiling despite the context, leading both modalities to appear less expressive.

It is difficult to draw any conclusive findings for the *Anger* scene (Figure 3.14e), possibly due to the impulsive nature of the emotion. Additionally, some participants mentioned in the open feedback section that the scene heavily relied on speech, which could explain why LE may be on par with FT in this context.

Similarly, no significant differences were found for the *Surprise* scene (Figure 3.14f). This outcome could be attributed to the fact that surprise, like anger or sadness, is a fleeting emotion that is highly expressive for short periods but challenging to maintain for a long time. Thus, it is possible that the presence of other elements such as body movements and the audio clip itself provided sufficient cues, making it difficult to discern differences in facial expressions between FT and LE, despite the theoretically higher fidelity of FT in expressions unrelated to speech.

Limitations and Future Developments

In future developments, the analysis presented in this study may be expanded. For example, the evaluation of additional hardware-based approaches may be considered, such as Meta Quest Pro's facial tracking, or software-based approaches, such as emotion detection algorithms combined with SALSA to improve the approximation of the entire face. The scenes used to represent emotions may also be refined in order to address limitations related to the short duration of some of them (e.g., disgust and surprise) compared to more persistent emotions.

Furthermore, the anger scene may be modified to be less reliant on speech and incorporate more anger-related facial expressions (e.g., shouts). It may also be worthwhile to investigate the incorporation of multiple scenes to represent each emotion, in order to broaden the evaluation of the two modalities from alternative perspectives. Additionally, other factors influencing non-verbal communication, such as body gestures and voice intonation, could be included in the evaluation.

Whole-body tracking solutions may be considered as alternatives to IK for this purpose.

Finally, it may be interesting to explore other avatar representations, such as cartoonish, non-human, or abstract avatars, possibly integrated with other techniques to convey emotions (e.g., UI elements, faceless avatars) to study their suitability for social VR scenarios and the metaverse.

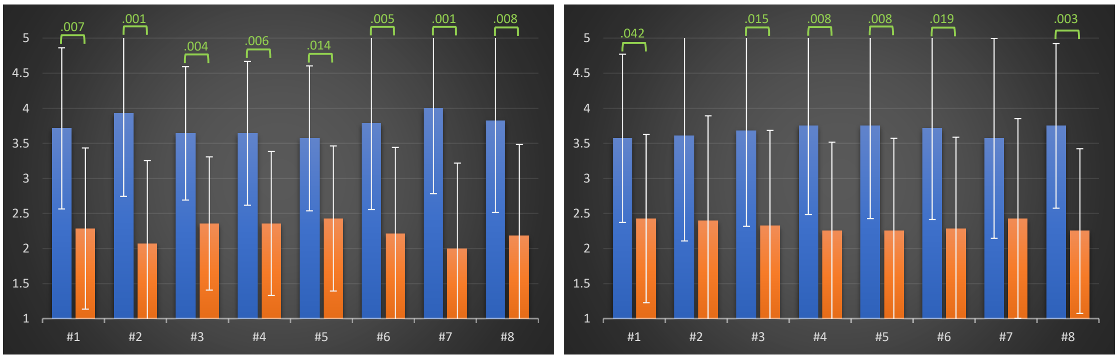
3.4 Considerations and Remarks

The work presented in this chapter seeks to investigate the utilization of voice and body language as a form of HMI in VEs, as well as to support communication between multiple users within shared VEs.

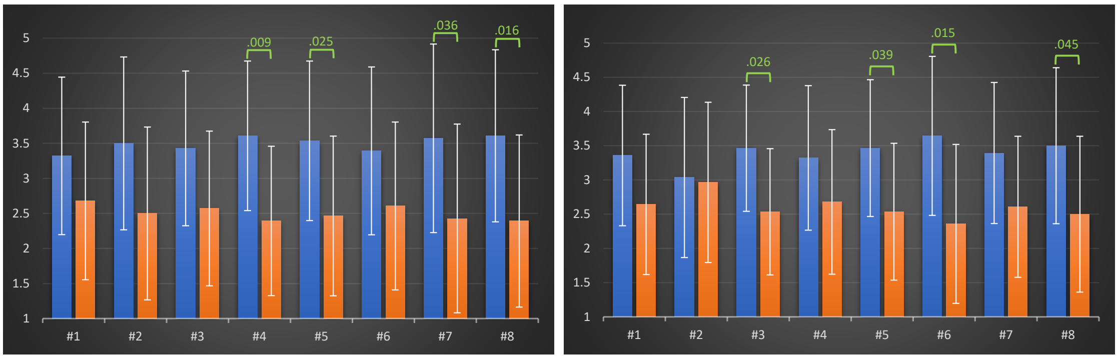
In the first investigation, three voice-based, hands-free navigation techniques for VR are implemented and compared in terms of performance using a within-subjects user study. A speech-only technique, which relies on detecting specific utterances combining intents and entities (e.g., “Take me to the X”) to identify the Point Of Interest (POI) and trigger the teleportation action, a speech with gaze variant, where the direction of the user’s gaze is used to identify the target POI, and voice commands are used to trigger the teleportation action (e.g., “Take me there”), and compound technique, which combines the functionalities of the previous two (utilizing both speech and gaze for POI identification and teleportation triggering). The performance of these three techniques are compared in the user study to evaluate their effectiveness in the context of a large indoor VE that simulates a common industrial training scenario. The results of the study revealed that the speech-only and the compound variants performed significantly better than the speech plus gaze one from various subjective perspectives, including likeability, cognitive demand, annoyance, and preference. On the other hand, the latter showed a reduction in errors, such as selecting the wrong POI and commands not being understood by the speech engine. However, this improvement in task performance was accompanied by lower efficiency, pleasantness, enjoyability, control, calmness, and higher repetitiveness, boredom, frustration, and inflexibility compared to the other two. Interestingly, the compound technique did not provide a significant advantage over the speech-based one, except for a slight reduction in the need for additional references to indicate the desired POI, but at the cost of potentially causing more confusion due to a less uniform interaction scheme.

In the second study, two different techniques for representing users’ avatars in multi-user VR experiences were evaluated in the context of an emergency training simulation, specifically a road tunnel fire simulation. The first technique, commonly used in commercial single-user VR applications, involved displaying only the head & hands of the user in form of VR Kit (VK), such as hand controllers and HMD, without an avatar body. The second technique, referred to as Full-Body (FB), utilized a combination of Inverse Kinematics (IK) algorithms and animation blending to create a whole-body realistic humanoid reconstruction of the user's avatar, based on the position and orientation of the user's head and hands. A within-subject user study involving 15 participants revealed different outcomes in terms of how the users perceived their own avatars and other users' avatars. For the representation of their own avatars, there was no clear winner in terms of embodiment or preference, although FB was perceived to be significantly better than VK in terms of appearance, response to external stimuli, and immersion. In terms of the other user's avatar, FB appeared to improve various social presence aspects, such as mutual awareness, mutual attention, and mutual understanding. Furthermore, FB was also preferred over VK for representing the other user in general, as well as for aesthetics and multiplayer interactions. These findings suggest that employing a whole-body approach to represent the avatar of other users in multi-user training simulations could greatly benefit the shared experience, as long as the combined use of IK and animations produces believable and realistic results. However, the situation is different when it comes to the user's own avatar representation, as the investigation yielded mixed results. Comments provided by some participants at the end of the experience offer possible interpretations. In particular, some participants reported being distracted by the FB avatar when the estimated pose of the body differed significantly from their real body pose. In such cases, they would have preferred not to see the avatar at all, similar to the VK approach. Therefore, some users may still perceive head & hands as better than whole-body in these specific situations.

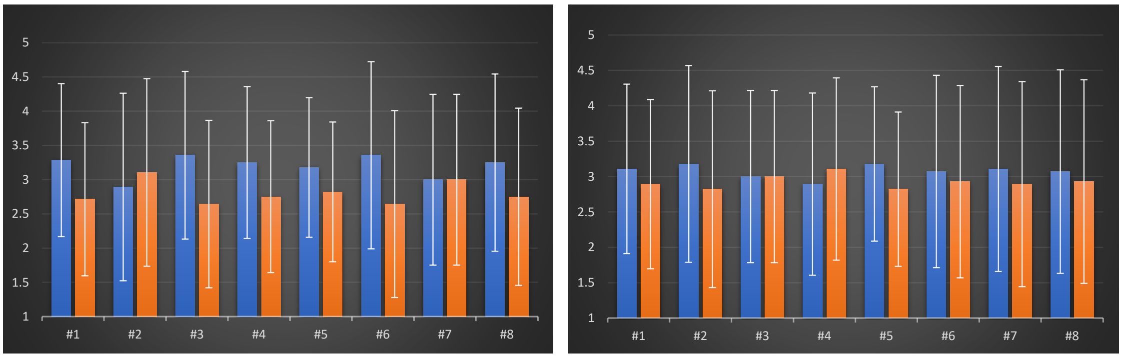
Finally, the third study focuses on comparing two avatar representations in terms of emotion conveyance within a social VR -oriented scenario. Both modalities feature a realistic whole-body avatar representation which emerged as superior from the previous investigation (achieved through a combination of IK and animations). The first modality (Facial+eye Tracking, FT), utilizes additional hardware to track the user's eyes and facial expressions. The second modality (Lip-sync Estimation, LE), utilizes a commercial software solution to generate visually-plausible facial movements related to speech based on the audio captured by the HMD microphone. The study involved 28 participants who blindly evaluated two configurations by observing an avatar performing six pre-determined scenes depicting Ekman's six basic emotions [176]. The evaluation covered important aspects of social presence, including comfort, expressiveness, realism, naturalness of expressions, likelihood, pleasantness, emotion conveyance, and overall preference. The experimental results indicated that FT had a clear superiority over LE in conveying sadness and disgust, as evidenced by most of the analyzed dimensions. This highlights the importance of having a faithful representation of the user's eyes and facial cues in the specific use case. Similar trends were observed for happiness and fear, although to a lesser extent, possibly due to issues with hardware detecting smiling actions and the more impulsive nature of these emotions. No conclusive findings were observed for the remaining emotions. For anger, this outcome may be related to the fact that the scene depicting anger was primarily speech-based, in which LE (via SALSA LipSync) performed well. As for surprise, the fleeting and ephemeral nature of the emotion may have made it less noticeable with both FT and LE.



