POLITECNICO DI TORINO
Repository ISTITUZIONALE

A testbed for studying cybersickness and its mitigation in immersive virtual reality

Terms of use:

(Article begins on next page)

18 November 2024

A Testbed for Studying Cybersickness and its Mitigation in Immersive Virtual Reality

Davide Calandra, *Student Member, IEEE,* Fabrizio Lamberti, *Senior Member, IEEE*

Abstract—**Cybersickness (CS) represents one of the oldest problems affecting Virtual Reality (VR) technology. In an attempt to resolve or at least limit this form of discomfort, an increasing number of mitigation techniques have been proposed by academic and industrial researchers. However, the validation of such techniques is often carried out without grounding on a common methodology, making the comparison between the various works in the state of the art difficult. To address this issue, the present paper proposes a novel testbed for studying CS in immersive VR and, in particular, methods to mitigate it. The testbed consists of four virtual scenarios, which have been designed to elicit CS in a targeted and predictable manner. The scenarios, grounded on available literature, support the extraction of objective metrics about user's performance. The testbed additionally integrates an experimental protocol that employs standard questionnaires as well as measurements typically adopted in state-of-the-art practice to assess levels of CS and other subjective aspects regarding User Experience. The paper shows a possible use case of the testbed, concerning the evaluation of a CS mitigation technique that is compared with the absence of mitigation as baseline condition.**

Index Terms—**cybersickness, testbed, virtual reality, evaluation, virtual environments, simulator sickness, taxonomy.**

I. INTRODUCTION

Over the last years, Virtual Reality (VR) has experienced a significant surge in popularity. This surge can be credited to the growing accessibility of powerful devices, specifically Head-Mounted Displays (HMDs) and their accessories, which are becoming more and more affordable. Nevertheless, despite these notable advancements, current VR technology continues to encounter several obstacles that can affect the overall User Experience (UX). A major obstacle is indeed represented by Cybersickness (CS), which should not be confused with Simulator Sickness (SS) [1].

SS refers to a condition linked to the use of advanced visual simulators and shares various symptoms with Motion Sickness (MS). Differently from MS, SS tends to be milder, less common, and arises from visual display elements and visuo-vestibular interactions that are not typical in real, i.e. physical, MS-inducing conditions [2]. CS, instead, consists of a combination of symptoms similar to those experienced in real MS, and is more generally associated with the use of VR. These symptoms may arise while using Virtual Environments (VEs) and often persist even once the exposure has ended, leaving individuals who are highly affected feel unsettled and disoriented [1]. Furthermore, unlike SS, in CS the symptoms

related to disorientation are more predominant, whereas those related to oculomotor issues are less prevalent [1]. Lastly, it has been demonstrated that the severity of CS effects is approximately three times greater than those of SS [1].

Hence, CS poses a significant challenge to the widespread acceptance and usability of the considered technology. In fact, it could hamper UX and limit the full potential of VR, especially considering that previous studies have reported that CS can be experienced by 22 to 80% of participants during or after exposure to immersive applications [3]. In response to this challenge, researchers and developers have been actively investigating and proposing various CS mitigation techniques [4], [5], [6]. These techniques aim to alleviate or reduce the adverse effects of CS, enhancing user comfort and overall VR experience. However, evaluating the effectiveness of these techniques presents its own set of difficulties [7]. In particular, upon reviewing the current literature on CS mitigation studies, it becomes evident that the proposed evaluation methodologies vary significantly, in terms of scenarios, tasks, and measurements [7]. Moreover, the complex nature of CS and its subjective effects make it imperative to establish common and robust procedures for studying the phenomenon and assessing techniques devised to mitigate it [8]. In this context, testbeds can provide standardized environments that can allow to replicate and control the factors contributing to CS, thus enabling rigorous investigations.

It is only very recently that the literature started to include works dealing with the mentioned needs. A notable example is represented by [8], where the authors presented a work-in-progress experimental environment named CS Assessment Framework (CSAF). CSAF is presented as a complete Unity-based package for studies on CS, equipped with the ability to manipulate factors that may contribute to it (e.g., locomotion, navigation path, visual stimuli), logging functionalities, as well as the possibility to load, save, and share its configuration. The framework is complemented by various standard locomotion and CS mitigation techniques, with the aim of providing developers and researchers with a tool to design new scenarios for CS studies (particularly of the "follow path of collectibles" type). While it represents a good starting point for designing new experiments, it does not, however, provide a representative set of scenarios, leaving their creation to the framework users.

By grounding on the above considerations, the present paper proposes a novel testbed for evaluating CS in fully immersive VR experiences. The decision to focus on fully immersive VR is motivated by fact that, in such context, CS holds much greater importance (and negative impact) compared to semi-immersive (e.g., based on CAVEs) and non-immersive setups (e.g., based on desktop PCs and monitors) [9].

The testbed aims to surpass the limitations of methodologies adopted in prior works by offering a comprehensive set of four scenarios, heavily inspired by the current literature, alongside normalized and standardized metrics across them. Additionally, it provides a protocol for conducting studies on the subject. To this purpose, a taxonomy of scenarios based on a survey of the state of the art in the field has been developed first, and then used to select the most relevant candidates to be included. Moreover, the testbed considers a wide range of factors that can contribute to CS. For instance, motion patterns (acceleration, deceleration, and rotation) and visual stimuli, which play a crucial role in triggering CS symptoms, are carefully tailored for each scenario.