(a) Sadness (b) Disgust



(c) Fear (d) Happiness



(e) Anger (f) Surprise

Figure 3.14: 🟦 FT 🟧 LE
Subjective outcomes concerning the questionnaires for each scene/emotion (a–f). Statistically significant results ($p < .05$) are marked with brackets. Error bars are used to indicate the SDs. Questions numbered from #1 to #8. Which of the two modalities: #1. Let you feel more comfortable? #2. Was more expressive? #3. Was more realistic? #4. Was more natural? #5. Was more plausible? #6. Was more pleasant? #7. Better conveyed the emotional content? #8. Did you prefer (overall)?

Chapter 4

Conclusions

In this chapter, a summary of the conducted studies and their outcomes is presented. Then, the main limitations of the research are discussed, along with potential directions for future work.

4.1 Summary of Contributions

The previous chapters provided an summary of the research conducted during the Ph.D. period, which aimed to contribute to the advancement in the field of interaction in the context of VR simulations. After three years of intensive research on the topic, valuable insights have been gained into the current state of the field and identified areas where improvements can be made. This journey has not only deepened the understanding of the challenges but has also shed light on the most promising directions for scientific advancements and innovations to address them.

Through this work, it has been learned that that interaction in VR is a multifaceted and complex domain that requires a deep understanding of human perception, cognition, and behavior. The importance of user-centered design principles and iterative evaluation processes cannot be overstated. A user-centric approach ensures that interaction techniques are intuitive, comfortable, and aligned with users' expectations, enhancing the overall user experience.

Moreover, it has been discovered that the embodiment and presence of users within VEs play a crucial role in facilitating effective interaction. Creating realistic and responsive interactions that mimic real-world actions and movements, while minimizing discomfort and motion sickness, is a key challenge. Balancing the need for immersion with the consideration of user comfort is paramount to delivering compelling VR experiences.

This research has highlighted several areas where improvements can be made to enhance interaction in VR. Firstly, addressing the challenges of locomotion and navigation in VR is vital. Developing efficient and intuitive locomotion techniques that mitigate motion sickness while allowing users to move freely and explore virtual worlds remains a key focus area. Additionally, exploring innovative haptic feedback mechanisms is crucial for creating a more immersive and engaging VR experience. Incorporating haptic devices that provide realistic tactile sensations and force feedback can significantly enhance the sense of presence and user engagement within VEs.

In this context, Chapter 2 focuses on these two relevant aspects, firstly by proposing a novel testbed built upon various prior studies for evaluating and comparing locomotion techniques, and then by exploring the benefits of the two main kinds of haptic interfaces (passive and active) in two relevant use cases related to the faithful simulation of manual tools for training purposes.

The testbed aims to fill the longstanding research issue represented by the lack of well-defined method for evaluating and comparing the various locomotion techniques in VR. The tool, which includes a research methodology for acquiring objective and subjective metrics and a scoring system to rank motion methods based on specified criteria, begins to be discussed or adopted and by other literature works on the subject [188]–[190].

Regarding the investigation on the use of PHs in interactive simulations, it aims at investigating the possible contribution in a field, i.e. emergency training, in which trainees often have to learn aspects related to the physicality of their equipment. In particular, when it comes to manual tools, a faithful PH component plays a fundamental role for the learning outcome of the VR training. To investigate the benefits in terms of learning, a VRTS for the training of forest fire firefighters was designed around the use of PH interfaces faithfully replicating the physical characteristics of as many real firefighting tools. The VRTS was then deployed as a complement to a standard, video-based, firefighting course, and compared against the learning outcome of the course alone. The study showed that using the devised VRTS alongside traditional course lessons significantly improved procedural learning. As a result, trainees were better able to recall the concepts related to the use of the simulated tool. The VR experience helped them correct errors before the actual exam, leading to better performance on the practice test compared to the control group. However, there was no significant benefit when it came to conceptual learning. Additionally, participants found the overall learning experience more attractive and stimulating thanks to the practice session in the VRTS, demonstrating the numerous benefits that the VR technology can bring to the simulation of physical objects, as long as their characteristics (e.g., shape and weight) are faithfully replicated.

For what it concerns the investigation on compound configurations of PH and AH technologies, the study is configured as an extension of a previous work in the given field. The reference work investigated the simulation, in VR, of an active electromechanical tool (an ES) by means of two configurations of consumer haptic devices [40]. The prior investigation concluded that the use of a single, fully-fledged exoskeleton-based VR glove did not stand up against the combined use of a simpler device (an ATF-based VR glove) and a custom-made 3D-printed prop representing the physical shape of the simulated too. However, it was not possible to isolate the contribution of each of the elements composing the winning configuration. On the basis of these premises, the new study performs a so-called breakdown analysis to identify a number of viable sub-sets of the original compound haptic configuration: a glove-only variant, a glove plus controller (as low fidelity mockup of the tool) one, and the original configuration, in which a high-fidelity mockup is obtained by enveloping a hand controller (used to provide 6-DOF tracking and vibrations) with a 3D-printed shell reproducing the physical characteristics of the real ES.

The results of the new user study can be summarized as follows. Firstly, the use of a physical prop, instead of just gloves, positively affects mental demand, frustration, attractiveness, perspicuity, efficiency, input, overall fidelity, comfort, interaction, PH fidelity (PTF and PKF), fidelity of various elements (such as hand and finger tracking, feedback of the ES trigger, and various phases of the screwing action), control, efficiency, and overall preference. Secondly, if the physical prop also has a higher-fidelity PH component, such as the 3D-printed shell of the ES, there is a further benefit in terms of task control, attractiveness, perspicuity, dependability, simulation, input, presence, satisfaction, interaction, PH fidelity, fidelity of the ES trigger, and overall preference. Thirdly, the use of only the standard hand controller as a prop negatively affects the perceived novelty of the setup compared to the absence of props. Thus, although the controller-based configuration scored lower than the controller plus mockup one, it can still be considered a viable compromise between setup complexity and performance for the evaluated dimensions, confirming again the importance of the presence of a PH component when simulating real hand equipment.

Chapter 3 moves the investigation to two other important interaction means in VR, that are the voice and body language. A first study aims at investigating the possible use of voice as a HMI method. This is done by evaluating three different speech-based techniques to achieve a completely hands-free teleport-based navigation of a large VE; a purely speech-based technique, which relies on the recognition of navigation intents to possible destinations starting from an uttered description, a second technique which relies on the head/gaze of the user to operate a selection of the desired POI, and uses the voice as input to trigger the teleporting, and a third technique obtained by combining the functionalities of the other two. Results showed the significant superiority of the two approaches supporting the use of descriptions (the speech-only and the compound ones) from a wide range of perspectives. At the same time, the speech-based technique proved to be equally effective as the compound technique with the exception of a minor decrease in the requirement for added references to identify the POI. In fact, utilizing the compound technique resulted in a less cohesive interaction scheme and increased the likelihood of confusion, showing that in this particular scenario, the combination of many too heterogeneous functionalities do not automatically lead to a better performance.

The second study focuses of the HHI aspects in multi-user shared VR simulations. Whereas the presence of voice communication (through VOIP) can be considered as a standard feature in these contexts, finding the best technique of representing VR user within a particular shared VE is not easy, and the conditions may dramatically change from one use case to another. To this purpose, two different techniques for avatar representation, differing in term of level of visibility (head & hands vs whole-body) and realism (minimal vs realistic), are compared against a relevant use case in the field the emergency training (i.e., a road tunnel fire simulation). Results suggest that utilizing a whole-body approach to depict the avatars of other users in multi-user training simulations can significantly enhance the shared experience as long as the use of IK and animations generates credible and realistic outcomes. However, the findings were mixed when it came to the representation of the user's own avatar. On the other hand, the whole-body avatar was perceived as distracting when the estimated posture of the body differed considerably from their actual body posture, and in such situations, they would prefer not to see the avatar at all, similar to the other approach. Hence, in some specific circumstances, some a minimalist approach may still be more effective than the whole-body one.

After verifying that the whole-body approach is generally preferred by users, the investigation was moved to the various possibilities to allow such avatars to effectively convey the emotional content in a multi-user scenario. In a second study, realistic whole-body avatars were paired with two common approaches for managing facial expressions in VR. The first approach consists in taking advantage of facial and eye-tracking hardware to directly maps the user's facial and eye movements to the 3D avatar. The second one consists in utilizing a software solution to approximate of the lip movements of the avatar on the basis of the voice input of the relative users. The focus was put on the conveyance of emotions, in the context of a simulated social-VR scenario. Results highlight the importance of having an accurate portrayal of the user's eye movements and facial expressions in the investigated scenario, but not for all the considered emotions. In fact, apart from sadness, disgust, and, to a lesser extent, happiness and fear, there were no conclusive findings for the other two emotions (anger and surprise). This outcome suggests that in some conditions, the contemporary use of voice and body movements could reduce the visibility of facial expressions, especially in frantic moments, making the faithful mapping between user's real movements and avatar's face less important and thus the less precise lip-sync approximation a good trade-off between additional complexity and faithful emotion conveyance.

4.2 Limitations and Future Developments

A first limitation of most of the reported studies is the lack of standardized approaches available at the time of the investigation. Apart from LET-VR, which specifically aimed at filling the relative research gap, the other studies relied on protocols partially inspired to methodologies of previous literature works, and partially customized to make them suitable for the investigated use case. Although this approach allows to operate comparisons without requiring a well-defined and broadly-accepted methodology, there is a risk of introducing sources of biases related to the necessity to perform arbitrary design choices, possibly undermining the generalizability of the obtained results.

To cope with this important issue, one of the further developments being pursued during the writing of this thesis is the definition of new standard approaches for these kinds of evaluations, similarly to what it was done for locomotion. In particular, an evaluation methodology for investigating the phenomenon of cybersickness in VR is being defined, with the aim to provide it again in form of public open souce tool (i.e., a testbed). At the same time, for what it concerns the speech-based interfaces, the evaluation protocol in being refined to correct a number of issues emerged during the experiments, mainly related to some arbitrary choices, in order to perform a new and more fairer comparison of this kind of paradigms. In other cases, like the study on emotion conveyance of avatars, the lack of similar previous investigations makes this type of refinement much more difficult.

Another restriction of the presented studies was the possible lack of statistical power, due to the fact that some experiments were performed with a limited sample size during the COVID-19 pandemic. Future works, aimed at investigating other still-unexplored areas of interaction in VR, will be designed and structured in order to ensure a sufficiently large sample size, thus possibly guaranteeing a higher generalizability of the measured outcomes.

Despite these limitations, the hope is that the performed work could lead to an advancement of the state of the art in the studied area, by contributing to the development of standardized evaluation methodologies (i.e., in the case of [45]) and to the validation of various HCI and HHI techniques and approaches when applied to the simulation of real life relevant use cases, from the training in emergency and industrial contexts, to social VR and the metaverse-oriented scenarios.

Looking ahead, there are several promising directions for scientific advancements and innovations in the field of VR interaction. One area of great potential lies in the integration of Augmented Reality (AR) and VR, enabling Mixed Reality (MR) experiences. By seamlessly blending the real and virtual worlds, users can interact with digital content while maintaining a connection with their physical surroundings. This fusion opens up new possibilities for collaboration, visualization, and interactive storytelling.

Furthermore, the exploration of Brain-Computer Interfaces (BCIs) holds immense promise for revolutionizing interaction in VR [191], [192]. By leveraging neurophysiological signals, BCIs can enable direct communication between the human brain and VE, bypassing traditional input devices. This avenue of research can lead to more immersive and intuitive interactions, where users can control and manipulate virtual objects through their thoughts and intentions.

Lastly, the field of social interaction in VR is an exciting area of research. Enabling realistic and compelling social interactions within VEs has the potential to revolutionize communication, remote collaboration, and shared experiences. Advancements in avatar representation, emotion recognition, and social presence can contribute to creating highly engaging and socially immersive VR environments.

In conclusion, through the three years of work on interaction in VR, a valuable lesson was learned about the significance of user-centered design, the challenges related to embodiment and presence, and the need for accurate haptic feedback, locomotion techniques, and social interactions. The most promising directions for scientific advancements and innovations lie in the integration of AR and VR, BCIs, and the exploration of social interaction. By pushing the boundaries of these areas, we can enhance the immersion, intuitiveness, and transformative potential of VR, opening up new horizons for a wide range of applications across various domains.

Glossary

AC Accuracy. 31, 32, 34, 38, 39, 45, 48

AH Active Haptic. 17, 21, 22, 76, 134

AKF Active Kinesthetic Feedback. 21, 75, 76

AQ Acrophobia Questionnaire. 35

AR Augmented Reality. 13, 120, 137, 138

AS Arm-Swinging. 24, 25, 41, 44, 48, 108, 110

ATF Active Tactile Feedback. 21, 74–79, 81, 85, 88, 135

ATT Attractiveness. 71

BCI Brain-Computer Interface. 137, 138

DK Development Kit. 78, 87

DOF Degrees-Of-Freedom. 25, 29, 61, 75, 78–80, 87, 88, 135

EP Error-Proneness. 31, 32, 34, 38, 39, 45, 48

ES Electric Screwdriver. 11, 22, 75–84, 86–88, 134, 135

FB Full-Body. 106–111, 115–123, 128, 129

FDS Fire Dynamics Simulation. 52, 53, 107

FR Functional Requirement. 26, 38–40, 44, 45, 48, 49

FT Facial+eye Tracking. 122, 124, 126, 127, 129–131

G Gloves-only. 79, 80, 83–88

G+C Gloves+Controller. 79, 80, 83–88

G+M Gloves+Mockup. 78–80, 83–88

HCI Human-Computer Interaction. 14, 91, 137

HHI Human-Human Interaction. 14, 17, 136, 137

HMD Head-Mounted Display. 13, 24, 26, 53, 54, 57, 61, 62, 72, 73, 84, 105, 109, 110, 120–122, 124, 125, 128, 129

HMI Human-Machine Interaction. 17, 128, 135

IK Inverse Kinematics. 106, 108–110, 119, 122, 128, 129, 136

IMMS Instructional Materials Motivation Survey. 11, 66, 69, 70

JS JoyStick. 24, 25, 41–44, 48

LE Lip-sync Estimation. 123, 124, 126, 127, 129–131

LET-VR Locomotion Evaluation Testbed VR. 10, 11, 23, 27, 30–33, 35–38, 41, 44, 49, 89, 136

M Mean. 37, 38, 40, 48, 63, 68, 70, 72, 82, 84, 126, 127

MCDA Multi-Criteria Decision Analysis. 37, 48

MR Mixed Reality. 13, 120, 137

NLP Natural Language Processing. 91, 94, 95

NMPS Networked Minds Social Presence. 12, 115, 118

NPC Non-Player Character. 57, 64, 109, 111, 112

NR Non-functional Requirement. 10, 26, 31–34, 38, 40, 44, 48, 49

OS Operation Speed. 31, 32, 34, 38–40, 45, 48

OT Other. 31, 34

PE Physical Effort. 32–35, 38, 39, 45, 46, 48

PH Passive Haptic. 17, 21, 22, 50, 51, 53–59, 72–78, 85, 89, 90, 134, 135

PKF Passive Kinesthetic Feedback. 21, 74–77, 79, 90, 135

POI Point Of Interest. 93, 95–99, 101–104, 128, 135

POV Point Of View. 23, 28, 98, 99, 108, 126

PTF Passive Tactile Feedback. 21, 74, 75, 79, 90, 135

RDB Raw Database. 40, 43, 44

S Speech-only. 11, 96, 97, 101–104

SASSI Subjective Assessment of Speech System Interfaces. 12, 99, 101, 102

SD Standard Deviation. 63, 67, 68, 70, 72, 82, 84–87, 101–103, 116, 118, 126, 131

SDK Software Development Kit. 95, 108

SG Speech with Gaze. 12, 96, 97, 101–104

SGD and Speech with Gaze & Descriptions. 12, 96–98, 101–104

SIM-TLX Simulation Task Load Index. 11, 83, 85

SM SlideMill. 24, 41, 42, 44, 48, 107, 119

SSQ Simulator Sickness Questionnaire. 33, 35, 36, 38, 47, 83, 84

SUD Subjective Units of Discomfort. 32, 35, 39

SUS System Usability Scale. 66, 71, 83, 84, 99, 103

SWOT Strengths, Weaknesses, Opportunities, and Threats. 51

TM TreadMill. 24, 107, 119

U-NET Unity High-level Network API. 109, 110

UEQ User Experience Questionnaire. 11, 83–85

UI User Interface. 14, 55, 96, 97, 100, 105, 128

UPD Unintended Positional Drift. 24

UX User eXperience. 10, 14, 17, 22, 25, 26, 28, 29, 34, 76, 83–85, 89, 104, 105, 107, 108, 121

V Video-only. 63, 64, 66–71

V+VR Video+VR. 11, 63, 64, 66–72

V/R Virtual/Real. 35, 46, 48

VE Virtual Environment. 12–15, 17, 19, 20, 22, 24, 25, 29, 42, 43, 49, 50, 53–55, 57, 58, 64, 74, 77–83, 88, 89, 92–98, 101, 104–107, 117, 119–122, 125, 128, 133–138

VK VR Kit. 109, 111, 114–118, 128, 129

VOIP Voice Over Internet Protocol. 109, 136

VR Virtual Reality. 9, 11–17, 19, 20, 22–29, 31, 36, 40–45, 49–58, 63, 66, 67, 69, 72–80, 82, 84, 87–96, 99, 105–109, 111, 113–115, 119–125, 127–129, 133–138, 140–142

VRTS VR Training Scenario. 22, 50–55, 57, 61–66, 69, 71–74, 89, 94, 134

WDB Weighted Database. 40, 41, 44, 48

WIP Walking-In-Place. 24, 25, 41, 42, 44, 48, 119

WSM Weighted Sum Model. 37, 38

XR eXtended Reality. 13, 16, 120

Bibliography

[1] H. Rheingold, *Virtual Reality.* Summit Books, 1991.

[2] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller, "State of the art of virtual reality technology", in *Proc. of 2016 IEEE Aerospace Conference*, 2016, pp. 1–19. DOI: `10.1109/AERO.2016.7500674`.

[3] C. Wee, K. M. Yap, and W. N. Lim, "Haptic interfaces for virtual reality: Challenges and research directions", *IEEE Access*, vol. 9, pp. 112 145–112 162, 2021. DOI: `10.1109/ACCESS.2021.3103598`.

[4] Y. Li, J. Huang, F. Tian, H.-A. Wang, and G.-Z. Dai, "Gesture interaction in virtual reality", *Virtual Reality & Intelligent Hardware*, vol. 1, no. 1, pp. 84–112, 2019. DOI: `10.3724/SP.J.2096-5796.2018.0006`.