To ensure a broad coverage of aspects possibly relevant for the analysis, the proposed testbed employs a combination of objective and subjective metrics. On the one hand, objective metrics provide quantitative insights regarding the physiological aspects of CS, as well as the impact of mitigation techniques on the proper execution of the tasks characterizing each scenario. On the other hand, subjective metrics, gathered through standard questionnaires, are meant to assess user's self-reported experience, allowing for a deeper understanding of CS and the effectiveness of adopted mitigation techniques.

Furthermore, the testbed is equipped with a set of standard CS mitigation techniques, and can be easily extended to support new ones. A use case demonstrating a possible way to leverage the testbed is also provided. In particular, it is used to demonstrate the evaluation of a new mitigation technique.

The ultimate goal of the proposed tool, available as open-source at https://github.com/VRatPolito/CET-VR, is to become a reference framework for researchers and developers carrying out studies on CS, letting them conduct systematic analyses on different setups and configurations. The goal is to contribute to the development and refinement of effective strategies to study and combat CS. The proposed testbed will facilitate the evaluation of the existing mitigation techniques, allowing researchers and developers to study their performance and user acceptance. These insights will also help in the design and implementation of improved mitigation strategies, ultimately leading to enhanced user comfort, improved VR experiences, and wider adoption of technology.

II. BACKGROUND

In order to understand and address the challenges posed by CS, extensive research has been conducted to investigate its causes, effects, and potential mitigation strategies. This section provides an overview of the current state of the art in CS mitigation in VR, highlighting key findings and advancements achieved so far.

A. Mitigation Techniques

Over the years, numerous techniques have been proposed in various literature works and VR applications with the aim of mitigating the undesirable effects caused by MS, SS, and CS. These techniques can be considered as belonging to some macro-groups, each distinguished by the type of approach employed.

1) Field-Of-View (FOV) reduction techniques: Alongside other factors, such as optical flow, the FOV directly influences the intensity and duration over time of the *vection* illusion [10]. A static reduction of the FOV, often achieved by applying a black vignette to frames (known as *vignetting*), aims to hide peripheral content in the user's visual field, potentially diminishing CS [10]. However, this comes at the expense of reduced visibility, possibly affecting also immersion and sense of presence. Alternatively, many studies focused on a dynamic reduction of the FOV [11], [12], [13], referred to as *Dynamic FOV*. This approach adjusts the FOV aperture based on the user's translational and/or rotational speed. The advantage is that, in stationary situations, the black vignetting is minimal or absent, reducing the negative impact of the technique when not necessary [12]. A variant known as *Circle effect* has also been presented, which replaces the black vignetting with static rendering of the scene captured from a slightly different viewpoint, providing a less noticeable alternative [14].

2) Blur techniques: These techniques aim to reduce vection by making the scene more abstract. Techniques like, e.g., *Rotational dynamic gaussian blur* [15] and *Saliency-based dynamic gaussian blur* [5] apply blur to frames based on user's motion or salient elements identified through AI methods. A particular kind of blur technique is the *Peripheral blur* [6], [16], which combines the benefits of vignetting (blurring is not applied across the entire frame) with the reduced visual impact of a blurred image, ensuring a good trade-off between mitigation and intrusiveness. Finally, *Texture blur* applies blur directly to 3D objects' textures, directing attention to relevant elements [17].

3) Rest frames: In these techniques, stable elements within the VE are introduced to reduce CS [18]. Examples include *Virtual Noses* [19], [20], *Virtual (or Simulated) CAVE* [21], [22], *Simulated Room* [22], and *Reticle* [6]. In some cases, rest frames can be also elements of the VE (e.g., cockpits, or headgears), and can be either static or dynamic [18].

4) Viewpoint repositioning: These techniques perform a recalibration of the camera based on user's movements in order to combat CS. *Head snapper* [23] involves fading to black and repositioning of the camera when rotational speed exceeds a threshold. *Head (or Vision) Lock* [21] uses eye tracking or manual input to lock the virtual world's movement to the user's line of sight during rotations.

5) Optical flow techniques: These techniques, like the *Dot effect* [14], manipulate the optical flow by introducing artificial elements (in that case, dots moving in the direction of the motion) to neutralize apparent motion in peripheral vision and reduce vection.

6) Alterations of the VE geometry: These techniques, such as *Geometric deformation* [24], [25], [26], [27] and *Geometric simplification* [28], can also reduce apparent motion in the peripheral FOV by applying various alterations directly to the VE. In Nie et al. [27], for example, tilting the user's viewpoint on virtual slopes is used to mimic real-world visual stimuli, reducing the risk of CS.

7) Depth-of-field techniques: Techniques like *Gaze-contingent depth-of-field* (or *Depth-of-field blur)* [6], [29]

attempt to alleviate the vergence-accommodation conflict by selectively blurring elements in the scene based on depth.

8) External hardware solutions: These techniques, based, e.g., on vibrating platforms [30] or Galvanic Vestibular Stimulation (GVS) [31], provide vestibular stimuli to reduce CS. While effective, they may be invasive, restrict movement, or involve additional costs.

Apart from this categorization, possible hybrid solutions can be adopted; these are combinations of other techniques that can draw characteristic elements from two or more macro-groups, aiming to achieve better results.

B. Taxonomy of Cybersickness Evaluation Scenarios

Alongside the increase in the number of CS mitigation techniques, there has also been the emergence of a vast and heterogeneous number of methodologies for evaluating CS and comparing the effectiveness of the above techniques, often completely different from one study to another. While the subjective metrics remained more or less consistent across various investigations (essentially, the Simulator Sickness Questionnaire [2]), the test scenario (i.e. the VE), as well as the tasks (i.e. the procedures to be executed in it), were often designed ex-novo with the aim of specifically stressing the characteristics of the considered technique, without necessarily drawing inspiration from what done in previous works. Furthermore, even when a previous scenario was taken into consideration, both intentional and non-intentional modifications were often introduced, e.g., due to the fact that source code was not available or implementation details were not fully described.

Nevertheless, by analyzing the scenarios employed in studies related to CS and its mitigation, a number of common features can be identified, allowing to organize them in a series of categories and subcategories. This analysis, which was actually performed to identify the most relevant scenarios to be included in the testbed proposed in the present paper, led to the definition of the taxonomy described in the following and depicted in Fig. 1.

1) Locomotion Technique: A primary categorization can be made based on the type of locomotion technique employed within the scenario. Locomotion is a significant aspect for most VR experiences [32], and can vary greatly depending on the specific scenario, the task the user needs to perform, and the size of the VE compared to the available physical space. Moreover, locomotion is closely associated with CS since, depending on the technique used, it can induce vection and consequently lead to discomfort symptoms. Referring to the categorization of locomotion techniques for VR presented in [33], the most suitable way for dividing scenarios in the study of CS is the one that classifies locomotion techniques as either "body-centric" or "vehicle-centric".

Body-centric scenarios are based on locomotion techniques that aim to simulate activities such as walking, running, etc., by having the user execute same or similar movements [33]. These movements can range from simple rotations in place to walking, running, and jumping.

Vehicle-centric scenarios take advantage of virtual vehicles to displace the user within the VE [33]. Generally, the vehicles used in CS studies are either motor vehicles or those involved in amusement rides (e.g., roller coasters).

A particular class of body-centric techniques is *redirected walking*, which enables natural movement in VR by altering the user's path to stay within the tracking area [32]. These techniques can be divided into subtle methods, which require specific VEs (e.g., indoor), and overt methods, which offer greater flexibility in terms of VE design but may reduce presence [26]. However, in works on redirected walking, CS is typically one of the aspects considered in the evaluation of the locomotion technique, not the main focus of the study.

2) Degrees-Of-Freedom (DOFs): Focusing on body-centric scenarios, it is possible to subdivide them into two main subcategories based on the DOFs of user's movements: "3-DOF (rotational)" and "6-DOF (with gravity)."

3-DOF (rotational) scenarios feature a fixed user's position within the 3D space, allowing primarily (or exclusively) rotations of the viewpoint around the three rotational axes (yaw, pitch, and roll). These scenarios are thus employed to evaluate CS mitigation techniques based on rotation.

The simplest task employed in this subcategory is undoubtedly the *Follow target*, for which an example is given in [12]. In that work, the authors assessed the performance of two Dynamic FOV implementations based on head motion (acceleration-based and velocity-based) in presence of amplified head rotations (the user's view rotates more in the VE than in the real world). To this purpose, they employed a cartoonish forest as test scenario, in which the user was immersed (standing, but without additional locomotion techniques). Taking inspiration from a previous work regarding re-orientation techniques in VR, an object moving around the users (a virtual butterfly) was used to stimulate their rotation as they follow it with their gaze.

Another example within this subcategory that is characterized, instead, by a high level of difficulty is the *Stationary shooter* scenario used in [23]. In this kind of scenarios, the user is positioned at the center of the VE, and has to face an increasing number of enemies approaching from all sides. The authors of [23] leveraged a First Person Shooter (FPS) as test scenario to evaluate viewpoint snapping while fighting a horde of zombies in an urban setting. In case the user's task is to protect an element of the VE from the enemies, this kind of scenarios is also referred to as *Tower Defense* [34].

Contrarily to 3-DOF scenarios, *6-DOF (with gravity)* scenarios allow the user to navigate through the 3D space of the VE. While allowing all six DOFs, movements along the vertical axis are generally more limited in these scenarios compared to the other two (typically, the user can ascend and descend by walking along a slope, and less frequently, can jump or climb scene elements). For this reason, one may refer to this subcategory as 6-DOF with gravity, to make it clear that it does not involve movements in total freedom (e.g., with a fly camera).