[5] C. Kyrlitsias and D. Michael-Grigoriou, "Social interaction with agents and avatars in immersive virtual environments: A survey", *Frontiers in Virtual Reality*, vol. 2, no. 786665, pp. 1–13, 2022. DOI: `10.3389/frvir.2021.786665`.

[6] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques", in *Proc. of IEEE 1997 Annual International Symposium on Virtual Reality (VR 1997)*, 1997, pp. 45–52. DOI: `10.1109/VRAIS.1997.583043`.

[7] A. Ferracani, D. Pezzatini, J. Bianchini, G. Biscini, and A. Del Bimbo, "Locomotion by natural gestures for immersive virtual environments", in *Proc. of 1st International Workshop on Multimedia Alternate Realities (AltMM '16)*, 2016, pp. 21–24. DOI: `10.1145/2983298.2983307`.

[8] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, *3D User Interfaces: Theory and practice.* 2004, pp. 1–85. DOI: `10.5555/993837`.

[9] E. Bouzbib, G. Bailly, S. Haliyo, and P. Frey, ""Can I touch this?": Survey of virtual reality interactions via haptic solutions", in *Proc. of 32nd Conference on l'Interaction Homme-Machine (IHM '21)*, 2021, pp. 1–16. DOI: `10.1145/3450522.3451323`.

[10] K. L. Nowak and J. Fox, "Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations", *Review of Communication Research*, vol. 6, pp. 30–53, 2018. DOI: `10.12840/issn.2255-4165.2018.06.01.015`.

[11] A. Zenner and A. Krüger, "Drag:On: A virtual reality controller providing haptic feedback based on drag and weight shift", in *Proc. of 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 2019, pp. 1–12. DOI: `10.1145/3290605.3300441`.

[12] J. Tromp and D. Snowdon, "Virtual body language: Providing appropriate user interfaces in collaborative virtual environments", in *Proc. of ACM symposium on Virtual reality software and technology (VRST '97)*, 1997, pp. 37–44. DOI: `10.1145/261135.261143`.

[13] J. Muller, C. Krapichler, L. S. Nguyen, K.-H. Englmeier, and M. Lang, "Speech interaction in virtual reality", in *Proc. of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '98)*, vol. 6, 1998, pp. 3757–3760. DOI: `10.1109/ICASSP.1998.679701`.

[14] A. Cannavó, D. Calandra, F. G. Pratticó, V. Gatteschi, and F. Lamberti, "An evaluation testbed for locomotion in virtual reality", *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 1871–1889, 2021. DOI: `10.1109/TVCG.2020.3032440`.

[15] D. Calandra, F. De Lorenzis, A. Cannavó, and F. Lamberti, "Immersive virtual reality and passive haptic interfaces to improve procedural learning in a formal training course for first responders", *Virtual Reality*, 2022. DOI: `10.1007/s10055-022-00704-9`.

[16] F. G. Pratticó, A. Cannavó, D. Calandra, and F. Lamberti, "A breakdown study of a mockup-based consumer haptic setup for virtual reality", *IEEE Consumer Electronics Magazine*, pp. 1–14, 2022. DOI: `10.1109/MCE.2022.3212571`.

[17] R. A. Ruddle and S. Lessels, "For efficient navigational search, humans require full physical movement, but not a rich visual scene", *Psychological Science*, vol. 17, no. 6, pp. 460–465, 2006. DOI: `10.1111/j.1467-9280.2006.01728.x`.

[18] D. Waller and E. Hodgson, "Sensory contributions to spatial knowledge of real and virtual environments", in *Human Walking in Virtual Environments*, 2013, pp. 3–26. DOI: `10.1007/978-1-4419-8432-6_1`.

[19] E. A. Suma, S. Babu, and L. F. Hodges, "Comparison of travel techniques in a complex, multi-level 3D environment", in *2007 IEEE Symposium on 3D User Interfaces (3DUI 2007)*, 2007, pp. 1–7. DOI: `10.1109/3DUI.2007.340788`.

[20] D. C. Niehorster, L. Li, and M. Lappe, "The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research", *I-Perception*, vol. 8, no. 3, pp. 1–23, 2017. DOI: `10.1177/2041669517708205`.

[21] J. N. Templeman, P. S. Denbrook, and L. E. Sibert, "Virtual locomotion: Walking in place through virtual environments", *Presence*, vol. 8, no. 6, pp. 598–617, 1999. DOI: `10.1162/105474699566512`.

[22] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey, "Point & teleport locomotion technique for virtual reality", in *Proc. of 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*, 2016, pp. 205–216. DOI: `10.1145/2967934.2968105`.

[23] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a WIM: Interactive worlds in miniature", in *Proc. of SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*, 1995, pp. 265–272. DOI: `10.1145/223904.223938`.

[24] S. L. Stoev, D. Schmalstieg, and W. Straßer, "Two-handed through-the-lens-techniques for navigation in virtual environments", in *Immersive Projection Tech. and Virtual Env.* 2001, pp. 51–60. DOI: `doi.org/10.1007/978-3-7091-6221-7_6`.

[25] L. P. Fiore, E. Coben, S. Merritt, P. Liu, and V. Interrante, "Towards enabling more effective locomotion in VR using a wheelchair-based motion platform", in *Joint Virtual Reality Conference of EGVE - EuroVR (JVRC 13)*, 2013, pp. 83–90. DOI: `10.2312/EGVE.JVRC13.083-090`.

[26] J. Wang and R. W. Lindeman, "Comparing isometric and elastic surfboard interfaces for leaning–based travel in 3D virtual environments", in *2012 IEEE Symposium on 3D User Interfaces (3DUI 2012)*, 2012, pp. 31–38. DOI: `10.1109/3DUI.2012.6184181`.

[27] S. Beckhaus, K. J. Blom, and M. Haringer, "ChairIO–the chair-based interface", *Concepts and Technologies for Pervasive Games: A Reader for Pervasive Gaming Research*, vol. 1, pp. 231–264, 2007.

[28] S. Fels, Y. Kinoshita, T.-P. G. Chen, *et al.*, "Swimming across the Pacific: A VR swimming interface", *IEEE Computer Graphics and Applications*, vol. 25, no. 1, pp. 24–31, 2005. DOI: `10.1109/MCG.2005.20`.

[29] N. C. Nilsson, S. Serafin, F. Steinicke, and R. Nordahl, "Natural walking in virtual reality: A review", *Computers in Entertainment*, vol. 16, no. 2, 8:1–8:22, 2018. DOI: `10.1145/3180658`.

[30] K. S. Hale and K. M. Stanney, "Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations", *IEEE Computer Graphics and Applications*, vol. 24, no. 2, pp. 33–39, 2004. DOI: `10.1109/MCG.2004.1274059`.

[31] J.-L. Rodríguez, R. Velázquez, C. Del-Valle-Soto, S. Gutiérrez, J. Varona, and J. Enríquez-Zarate, "Active and passive haptic perception of shape: Passive haptics can support navigation", *Electronics*, vol. 8, no. 3, pp. 1–12, 2019. DOI: `10.3390/electronics8030355`.

[32] A. Zenner and A. Krüger, "Shifty: A weight-shifting dynamic passive haptic proxy to enhance object perception in virtual reality", *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1285–1294, 2017. DOI: `10.1109/TVCG.2017.2656978`.

[33] M. White, J. Gain, U. Vimont, and D. Lochner, "The case for haptic props: Shape, weight and vibro-tactile feedback", in *Proc. of 12th ACM SIGGRAPH Conference on Motion, Interaction and Games (MIG '19)*, 2019, pp. 1–10. DOI: `10.1145/3359566.3360058`.

[34] A. Lécuyer, S. Coquillart, A. Kheddar, P. Richard, and P. Coiffet, "Pseudo-haptic feedback: Can isometric input devices simulate force feedback?", in *In Proc. of IEEE Virtual Reality 2000*, 2000, pp. 83–90. DOI: `10.1109/VR.2000.840369`.

[35] A. Lécuyer, J.-M. Burkhardt, and L. Etienne, "Feeling bumps and holes without a haptic interface: The perception of pseudo-haptic textures", in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*, 2004, pp. 239–246. DOI: `10.1145/985692.985723`.

[36] A. Lécuyer, "Simulating Haptic Feedback Using Vision: A Survey of Research and Applications of Pseudo-Haptic Feedback", *Presence: Teleoperators and Virtual Environments*, vol. 18, no. 1, pp. 39–53, 2009. DOI: `10.1162/pres.18.1.39`.

[37] Y. Ujitoko and Y. Ban, "Survey of pseudo-haptics: Haptic feedback design and application proposals", *IEEE Transactions on Haptics*, vol. 14, no. 4, pp. 699–711, 2021. DOI: `10.1109/TOH.2021.3077619`.

[38] R. D. Joyce and S. Robinson, "Passive haptics to enhance virtual reality simulations", in *Proc. of AIAA Modeling and Simulation Technologies Conference (AIAA 2017)*, 2017, ch. 1, pp. 1–10. DOI: `10.2514/6.2017-1313`.

[39] D. Calandra, F. G. Pratticò, A. Cannavò, L. Micelli, and F. Lamberti, "Building reconfigurable passive haptic interfaces on demand using off-the-shelf construction bricks", in *Proc. of 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, 2019, pp. 1403–1404. DOI: `10.1109/VR.2019.8797865`.

[40] F. G. Prатticó, D. Calandra, M. Piviotti, and F. Lamberti, "Assessing the user experience of consumer haptic devices for simulation-based virtual reality", in *Proc. of 11th International Conference on Consumer Electronics (ICCE-Berlin 2021)*, 2021, pp. 1–6. DOI: 10.1109/ICCE-Berlin53567.2021.9719998.

[41] A. Garg, J. A. Fisher, W. Wang, and K. P. Singh, "ARES: An application of impossible spaces for natural locomotion in VR", in *Proc. of 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*, 2017, pp. 1–4. DOI: 10.1145/3027063.3048416.

[42] N. C. Nilsson, S. Serafin, and R. Nordahl, "The perceived naturalness of virtual locomotion methods devoid of explicit leg movements", in *Proc. of Motion on Games (MIG '13)*, 2013, pp. 155–164. DOI: 10.1145/2522628.2522655.

[43] M. C. Whitton, J. V. Cohn, J. Feasel, *et al.*, "Comparing VE locomotion interfaces", in *Proc. of IEEE 2005 Annual International Symposium Virtual Reality (VR 2005)*, 2005, pp. 123–130. DOI: 10.1109/VR.2005.1492762.

[44] M. J. Schuemie, B. Abel, C. van der Mast, M. Krijn, and P. M. G. Emmelkamp, "The effect of locomotion technique on presence, fear and usability in a virtual environment", in *Proc. of 2005 Euromedia (EUROMEDIA'2005)*, 2005, pp. 129–135, ISBN: ISBN 90-77381-17-1.

[45] VR@POLITO, *Locomotion Evaluation Testbed VR (LET-VR)*, [accessed 19 April 2023]. [Online]. Available: https://github.com/VRatPolito/LET-VR.

[46] D. Calandra, M. Billi, F. Lamberti, A. Sanna, and R. Borchiellini, "Arm swinging vs treadmill: A comparison between two techniques for locomotion in virtual reality", in *EUROGRAPHICS 2018 - Short Papers (EG 2018)*, 2018, pp. 53–56. DOI: 10.2312/egs.20181043.

[47] D. Calandra, F. Lamberti, and M. Migliorini, "On the usability of consumer locomotion techniques in serious games: Comparing arm swinging, treadmills and walk-in-place", in *Proc. of 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin 2019)*, 2019, pp. 348–352. DOI: 10.1109/ICCE-Berlin47944.2019.8966165.

[48] E. Schiza, M. Matsangidou, K. Neokleous, and C. S. Pattichis, "Virtual reality applications for neurological disease: A review", *Frontiers in Robotics and AI*, vol. 6, pp. 1–14, 2019. DOI: 10.3389/frobt.2019.00100.

[49] N. C. Nilsson, S. Serafin, M. H. Laursen, K. S. Pedersen, E. Sikstrom, and R. Nordahl, "Tapping-in-place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input", in *2013 IEEE Symposium on 3D User Interfaces (3DUI 2013)*, 2013, pp. 31–38. DOI: 10.1109/3DUI.2013.6550193.

[50] Y. S. Pai and K. Kunze, “Armswing: Using arm swings for accessible and immersive navigation in AR/VR spaces”, in *Proc. of 16th International Conference on Mobile and Ubiquitous Multimedia (MUM ’17)*, 2017, pp. 189–198. DOI: `10.1145/3152832.3152864`.

[51] J.-F. Lapointe and P. Savard, “A comparative study of three bimanual travel techniques for desktop virtual walkthroughs”, in *2009 IEEE International Workshop on Haptic Audio visual Environments and Games (HAVE 2009)*, 2009, pp. 182–185. DOI: `10.1109/HAVE.2009.5356144`.

[52] G. Loup and E. Loup-Escande, “Effects of travel modes on performances and user comfort: A comparison between ArmSwinger and Teleporting”, *International Journal of Human–Computer Interaction*, vol. 35, no. 14, pp. 1270–1278, 2019. DOI: `10.1080/10447318.2018.1519164`.

[53] E. Suma, S. Finkelstein, M. Reid, S. Babu, A. Ulinski, and L. F. Hodges, “Evaluation of the cognitive effects of travel technique in complex real and virtual environments”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 4, pp. 690–702, 2010. DOI: `10.1109/TVCG.2009.93`.

[54] M. Nabiyouni, A. Saktheeswaran, D. A. Bowman, and A. Karanth, “Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality”, in *2015 IEEE Symposium on 3D User Interfaces (3DUI 2015)*, 2015, pp. 3–10. DOI: `10.1109/3DUI.2015.7131717`.

[55] M. Asnabrygg, *Unbreakable VR runner*, [accessed 19 April 2023], 2016. [Online]. Available: `https://store.steampowered.com/app/494310/Unbreakable_Vr_Runner`.

[56] R. Paris, M. Joshi, Q. He, G. Narasimham, T. P. McNamara, and B. Bodenheimer, “Acquisition of survey knowledge using walking in place and resetting methods in immersive virtual environments”, in *Proc. of ACM Symposium on Applied Perception (SAP ’17)*, 2017, pp. 1–8. DOI: `10.1145/3119881.3119889`.

[57] Bethesda Game Studios, *Fallout 4 VR*, [accessed 19 April 2023], 2017. [Online]. Available: `https://store.steampowered.com/app/611660/Fallout_4_VR`.

[58] Ready At Dawn, *Echo VR*, [accessed 19 April 2023], 2017. [Online]. Available: `https://www.oculus.com/experiences/rift/1369078409873402/`.

[59] Metricminds GmbH and Co KG, *Catch & Release*, [accessed 19 April 2023], 2018. [Online]. Available: `https://store.steampowered.com/app/679750/Catch__Release/`.

[60] Beat Games, *Beat Saber*, [accessed 19 April 2023], 2018. [Online]. Available: `https://beatsaber.com/`.

[61] SUPERHOT Team, *SUPERHOT VR*, [accessed 19 April 2023], 2017. [Online]. Available: `https://store.steampowered.com/app/617830/SUPERHOT_VR/`.

[62] Owlchemy Labs, *Job simulator*, [accessed 19 April 2023], 2016. [Online]. Available: `https://store.steampowered.com/app/448280/Job_Simulator`.

[63] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness.", *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993. DOI: `10.1207/s15327108ijap0303\_3`.

[64] S. L. Fischer, P. B. Watts, R. L. Jensen, and J. Nelson, "Energy expenditure, heart rate response, and metabolic equivalents (METs) of adults taking part in children's games", *Journal of sports medicine and physical fitness*, vol. 44, p. 398, 2004.

[65] K. R. Westerterp, "Assessment of physical activity: A critical appraisal", *European Journal of Applied Physiology*, vol. 105, no. 6, pp. 823–828, 2009. DOI: `10.1007/s00421-009-1000-2`.

[66] A. P. Hills, N. Mokhtar, and N. M. Byrne, "Assessment of physical activity and energy expenditure: An overview of objective measures", *Frontiers in Nutrition*, vol. 1, p. 5, 2014. DOI: `10.3389/fnut.2014.00005`.

[67] R. S. Kalawsky, "VRUSE – A computerised diagnostic tool: For usability evaluation of virtual/synthetic environment systems", *Applied Ergonomics*, vol. 30, no. 1, pp. 11–25, 1999. DOI: `10.1016/S0003-6870(98)00047-7`.

[68] "ISO 9241-400:2007: Ergonomics of human—-system interaction – Part 400: Principles and requirements for physical input devices", International Organization for Standardization, Standard, 2007.

[69] ElectricNightOwl, *Arm Swinger*, [accessed 19 April 2023]. [Online]. Available: `https://github.com/ElectricNightOwl/ArmSwinger`.

[70] J. Feasel, M. C. Whitton, and J. D. Wendt, "LLCM-WIP: Low-latency, continuous-motion walking-in-place", in *2008 IEEE Symposium on 3D User Interfaces (3DUI 2008)*, 2008, pp. 97–104. DOI: `10.1109/3DUI.2008.4476598`.

[71] T. Cakmak and H. Hager, "Cyberith virtualizer: A locomotion device for virtual reality", in *Proc. of ACM SIGGRAPH 2014 Emerging Technologies (SIGGRAPH '14)*, 2014, p. 1. DOI: `10.1145/2614066.2614105`.

[72] C. Boletsis and J. E. Cedergren, "VR locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques", *Advances in Human-Computer Interaction*, vol. 2019, 2019. DOI: `10.1155/2019/7420781`.

[73] J. B. Cronin and K. T. Hansen, “Strength and power predictors of sports speed”, *ournal of strength and conditioning research*, vol. 19, pp. 349–357, 2005. DOI: `10.1519/14323.1`.

[74] B. K. Jaeger and R. R. Mourant, “Comparison of simulator sickness using static and dynamic walking simulators”, in *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, 2001, pp. 1896–1900. DOI: `10.1177/154193120104502709`.

[75] Valve, *Half-Life: Alyx*, [accessed 19 April 2023], 2020. [Online]. Available: `https://store.steampowered.com/app/546560/HalfLife_Alyx/`.

[76] B. Lang, *“The Boneworks thing is weird” - Valve says it took more inspiration from Budget Cuts for Half-Life: Alyx*, [accessed 19 April 2023], 2020. [Online]. Available: `https://www.roadtovr.com/half-life-alyx-best-rated-pc-game-2020-steam-vr-game-all-time/`.

[77] K. McKeand, *“Half-Life: Alyx” is the Best Rated PC Game of 2020 & Best Rated Steam VR Game of All Time*, [accessed 19 April 2023], 2020. [Online]. Available: `https://www.vg247.com/budget-cuts-half-life-alyx-boneworks-valve`.

[78] Z. Feng, V. A. González, R. Amor, R. Lovreglio, and G. Cabrera-Guerrero, “Immersive virtual reality serious games for evacuation training and research: A systematic literature review”, *Computers & Education*, vol. 127, pp. 252–266, 2018. DOI: `10.1016/j.compedu.2018.09.002`.