In this case, the scenario characterized by the greatest simplicity is that of *Free exploration* [19], [25], [29], [27]. The user is given the freedom to navigate through the VE, typically in rich settings such as urban landscapes [24], [25], [27], or buildings [19]. In [19], for instance, the Virtual Nose
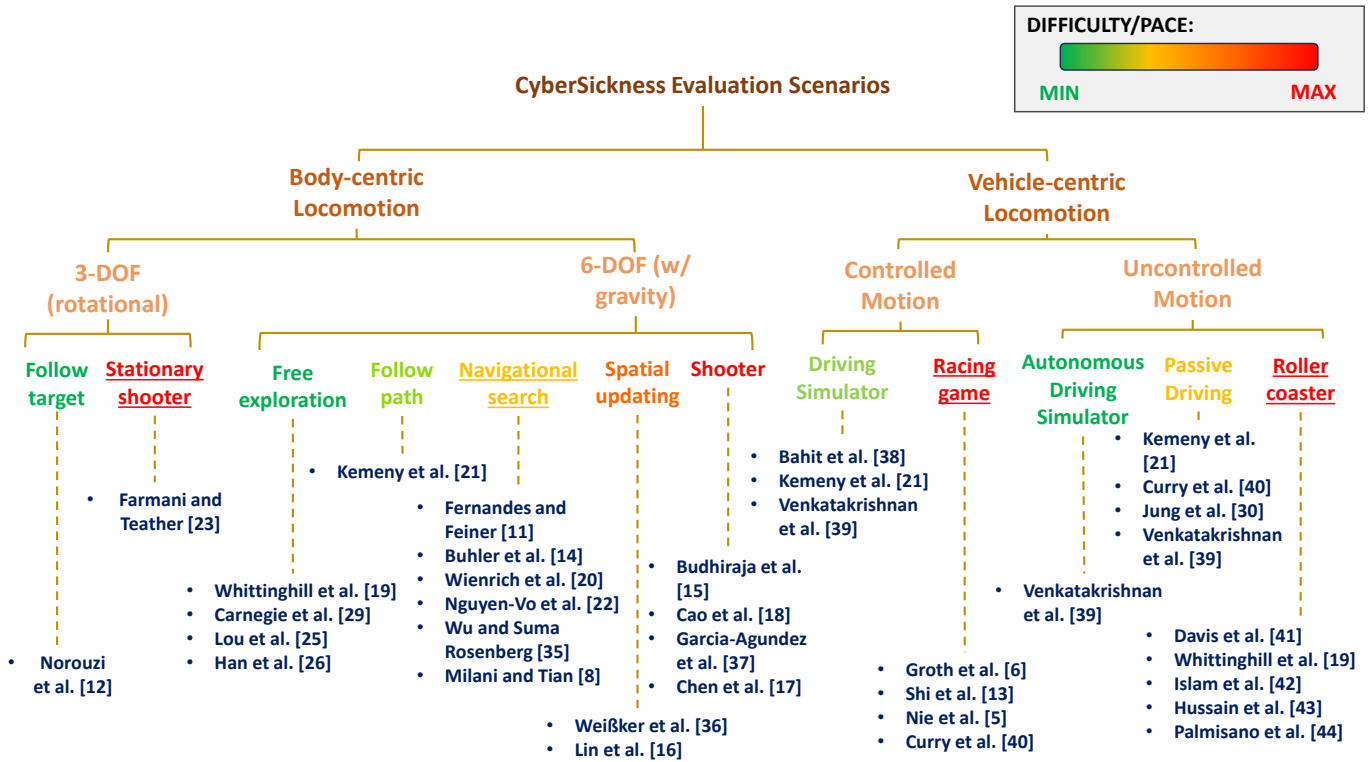
Fig. 1: Taxonomy of scenarios for the evaluation of CS (the scenarios whose name is underlined have been chosen for the development of the proposed testbed).

technique is evaluated by letting the user freely explore the interiors of the famous and no longer available Tuscany demo for the Oculus Rift DK1/DK2, realistically depicting a virtual villa set in the Tuscan countryside. No specific objectives are assigned to the user, except for the aim of spending as much time as possible in motion during the VR experience.

Moving up in difficulty, the *Follow path* scenarios [21] request the user to navigate a constantly visible virtual path. The aim is to minimize aimless wandering, which could hinder the generation of visual stimuli capable of triggering CS. For instance, in [21], the authors utilized a hangar environment to assess the performance of two mitigation techniques (VirtualCave and Head Lock). The hangar incorporated realistic distances and sizes, with scattered obstacles defining a specific route; additionally, red arrows were strategically positioned on the floor and walls to guide the user along the path.

If, instead of following a path, the user is asked to search for something within the VR environment, then the scenario is referred to as *Navigational search* [8], [11], [14], [20], [22], [35]. This kind of scenarios proves to be the most common in CS studies, and comes in several variants. The most prevalent one is the "Collect coins to buy your way" [8], [11], [20], [35], where the user needs to gather a specific number of objects scattered throughout the VE. In this variant, the next coin is often placed in a location easily visible or accessible to the user, preventing free roaming. An example of such a scenario can be found in the study reported in [20] to investigate the Virtual Nose technique. The authors developed an outdoor woodland setting with simplified graphics, requiring the user to collect coins through activities like walking, balancing on bridges of varying sizes, and jumping from platform to platform, using a Xbox controller to perform smooth locomotion within the VE. A similar approach was followed by the authors of [11] and [35] to evaluate their Dynamic FOV implementation. In particular, in [11], the user was kept seated, locomotion was managed using a joystick, and the vignetting rate was driven by both angular and translational velocities. In this case the task was a sequential waypoint navigation. The test scenario was again the Tuscany demo, but modified to provide access to the land outside the villa. To enhance the outskirts, several objects were introduced, such as an additional house featuring a smoking chimney. The waypoints were represented by posts surrounded by a particle effect, visible one at a time (the next one appeared upon reaching the current one). In [35], instead, the user was placed inside a maze and had to collect a certain number of always present coins distributed in the VE.

In other cases, the user is asked to search for specific objects, all present but in a smaller number and more challenging to find. As example of this scenario is provided in [14], where the authors leveraged another popular Unity demo scenario (*Viking Village*, set in a realistic and rich Viking-inspired settlement) to evaluate their two mitigation techniques (Circle and Dot Effect). The considered VE features real-time lighting, dynamic shadows, and advanced rendering techniques, and it is filled with numerous assets like houses, boats, and other props. In this case, 20 numbered blue boxes were scattered around the village, and the user was asked to collect them in ascending order.

A particular kind of Navigational search scenarios, characterized by an increased complexity in terms of cognitive load, is the *Spatial updating* one [16], [36]. The user is shown two identical objects at different times and positions within the VE. When an object of a certain type is seen for the first time, the user must try to remember its position; as soon as the second object of the pair is seen, an audio signal is played, prompting him or her to indicate the position of the identical object seen previously. As a matter of example, the authors of [16] used this task within a maze scenario to study the impact of a Static peripheral blur in the user's FOV on CS reduction. This particular task was chosen to assess the user's acquisition of spatial knowledge and to force rotational movements. Previously, the same task had been used in [36] to study the CS induced by two common locomotion techniques (smooth locomotion and teleporting in a virtual urban setting).