[79] M. Andrade, C. Souto Maior, E. Silva, M. Moura, and I. Lins, “Serious games & human reliability. The use of game-engine-based simulator data for studies of evacuation under toxic cloud scenario”, in *Proc. of Probabilistic Safety Assessment and Management (PSAM 14)*, Sep. 2018, pp. 1–12.

[80] S. Pedram, S. Palmisano, R. Skarbez, P. Perez, and M. Farrelly, “Investigating the process of mine rescuers’ safety training with immersive virtual reality: A structural equation modelling approach”, *Computers & Education*, vol. 153, p. 103 891, 2020. DOI: `10.1016/j.compedu.2020.103891`.

[81] F. Lamberti, F. De Lorenzis, F. G. Pratticò, and M. Migliorini, “An immersive virtual reality platform for training CBRN operators”, in *Proc. of 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC 2021)*, 2021, pp. 133–137. DOI: `10.1109/COMPSAC51774.2021.00030`.

[82] F. Buttussi and L. Chittaro, “A comparison of procedural safety training in three conditions: Virtual reality headset, smartphone, and printed materials”, *IEEE Transactions on Learning Technologies*, vol. 14, no. 1, pp. 1–15, 2021. DOI: `10.1109/TLT.2020.3033766`.

[83] R. Lovreglio, “Virtual and augmented reality for human behaviour in disasters: A review”, in *Proc. of Fire and Evacuation Modeling Technical Conference (FEMTC 2020)*, Aug. 2020, pp. 1–14.

[84] B. Lok, S. Naik, M. Whitton, and F. P. Brooks, “Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments”, *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 6, pp. 615–628, 2003. DOI: `10.1162/105474603322955914`.

[85] M. Suhail, S. Gainer, J. Haskins, *et al.*, “Simulating a futuristic fire pump panel in virtual reality”, in *Proc. of 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, 2019, pp. 1415–1416. DOI: `10.1109/VR.2019.8798280`.

[86] H. Engelbrecht, R. W. Lindeman, and S. Hoermann, “A SWOT analysis of the field of virtual reality for firefighter training”, *Frontiers in Robotics and AI*, vol. 6, p. 101, 2019. DOI: `10.3389/frobt.2019.00101`.

[87] L. Chittaro, F. Buttussi, and N. Zangrando, “Desktop virtual reality for emergency preparedness: User evaluation of an aircraft ditching experience under different fear arousal conditions”, in *Proc. of 20th ACM Symposium on Virtual Reality Software and Technology (VRST ’14)*, 2014, pp. 141–150. DOI: `10.1145/2671015.2671025`.

[88] S. M. V. Gwynne, E. D. Kuligowski, K. E. Boyce, *et al.*, “Enhancing egress drills: Preparation and assessment of evacuee performance”, *Fire and Materials*, vol. 43, no. 6, pp. 613–631, 2019. DOI: `10.1002/fam.2448`.

[89] M. N. Louka and C. Balducelli, “Virtual reality tools for emergency operation support and training”, in *Proc. of International Conference on Emergency Management Towards Co-operation and Global Harmonization (TIEMS 2001)*, Jun. 2001, pp. 1–10.

[90] T. U. St Julien and C. D. Shaw, “Firefighter command training virtual environment”, in *Proc. of Conference on Diversity in Computing (TAPIA ’03)*, 2003, pp. 1–4. DOI: `10.1145/948542.948549`.

[91] X. Lu, Z. Yang, Z. Xu, and C. Xiong, “Scenario simulation of indoor post-earthquake fire rescue based on building information model and virtual reality”, *Advances in Engineering Software*, vol. 143, p. 102 792, 2020. DOI: `10.1016/j.advengsoft.2020.102792`.

[92] J. Haskins, B. Zhu, S. Gainer, *et al.*, “Exploring VR training for first responders”, in *Proc. of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2020)*, 2020, pp. 57–62. DOI: `10.1109/VRW50115.2020.00018`.

[93] F. Corelli, E. Battegazzorre, F. Strada, A. Bottino, and G. P. Cimellaro, "Assessing the usability of different virtual reality systems for firefighter training", in *Proc. of 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (HUCAPP 2020)*, 2020, pp. 146–153. DOI: `10.5220/0008962401460153`.

[94] R. Querrec, C. Buche, E. Maffre, and P. Chevaillier, "SécuRéVi: Virtual environments for fire fighting training", in *Proc. of the 5th Virtual Reality International Conference (VRIC 2003)*, 2003, pp. 169–175.

[95] M. Cha, S. Han, J. Lee, and B. Choi, "A virtual reality based fire training simulator integrated with fire dynamics data", *Fire Safety Journal*, vol. 50, pp. 12–24, 2012. DOI: `10.1016/j.firesaf.2012.01.004`.

[96] D. Calandra, F. G. Pratticó, M. Migliorini, V. Verda, and F. Lamberti, "A multi-role, multi-user, multi-technology virtual reality-based road tunnel fire simulator for training purposes", in *Proc. of 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (GRAPP 2019)*, 2021, pp. 96–105. DOI: `10.5220/0010319400960105`.

[97] Ü. Çakiroğlu and S. Gökoğlu, "Development of fire safety behavioral skills via virtual reality", *Computers & Education*, vol. 133, pp. 56–68, 2019, ISSN: 0360-1315. DOI: `10.1016/j.compedu.2019.01.014`.

[98] S. Morélot, A. Garrigou, J. Dedieu, and B. N'Kaoua, "Virtual reality for fire safety training: Influence of immersion and sense of presence on conceptual and procedural acquisition", *Computers & Education*, vol. 166, p. 104 145, 2021, ISSN: 0360–1315. DOI: `10.1016/j.compedu.2021.104145`.

[99] R. C. Rothermel, *A mathematical model for predicting fire spread in wildland fuels*, Res. Pap. INT-115. Ogden, UT: U.S. Department of Agriculture, Intermountain Forest and Range Experiment Station, 1972.

[100] F. De Lorenzis, Itratticò, and F. Lamberti, "Work-in-Progress—Blower VR: A virtual reality experience to support the training of forest firefighter", in *Proc. of 2022 8th International Conference of the Immersive Learning Research Network (iLRN 2022)*, 2022, pp. 1–3. DOI: `10.23919/iLRN55037.2022.9815975`.

[101] J. Keller, *Motivational Design for Learning and Performance: The ARCS Model Approach*. 2010, pp. 1–353. DOI: `10.1007/978-1-4419-1250-3`.

[102] F. Strada, A. Bottino, F. Lamberti, G. Mormando, and P. L. Ingrassia, "Holo-BLSD - A holographic tool for self-training and self-evaluation of emergency response skills", *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2019. DOI: `10.1109/TETC.2019.2925777`.

[103] M. Hassenzahl, F. Koller, and M. Burmester, “Der user experience (UX) auf der spur: Zum einsatz von www.attrakdiff.de”, in *Tagungsband UP08*, H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner, and K. Röse, Eds., 2008, pp. 78–82.

[104] P. Jost, S. Cobb, and I. Hämmerle, “Reality-based interaction affecting mental workload in virtual reality mental arithmetic training”, *Behaviour & Information Technology*, vol. 39, no. 10, pp. 1062–1078, 2020. DOI: 10.1080/0144929X.2019.1641228.

[105] J. Brooke, “SUS: A ‘quick and dirty’ usability scale”, *Usability Evaluation in Industry*, p. 189, 1996. DOI: 10.1201/9781498710411-35.

[106] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale”, *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.

[107] B. Boos and H. Brau, “Erweiterung des UEQ um die dimensionen akustik und haptik”, in *Mensch und Computer 2017 - Usability Professionals (UP 17)*, S. Hess and H. Fischer, Eds., 2017. DOI: 10.18420/muc2017-up-0236.

[108] M. Nguyen, M. Melaisi, B. Cowan, A. J. Uribe Quevedo, and B. Kapralos, “Low-end haptic devices for knee bone drilling in a serious game”, *World Journal of Science, Technology and Sustainable Development*, vol. 14, no. 2/3, pp. 241–253, 2017. DOI: 10.1108/WJSTSD-07-2016-0047.

[109] M. Jeong, S. Lim, T. Lim, and J. Ryu, “Work-in-Progress–Is virtual reality simulation ineffective for skill acquisition training?”, in *Proc. of 7th International Conference of the Immersive Learning Research Network (iLRN 2021)*, 2021, pp. 1–3. DOI: 10.23919/iLRN52045.2021.9459402.

[110] P. Fratczak, Y. M. Goh, P. Kinnell, L. Justham, and A. Soltoggio, “Virtual reality study of human adaptability in industrial human-robot collaboration”, in *Proc. of IEEE International Conference on Human-Machine Systems (ICHMS 2020)*, 2020, pp. 1–6. DOI: 10.1109/ICHMS49158.2020.9209558.

[111] J. Perret and E. B. Vander Poorten, “Touching virtual reality: A review of haptic gloves”, in *Proc. of 16th International Conference on New Actuators (ACTUATOR 2018)*, 2018, pp. 1–5.

[112] D. Wang, M. Song, A. Naqash, Y. Zheng, W. Xu, and Y. Zhang, “Toward whole-hand kinesthetic feedback: A survey of force feedback gloves”, *IEEE Transactions on Haptics*, vol. 12, no. 2, pp. 189–204, 2019. DOI: 10.1109/TOH.2018.2879812.

[113] S. Lontschar, D. Deegan, I. Humer, K. Pietroszek, and C. Eckhardt, "Analysis of haptic feedback and its influences in virtual reality learning environments", in *Proc. of 6th International Conference of the Immersive Learning Research Network (iLRN 2020)*, 2020, pp. 171–177. DOI: `10.23919/iLRN47897.2020.9155087`.