Finally, *Shooter*-type scenarios used in [15], [17], [18], [37] can be regarded as characterized by both higher difficulty and pace compared to the previous ones. For instance, the authors of [15] leveraged an open-source Unity scenario (*AngryBots*) to evaluate Rotational dynamic gaussian blur. The scenario was a shooter game set in an industrial environment and featuring various bots scattered throughout the VE. Although this can be considered as a very representative example of a gaming scenario, it employed a now-dated VR setup consisting of HMD, mouse, and keyboard. The evolution of VR interfaces, with the introduction of hand controllers, and the advent of room-scale movements have led to a significant differentiation, in terms of game design, between classic shooters, playable with mouse and keyboard on a flat screen, and those oriented towards immersive VR. Therefore, this particular scenario may be no longer suitable for the purpose today. This limitation was overcome by the authors of [18], who opted for a shooter game specifically developed for VR (*VR Apocalypse*), which however was not free nor open-source, and is no longer available for purchase. The goal was to evaluate two Rest frame techniques, a classical (static) one and a dynamic variant that increases visibility based on the user's motion. A shooter game was employed also in [17] to assess the Texture blur technique. In this case, a straight urban track setting was adopted, in which the user was required to shoot enemies while avoiding their attacks and collecting jewels.

3) Degree of Control: Moving into the category of vehicle-centric scenarios, a further subdivision can be based on the level of control the user has over the virtual vehicle's motion. The scenarios can be either based on "Controlled motion", as the user has complete control over the vehicle, or on "Uncontrolled motion", as there is little to no possibilities to intervene in its movement.

Controlled motion scenarios are the most common in works that investigated CS. Generally, the vehicles used are cars. However, theoretically, other vehicles could be considered too. Such scenarios can be of two types, based on the level of difficulty and pace of the experience: "driving simulators" or "racing games".

Driving simulator scenarios [21], [38], [39] are characterized by a simpler and more relaxed experience. The user drives a vehicle, e.g., in an urban scenario, while adhering to traffic rules and staying aware of other vehicles. These scenarios are generally not used to study the CS mitigation techniques but, rather, CS in general. In [38] and [21], for instance, this task is employed to assess the CS induced by driving simulators in immersive VR setups. In [40] and [39], instead, the authors study the difference in terms of CS in a driving simulation scenario with and without motion control (as a driver and as a passenger).

Racing game scenarios [5], [6], [13], [40] are characterized by significantly higher difficulty and pace compared to the previous ones. These very common scenarios draw inspiration from racing games, thus not limiting themselves to realistically simulating a race, but also incorporating gaming mechanics such as collecting (or destroying) objects. In [6], this kind of scenario is used to evaluate two mitigation techniques, i.e. Peripheral blurring and Vignetting, asking the user to drive around a circular track for 10 minutes, with the option to withdraw from the experiment in case of severe CS symptoms. The authors of [13] selected a racing game as scenario for comparing three other CS mitigation techniques (Dynamic FOV, Depth-of-field blur, and Reticle). A seated user either drove a virtual car using a joystick, or assumed a passive role, experiencing the scenario from a first-person perspective. The virtual car's model was intentionally kept hidden to prevent it from serving as an additional rest frame, which could be a confounding factor. Various red balloons were strategically positioned in mid-air, and the user was required to press the controller trigger upon sighting each one of them. To prevent multiple balloons from appearing simultaneously in the same view, one balloon was showed at a time (only after the user had passed a balloon's location and it was no longer visible, the next balloon became visible). The authors of [5] also made a similar choice for the evaluation of their mitigation technique (Saliency-based dynamic gaussian blur). The task consisted of trying to hit as many pink walls (scattered around the track, each composed of various physics-enabled bricks) as possible. The walls were automatically repaired shortly after the collision to be available for the next lap. The track was also equipped with signs indicating dangerous curves. These elements (walls and signs) were particularly suitable to be recognized as salient by the proposed algorithm and, therefore, excluded from blurring. Again, to avoid unintended rest frames, the scenario did not include the visualization of the vehicle driven by the user.

Uncontrolled motion scenarios, in turn, include all the cases in which the user is placed on a moving vehicle within the VE without having complete or even minimal control over its movement. Further subdivisions can be made based on the pace of the experienced movements.

The least demanding scenario is the *Autonomous driving simulation*, used, e.g., in [39]. In this case, the user is positioned in the driver's seat of an autonomous vehicle, which follows a specific route in a typical urban context.

Passive driving scenarios are typically more intense [21], [30], [39], [40]. Differently from the previous case, the vehicle is not simulated as a real autonomous vehicle. Instead, a session of manual driving is pre-simulated (or pre-recorded) and then played to be experienced passively by the user. The

user can be positioned either in the driver's seat [39] or the passenger's seat [30]; if the vehicle's cockpit is hidden, this difference becomes imperceptible. The setting can also vary, from an urban environment [21], [39], to countryside roads [30], up to race tracks [40].

The most extreme level for this subcategory is reached by *Roller coaster* scenarios [19], [41], [42], [43], [44]. This category has been extensively studied, both for the intensity of visual stimuli on the user, as well as because it is a very common scenario among VR entertainment applications (especially demos). The user is placed on the cart of the aforementioned amusement ride, and is moved along a path consisting of sudden ascents, descents, curves, loops, and twists, thus providing a much more varied 3D motion compared to vehicle driving scenarios. Usually, the duration of the experience is pre-planned, and the user has the option to end it prematurely if experiencing severe CS symptoms.