[114] P. L. Strandholt, O. A. Dogaru, N. C. Nilsson, R. Nordahl, and S. Serafin, "Knock on wood: Combining redirected touching and physical props for tool-based interaction in virtual reality", in *Proc. of 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, 2020, pp. 1–13. DOI: `10.1145/3313831.3376303`.

[115] D.-S. Choi, S. Ryu, Y. Do, K.-U. Kyung, K. Jin, and S.-Y. Kim, "Affordable drilling interface for haptic interaction in virtual environment", in *Proc. of IEEE International Conference on Consumer Electronics (ICCE 2019)*, 2019, pp. 1–2. DOI: `10.1109/ICCE.2019.8662116`.

[116] J. Jose, R. Unnikrishnan, D. Marshall, and R. R. Bhavani, "Haptics enhanced multi-tool virtual interfaces for training carpentry skills", in *Proc. of 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA 2016)*, 2016, pp. 1–6. DOI: `10.1109/RAHA.2016.7931900`.

[117] D. Escobar-Castillejos, J. Noguez, L. Neri, A. Magana, and B. Benes, "A review of simulators with haptic devices for medical training", *Journal of Medical Systems*, vol. 40, no. 4, p. 104, 2016. DOI: `10.1007/s10916-016-0459-8`.

[118] C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo, "Wearable haptic systems for the fingertip and the hand: Taxonomy, review, and perspectives", *IEEE Transactions on Haptics*, vol. 10, no. 4, pp. 580–600, 2017. DOI: `10.1109/TOH.2017.2689006`.

[119] D. Wang, K. Ohnishi, and W. Xu, "Multimodal haptic display for virtual reality: A survey", *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 610–623, 2020. DOI: `10.1109/TIE.2019.2920602`.

[120] M. Caeiro-Rodríguez, I. Otero-González, F. A. Mikic-Fonte, and M. Llamas-Nistal, "A systematic review of commercial smart gloves: Current status and applications", *Sensors*, vol. 21, no. 8, p. 2667, 2021. DOI: `10.3390/s21082667`.

[121] A. Okamura, M. Cutkosky, and J. Dennerlein, "Reality-based models for vibration feedback in virtual environments", *IEEE/ASME Transactions on Mechatronics*, vol. 6, no. 3, pp. 245–252, 2001. DOI: `10.1109/3516.951362`.

[122] C. H. Leong, R. Mohd-Mokhtar, N. S. Ahmad, and C. W. Leow, "Modeling and control of torque and impact rate during screwing process", in *Proc. of 2017 IEEE Region 10 Conference (TENCON 2017)*, 2017, pp. 1997–2002. DOI: `10.1109/TENCON.2017.8228188`.

[123] D. Harris, M. Wilson, and S. Vine, "Development and validation of a simulation workload measure: The simulation task load index (SIM-TLX)", *Virtual Reality*, vol. 24, no. 4, pp. 557–566, 2020. DOI: `10.1007/s10055-019-00422-9`.

[124] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire", in *HCI and Usability for Education and Work*, A. Holzinger, Ed., 2008, pp. 63–76. DOI: `10.1007/978-3-540-89350-9_6`.

[125] C. Mizera, T. Delrieu, V. Weistroffer, C. Andriot, A. Decatoire, and J.-P. Gazeau, "Evaluation of hand-tracking systems in teleoperation and virtual dexterous manipulation", *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1642–1655, 2019. DOI: `10.1109/JSEN.2019.2947612`.

[126] P. T. Wilson, W. Kalescky, A. MacLaughlin, and B. Williams, "VR locomotion: Walking> walking in place> arm swinging", in *Proc. of 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - Volume 1 (VRCAI '16)*, 2016, pp. 243–249. DOI: `10.1145/3013971.3014010`.

[127] J. Albert and K. Sung, "User-centric classification of virtual reality locomotion", in *Proc. of 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*, 2018, pp. 1–2. DOI: `10.1145/3281505.3283376`.

[128] D. Calandra, F. G. Pratticó, and F. Lamberti, "Comparison of hands-free speech-based navigation techniques for virtual reality training", in *Proc. of 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON 2022)*, 2022, pp. 85–90. DOI: `10.1109/MELECON53508.2022.9842994`.

[129] D. Calandra, F. G. Pratticó, G. Lupini, and F. Lamberti, "Impact of avatar representation in a virtual reality-based multi-user tunnel fire simulator for training purposes", in *Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 1691, 2023, pp. 3–20. DOI: `10.1007/978-3-031-25477-2_1`.

[130] A. Visconti, D. Calandra, and F. Lamberti, "Comparing technologies for conveying emotions through realistic avatars in virtual reality-based metaverse experiences", *Computer Animation and Virtual Worlds*, vol. 34, no. 3-4, e2188, 2023. [Online]. Available: `https://doi.org/10.1002/cav.2188`.

[131] D. Perez-Marin and I. Pascual-Nieto, *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, I. IGI Global, 2011. DOI: `10.4018/978-1-60960-617-6`.

[132] J. Hombeck, H. Voigt, T. Heggemann, R. Datta, and K. Lawonn, "Tell me where to go: Voice-controlled hands-free locomotion for virtual reality systems", in *Proc. of 2023 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2023)*, vol. (in press), 2023, pp. 1–13.

[133] D. Maloney, G. Freeman, and D. Y. Wohn, “"Talking without a Voice": Understanding non-verbal communication in social virtual reality”, *Proc. of ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–25, 2020. DOI: 10.1145/3415246.

[134] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, and M. Bessa, “Hands-free interaction in immersive virtual reality: A systematic review”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2702–2713, 2021. DOI: 10.1109/TVCG.2021.3067687.

[135] D. Hepperle, Y. Weiß, A. Siess, and M. Wölfel, “2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality”, *Computers & Graphics*, vol. 82, pp. 321–331, 2019. DOI: 10.1016/j.cag.2019.06.003.

[136] M. Friedl, *Online Game Interactivity Theory with Cdrom*. Charles River Media, Inc., 2002, ISBN: 1584502150. DOI: 10.5555/579311.

[137] A. Ferracani, M. Faustino, G. X. Giannini, L. Landucci, and A. Del Bimbo, “Natural experiences in museums through virtual reality and voice commands”, in *Proc. of 25th ACM international conference on Multimedia (MM '17)*, 2017, pp. 1233–1234. DOI: 10.1145/3123266.3127916.

[138] J. Sin and C. Munteanu, “Let’s Go There: Voice and pointing together in VR”, in *Proc. of 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20)*, 2020, pp. 1–3. DOI: 10.1145/3406324.3410537.

[139] R. Mehra, V. S. Sharma, V. Kaulgud, S. Podder, and A. P. Burden, “Immersive IDE: Towards leveraging virtual reality for creating an immersive software development environment”, pp. 1–4, 2020. DOI: 10.1145/3387940.3392234.

[140] R. A. Bolt, ““Put-That-There”: Voice and gesture at the graphics interface”, in *Proc. of 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80)*, 1980, pp. 1–9. DOI: 10.1145/800250.807503.

[141] R. Sharma, M. Zeller, V. Pavlovic, *et al.*, “Speech/gesture interface to a visual-computing environment”, *IEEE Computer Graphics and Applications*, vol. 20, no. 2, pp. 29–37, 2000. DOI: 10.1109/38.824531.

[142] M. Thurber, *Rolls-Royce Opens BR725 Virtual Training Hangar*, https://www.ainonline.com/aviation-news/business-aviation/2020-10-09/rolls-royce-opens-br725-virtual-training-hangar, [accessed 19 April 2023], 2020.

[143] K. S. Hone and R. Graham, “Towards a tool for the subjective assessment of speech system interfaces (SASSI)”, *Natural Language Engineering*, vol. 6, no. 3-4, pp. 287–303, 2000. DOI: 10.1017/S1351324900002497.

[144] S. Fathima S J and J. Aroma, "Simulation of fire safety training environment using immersive virtual reality", *International Journal of Recent Technology and Engineering*, vol. 7, pp. 347–350, Jan. 2019.

[145] J.-L. Lugrin, M. Ertl, P. Krop, *et al.*, "Any "body" there? avatar visibility effects in a virtual reality game", in *Proc. of 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2018)*, 2018, pp. 17–24. DOI: `10.1109/VR.2018.8446229`.

[146] M. Parger, J. H. Mueller, D. Schmalstieg, and M. Steinberger, "Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality", in *Proc. of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*, Association for Computing Machinery, 2018, pp. 1–10. DOI: `10.1145/3281505.3281529`.

[147] Valve, *SteamVR*, [accessed 19 April 2023]. [Online]. Available: `https://store.steampowered.com/app/250820/SteamVR`.

[148] Oculus, *First Steps*, [accessed 19 April 2023]. [Online]. Available: `https://www.oculus.com/experiences/quest/1863547050392688`.

[149] E. Kokkinara and M. Slater, "Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion", *Perception*, vol. 43, no. 1, pp. 43–58, 2014. DOI: `10.1068/p7545`.

[150] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, "An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment", *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1406–1414, 2016. DOI: `10.1109/TVCG.2016.2518135`.

[151] Gunfire Games, *Dead and Buried*, [accessed 19 April 2023]. [Online]. Available: `https://www.oculus.com/experiences/rift/1198491230176054`.

[152] VRChat Inc., *VRChat*, [accessed 19 April 2023]. [Online]. Available: `https://hello.vrchat.com`.

[153] M. Kinateder, E. Ronchi, D. Nilsson, *et al.*, "Virtual reality for fire evacuation research", in *Proc. of Federated Conference on Computer Science and Information Systems (FedCSIS 2014)*, Jan. 2014, pp. 313–321. DOI: `10.13140/2.1.3380.9284`.