As confirmed from this review of the state of the art, over the years the nature of CS has led to the development of numerous techniques for mitigating the negative effects experienced by the users during prolonged exposure to VEs. However, these techniques often consist of solutions to specific problems, and may require particular conditions for optimal functioning. For the above reasons, the scenarios adopted to demonstrate their effectiveness vary dramatically from one work to another, making it almost impossible to compare results from different studies. Therefore, there is a need for a mechanism to constrast individual techniques under a common range of scenarios, in order to assess their effectiveness in a more generalized way. The goal of the present work was to develop a testbed able to induce CS in a relevant set of scenarios representing the majority of situations typically considered in the literature, in order to serve as a reference for future studies in the context of CS and its mitigation.

III. METHODOLOGY

This section outlines the methodology adopted to develop the proposed testbed, presenting the scenarios and tasks selected for the purpose, describing the objective and subjective metrics considered to support evaluation, and finally discussing the experimental protocol devised for utilizing it.

A. Selection of the Scenarios

Building upon the analysis of the state of the art in the field and, in particular, on the proposed taxonomy, a representative set of four scenarios was identified. The selection process involved choosing one scenario per taxonomy sub-category, prioritizing the most popular and comprehensive ones. When popularity was deemed comparable, the choice was based on interaction complexity, task variety, and pace. This approach should reasonably ensure that the discarded scenarios can be regarded as simplified versions of the selected ones.

1) Tower Defense (TD): For the 3-DOF (rotational) body-centric subcategory, the choice fell on a Stationary shooter. Such scenario, compared to a simple Follow target, allows for more intense and less predictable rotational movements. It also ensures greater engagement for the user, given the primarily gaming task, further motivating him or her to stay in motion. In this case, a configuration in which the user has to defend one of the elements in the VE was considered.

2) Navigational Search (NS): As for the 6-DOF (with gravity) body-centric subcategory, purely considering the level of difficulty and pace of the experience the most suitable scenario may appear to be the Shooter. However, it has been found that Shooter scenarios have two main drawbacks. Firstly, the combined use of standard locomotion techniques (based on controllers) and of a firearm can create quite a few problems for VR users with little or no experience, who are often the target of studies on CS. Secondly, previous investigations are almost all carried out on outdated VR configurations (e.g., seated user with a gamepad [37] or mouse and keyboard [15]), and do not represent the most commonly employed scenario in this subcategory. Hence, the focus was put on the most explored type of scenarios, i.e. the NS. Due to the large number of works that make use of these scenarios, the related characteristics can be very heterogeneous. Thus, inspiration was sought from various alternatives to define a scenario as much comprehensive as possible; these choices will be detailed further below.

3) Track Race (TR): Regarding the Controlled-motion vehicle-centric subcategory, it was decided to focus on the type of scenarios that emerged to be both more explored and more complex from the experiential point of view, i.e. Racing games. Like in the previous case, efforts were made to draw features from various implementations in order to achieve a high level of representativeness.

4) Roller Coaster (TR): A similar reasoning was made to chose the scenario representing the Uncontrolled-motion vehicle-centric subcategory, where Roller coasters emerged as the best candidates. In this case, in addition to drawing inspiration from previous works based on this setting, an attempt was made to introduce a task originally proposed for a different type of scenarios (a Racing game [13]) and involving the use of controllers, thus giving the user a specific reason to keep his or her gaze on the VE during motion.

The current version of the testbed was developed with Unity 2021.3 and OpenXR for enhanced interoperability. The testbed operates in tethered mode, allowing experimenters to manage settings and commands using mouse and keyboard. Some screenshots of the four scenarios are shown in Fig 2, whose caption also includes a link to videos of executions.

To facilitate comparisons among existing mitigation approaches, the testbed was also endowed with a significant number of publicly available mitigation techniques. In particular:

- The *GingerVR* repository [4] was integrated, to provide already numerous techniques proposed in the literature.
- The free *VR Tunnelling Pro* Unity asset was included, allowing interesting variations and customizations of the original *Dynamic FOV*;
- Additional techniques, not openly available (i.e., *Circle effect* [14] and *Gaze-contingent depth-of-field* [29]), were implemented, following the authors' descriptions in the corresponding papers.
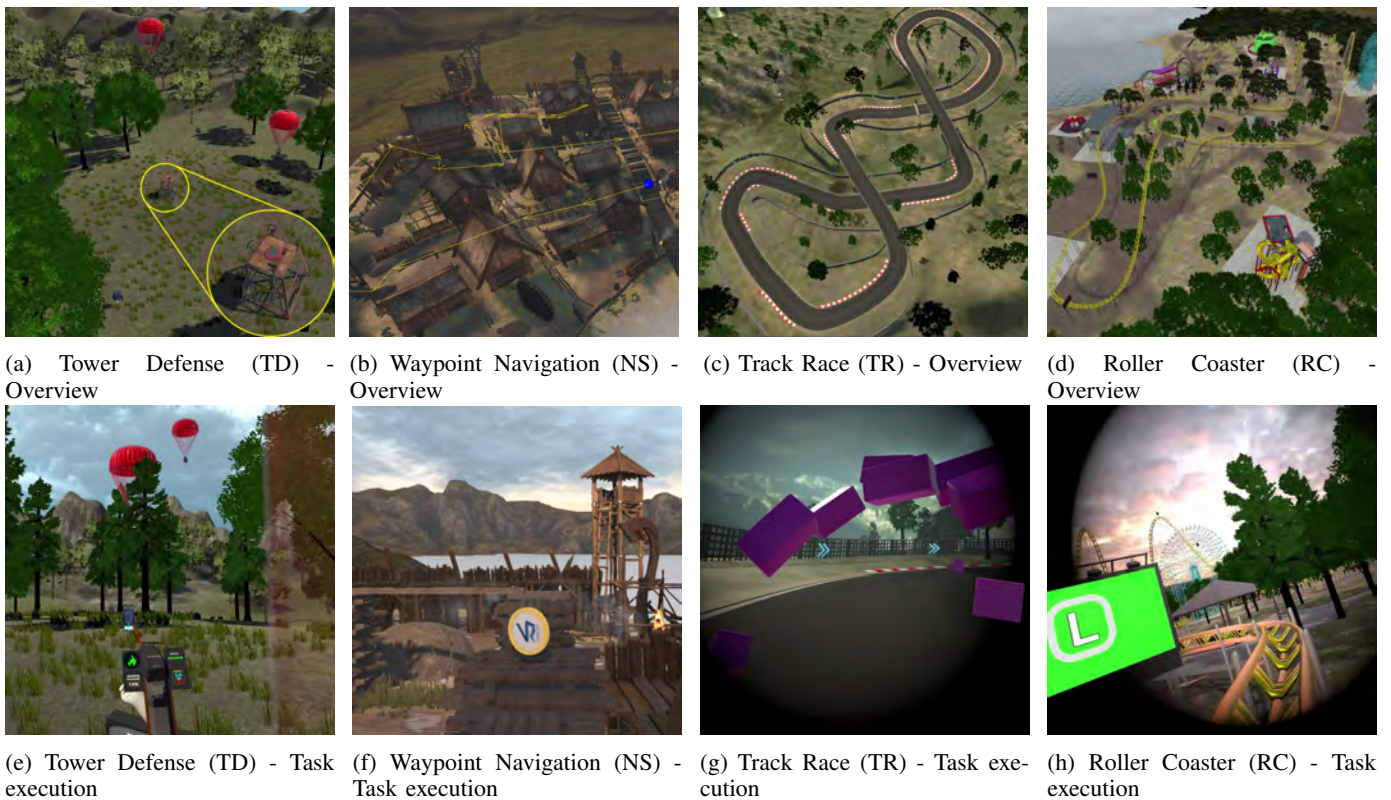
(a) Tower Defense (TD) - Overview

(b) Waypoint Navigation (NS) - Overview

(c) Track Race (TR) - Overview

(d) Roller Coaster (RC) - Overview

(e) Tower Defense (TD) - Task execution

(f) Waypoint Navigation (NS) - Task execution

(g) Track Race (TR) - Task execution

(h) Roller Coaster (RC) - Task execution

Fig. 2: The four scenarios of the proposed testbed (some videos of task execution are available at http://tiny.cc/64rdzz).

B. Details of the Scenarios and Tasks

Hereafter, a detailed description of the four scenarios and tasks to be performed is provided.

1) Tower Defense (TD): As mentioned before, the 3-DOF (rotational) scenario is a Stationary shooter [23], in the context of a TD setting. The VE comprises a natural landscape (Fig. 2a) with numerous hills, trees, and bushes in close proximity to the user. On the one side, they are clustered in a dense forest, on the other side they form a more open field. Additionally, there are mountains at greater distances, marking the boundaries of the environment.

At the center of this landscape there is a watchtower, about 4m high. At the top of the tower there is a rotating platform enclosed by a glass wall, covering three sides out of four, leaving only one frontal opening. The user is placed inside these walls, and can move only within a cylindrical volume with a diameter of 80cm positioned at the center of the platform, allowing him or her to freely rotate his or her head but not to leave the enclosed area. The platform can be freely rotated with a speed between 0 and 360°/s by using the touchpad (or the thumbstick) of the non-dominant hand's controller (left/right). However, the rotation is allowed only when the user is facing the frontal opening. In such a way, the user is forced to experience the visual effect of the artificial rotations, which could have been otherwise mitigated by keeping the eyes on the target while the platform turns. This approach combines stimuli from real-world self-rotations with those from artificial rotations, similar to those examined in earlier literature on VR setups with seated users [23].

The user is provided with a virtual weapon (Fig. 2e), positioned in the dominant hand, to complete the task associated with the scenario. This weapon is a semi-automatic handgun with an overheating function (if too many shots are fired consecutively, the user must wait a few seconds before resuming firing). The weapon also has a laser pointer that is disabled automatically when pointed against one of the platform's walls or during weapon overheating.

The task consists of two rounds, each lasting five minutes, during which the user must shoot down moving targets. These targets are displayed with bright, simple colors for easy identification. In both the rounds, the targets, represented as blue robots, are spawned one by one at distant points in the environment. The time intervals between spawns progressively decrease, and the robots move toward their objective with speeds that gradually increase over time.

In the first round, a horde of ground enemies approaches the tower from different directions in a 360° angle, and the user's goal is to shoot them down using the virtual weapon before they reach the tower. Each target that reaches its objective marks a point for the enemies. In the second round, the goal remains the same, but enemies now include both ground and aerial targets (i.e. robots equipped with parachutes). While the ground targets behave as in the previous round, for the aerial ones the user must hit them (or their parachutes) to bring them down before they land (each landing marks one point for the enemies too). The user earns points for each enemy shot down, aiming to outscore the enemies by eliminating as many as possible using the fewest bullets. Only one bullet per

target is required to eliminate it.

This scenario is designed to evaluate CS in a situation in which the users experience abrupt rotational changes, either natural or artificial, to engage targets at different height levels. Studies suggest that such rotational movements are more likely to cause CS compared to other types of motion [15].