[154] J. N. Bailenson, N. Yee, D. Merget, and R. Schroeder, "The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction", *Presence*, vol. 15, no. 4, pp. 359–372, 2006. DOI: `10.1162/pres.15.4.359`.

[155] D. Roth, J.-L. Lugrin, D. Galakhov, *et al.*, "Avatar realism and social interaction quality in virtual reality", in *Proc. of 2016 IEEE Virtual Reality (VR 2016)*, 2016, pp. 277–278. DOI: `10.1109/VR.2016.7504761`.

[156] V. Kasapakis and E. Dzardanova, "Using high fidelity avatars to enhance learning experience in virtual learning environments", in *Proc. of 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2021)*, 2021, pp. 645–646. DOI: `10.1109/VRW52623.2021.00205`.

[157] D. E. Benrachou, M. Masmoudi, O. Djekoune, N. Zenati, and M. Ousmer, "Avatar-facilitated therapy and virtual reality: Next-generation of functional rehabilitation methods", in *Proc. of 2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP 2020)*, 2020, pp. 298–304. DOI: `10.1109/CCSSP49278.2020.9151528`.

[158] A. Schäfer, G. Reis, and D. Stricker, "A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems", *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–27, 2022. DOI: `10.1145/3533376`.

[159] E. Molina, A. R. Jerez, and N. P. Gómez, "Avatars rendering and its effect on perceived realism in virtual reality", in *Proc. of 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR 2020)*, 2020, pp. 222–225. DOI: `10.1109/AIVR50618.2020.00046`.

[160] M. Gonzalez-Franco and T. C. Peck, "Avatar embodiment. towards a standardized questionnaire", *Frontiers in Robotics and AI*, vol. 5, 2018. DOI: `10.3389/frobt.2018.00074`.

[161] F. Biocca, C. Harms, and J. L. Gregg, "The networked minds measure of social presence : Pilot test of the factor structure and concurrent validity", in *International Workshop on Presence, Philadelphia*, 2001, pp. 1–9.

[162] P. Caserman, P. Achenbach, and S. Göbel, "Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality", in *Proc. of 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH 2019)*, 2019, pp. 1–8. DOI: `10.1109/SeGAH.2019.8882429`.

[163] L. Gu, L. Yin, J. Li, and D. Wu, "A real-time full-body motion capture and reconstruction system for vr basic set", in *Proc. of 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC 2021)*, vol. 5, 2021, pp. 2087–2091. DOI: `10.1109/IAEAC50856.2021.9390617`.

[164] Placeholder Software, *Dissonance Voice Chat*, [accessed 19 April 2023]. [Online]. Available: `https://assetstore.unity.com/packages/tools/audio/dissonance-voice-chat-70078`.

[165] RootMotion, *Final IK*, [accessed 19 April 2023]. [Online]. Available: `https://assetstore.unity.com/packages/tools/animation/final-ik-14290`.

[166] N. Xi, J. Chen, F. Gama, M. Riar, and J. Hamari, "The challenges of entering the metaverse: An experiment on the effect of extended reality on workload", *Information Systems Frontiers*, vol. 25, no. 2, pp. 659–680, 2023. DOI: `10.1007/s10796-022-10244-x`.

[167] J. D. Hart, T. Piumsomboon, G. A. Lee, R. T. Smith, and M. Billinghurst, "Manipulating avatars for enhanced communication in extended reality", in *Proc. of 2021 IEEE International Conference on Intelligent Reality (ICIR 2021)*, 2021, pp. 9–16. DOI: `10.1109/ICIR51845.2021.00011`.

[168] G. Freeman, S. Zamanifard, D. Maloney, and A. Adkins, "My body, my avatar: How people perceive their avatars in social virtual reality", in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, 2020, pp. 1–8. DOI: `10.1145/3334480.3382923`.

[169] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo, "The effect of avatar appearance on social presence in an augmented reality remote collaboration", in *Proc. of 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, 2019, pp. 547–556. DOI: `10.1109/VR.2019.8797719`.

[170] C.-H. Rhee and C. H. Lee, "Cartoon-like avatar generation using facial component matching", *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 69–78, 2013.

[171] F. De Lorenzis, F. G. Pratticò, and F. Lamberti, "HCP-VR: Training first responders through a virtual reality application for hydrogeological risk management.", in *Proc. of 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2 (HUCAPP)*, 2022, pp. 273–280. DOI: `10.5220/0011007800003124`.

[172] G. C. Dobre, M. Wilczkowiak, M. Gillies, X. Pan, and S. Rintel, "Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings", in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*, 2022, pp. 1–7. DOI: `10.1145/3491101.3519628`.

[173] J. D. Hart, T. Piumsomboon, L. Lawrence, G. A. Lee, R. T. Smith, and M. Billinghurst, "Emotion sharing and augmentation in cooperative virtual reality games", in *Proc. Annual Symposium on Computer-Human Interaction in Play Companion (Extended Abstracts)*, 2018, pp. 453–460. DOI: `10.1145/3270316.3271543`.

[174] N. Hube, K. Vidackovic, and M. Sedlmair, "Using expressive avatars to increase emotion recognition: A pilot study", in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*, 2022, pp. 1–7. DOI: `10.1145/3491101.3519822`.

[175] Crazy Minnow Studio, LLC, *SALSA LipSync Suite*, [accessed 19 April 2023]. [Online]. Available: `https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442`.

[176] P. Ekman, “Are there basic emotions?”, *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992. DOI: `10.1037/0033-295X.99.3.550`.

[177] P. Heidicker, E. Langbehn, and F. Steinicke, “Influence of avatar appearance on presence in social VR”, in *Proc. of 2017 IEEE Symposium on 3D User Interfaces (3DUI 2017)*, 2017, pp. 233–234. DOI: `10.1109/3DUI.2017.7893357`.

[178] O. Otto, D. Roberts, and R. Wolff, “A review on effective closely-coupled collaboration using immersive CVE’s”, in *Proc. of Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications (VRCIA ’06)*, 2006, pp. 145–154. DOI: `10.1145/1128923.1128947`.

[179] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch, “Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR”, in *Proc. of 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2018)*, 2018, pp. 105–114. DOI: `10.1109/ISMAR.2018.00039`.

[180] A. Krekhov, S. Cmentowski, and J. Krüger, “The illusion of animal body ownership and its potential for virtual reality games”, in *Proc. of 2019 IEEE Conference on Games (CoG 2019)*, 2019, pp. 1–8. DOI: `10.1109/CIG.2019.8848005`.

[181] B. Salem and N. Earle, “Designing a non-verbal language for expressive avatars”, in *Proc. of 3rd international conference on Collaborative virtual environments (CVE ’00)*, 2000, pp. 93–101. DOI: `10.1145/351006.351019`.

[182] I. Kotlyar and D. Ariely, “The effect of nonverbal cues on relationship formation”, *Computers in Human Behavior*, vol. 29, no. 3, pp. 544–551, 2013. DOI: `10.1016/j.chb.2012.11.020`.

[183] G. Bente and N. C. Krämer, “Virtual gestures: Embodiment and nonverbal behavior in computer-mediated communication”, *Face-to-face communication over the internet: Issues, research, challenges*, pp. 176–209, 2011. DOI: `10.1017/CBO9780511977589.010`.

[184] S. J. Ahn and J. N. Bailenson, “Self-endorsing versus other-endorsing in virtual environments”, *Journal of Advertising*, vol. 40, no. 2, pp. 93–106, 2011. DOI: `10.2753/JOA0091-3367400207`.

[185] N. Hube, O. Lenz, L. Engeln, R. Groh, and M. Sedlmair, “Comparing methods for mapping facial expressions to enhance immersive collaboration with signs of emotion”, in *Proc. of 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct 2020)*, 2020, pp. 30–35. DOI: `10.1109/ISMAR-Adjunct51615.2020.00023`.

[186] I. C. Campbell, *This VR doc shows how vibrant virtual life already is without Meta’s meddling*, `https://www.inverse.com/input/reviews/we-met-in-virtual-reality-documentary-shows-metaverse-vibrant-without-facebook`, [accessed 19 April 2023], 2022.

[187] J. A. Russell, A. Weiss, and G. A. Mendelsohn, “Affect grid: A single-item scale of pleasure and arousal”, *Journal of personality and social psychology*, vol. 57, no. 3, p. 493, 1989. DOI: `10.1037/0022-3514.57.3.493`.

[188] L. Zhao, X. Lu, M. Zhao, and M. Wang, “Classifying in-place gestures with end-to-end point cloud learning”, in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2021)*, 2021, pp. 229–238. DOI: `10.1109/ISMAR52148.2021.00038`.

[189] L. Zhao, X. Lu, Q. Bao, and M. Wang, “In-place gestures classification via long-term memory augmented network”, in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2022)*, 2022, pp. 224–233. DOI: `10.1109/ISMAR55827.2022.00037`.

[190] A. Viegas, N. Tian, and R. Boulic, “Cybersickness assessment framework(CSAF): An open source repository for standardized cybersickness experiments”, in *Proc. of 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2023)*, vol. (in press), 2023, pp. 1–4. DOI: `10.1109/VRW58643.2023.00103`.

[191] H. Grace, *Neuralink Microchip Implant Can Put Users in Full Virtual Reality, Says Elon Musk: How True Is It?*, [accessed 14 June 2023], 2021. [Online]. Available: `https://www.itechpost.com/articles/108515/20211229/neuralink-microchip-implant-put-users-full-virtual-reality-elon-musk.htm`.

[192] D. Gilbert, *Elon Musk’s Neuralink says it has FDA approval for human trials: What to know*, [accessed 14 June 2023], 2023. [Online]. Available: `https://www.washingtonpost.com/business/2023/05/25/elon-musk-neuralink-fda-approval/`.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.