The environment appears simple, but virtual objects have moderate details to increase perceived optical flow during rotational movements, potentially leading to higher levels of CS for the users [14], [15], [17].

2) Navigational Search (NS): The 6-DOF (with gravity) body-centric scenario included in the testbed is a NS in the form of sequential waypoint navigation, where waypoints are represented by coins [20]. The VE that constitutes this scenario consists of the large Viking-style village[1] already used in [14] (Fig. 2b), located along a coastal area with mountains in the background acting as a visual barrier. The user is tasked to explore an environment filled in with visual nuances, encountering various structures like houses, huts, and piers. The VE comprises well-ornamented outdoor spaces with natural daylight simulation, as well as indoor spaces with intricate details and less intense lighting. It should be noted that various areas of the original *Viking Village*, particularly those related to the final part of the coin collection lap, have been modified to add complexity from the movement perspective. Within the path that the user must follow to complete the task there are also some elevations and obstacles, such as axes, stairs, bridges, wooden beams to avoid, rocks, and narrow passages (Fig. 2f), which the user must overcome by running, jumping (using a button on the hand controller), or crouching, taking inspiration from the task proposed in [20].

The user can move freely within the boundaries of the village using smooth locomotion at speeds ranging from 0 to 3.5 m/s, increasing to 7 m/s when sprinting, as per the implementation in [32]. No artificial rotation is provided. Movement and speed modulation are managed using the hand controller, particularly the thumb-stick (or, if not available, the pad). The chosen implementation follows the one employed in [32], where movement occurs in the direction indicated through the controller. In the development of the testbed, the ability to jump by pressing the trigger and to sprint by holding the grip were added. This locomotion technique was chosen mainly because it is now commonplace in VR applications; however, it could be easily replaced with any other implementation that allows for generating different inputs for walking, running, and jumping without necessarily requiring controllers.

Regarding the task, the user has to follow a predefined path marked by a sequence of virtual coins of considerable size (Fig. 2f) which have to collected. The coins appear one by one and represent individual points of interest that the user must reach during the experience, as done in [11]. The objective is to complete two laps of a cyclic path while gathering these coins. A total of 53 coins are positioned at comparable distances (min: 4.9 m, max: 33.4 m, avg: 12.8 m), ensuring a path that spans the entire village (tot: 667.4 m). Importantly, these coins remain in the same locations for each lap. These coins, revealed sequentially upon collection, are positioned in close proximity to each other, although in some cases an exploratory activity is required to locate the next one. While executing the task, the user may be requested to overcome a certain number of obstacles, like climbing stairs, walking on gangways, balancing on wooden bars, jumping ditches and onto rooftops, physically crouching, and crawling. To prevent the user from getting lost in this search, if the next coin is not collected within one minute after the previous one, a blue arrow indicating the direction to proceed in the experience is shown. The main goal for the user is, therefore, to complete the task by collecting all the coins in the shortest time possible. The first lap is mainly intended for understanding the positions of the coins and the actions necessary to collect each of them, allowing for an increased execution speed in the second lap.

The devised NS scenario includes features of Follow path scenarios (the coins form a unique path), of Waypoint navigation scenarios (the coins appear sequentially), of Spatial updating scenarios (the user needs to try to remember the position of each coin), as well as of Navigational search scenarios (not all the coins are immediately visible when they appear). Like the other scenarios in this subcategory, the NS one allows for the evaluation and identification of the difficulties encountered by the users during experiences involving obstacle overcoming and object searching within a large VE, where they move freely without the use of vehicles or other supports. Specifically, this choice enables the observation of the intensity of symptoms caused by the contrast between translational movements within the VE (performed through the joystick) and the stationary standing position maintained by the users in the real world.

3) Track Race (TR): In the Controlled-motion vehicle-centric scenario, the user find himself or herself inside a race circuit (Fig. 2c) consisting of curves, ascents, descents, and bridges. The particular task selected for this scenario is inspired by the one presented in [5].

As a base for the VE, one of the tracks available in the *Race Tracks*[2] Unity asset was used, to which the *MS Vehicle System Free*[3] asset was added for vehicle management, appropriately modified to work in seated VR. In particular, the two main input axes (accelerate/brake and steering) were mapped to the two hand controllers' touchpads (or thumbsticks). The right controller's pad (up/down) was used to accelerate or brake/reverse gear, whereas the left controller's pad (left/right) was used to control the steering wheel. Finally, the handbrake (useful for drifting) was bound to the two controllers' grips. Like in the previous scenario, the majority of inputs could be easily reassigned to another type of interface (e.g., a gamepad).

The boundaries of the circuit are represented by metal grids, on which further indications of the direction to be maintained within the circuit were added (in form of intense blue arrows). Trees and road signs are also present at the edges of the track, serving as both decorative elements and virtual obstacles that can affect the vehicle's movement. The car in which the user is placed is kept invisible to him or her, in order not to

---

[1]Viking Village: http://tiny.cc/ey09zz

[2]Race Tracks: http://tiny.cc/lo98zz

[3]MS Vehicle System: http://tiny.cc/no98zz

provide a reference frame. However, it still interacts with the surrounding environment.

Within the track there are 15 pink brick walls (Fig. 2g), easily identifiable in the VE, which are associated with a physics simulation triggered by the impact of the user's vehicle, similarly to what is done in [5]. In case of impact, they return to their initial state after a predetermined time, allowing them to be interacted with in subsequent laps. Finally, auditory and visual cues regarding the vehicle's acceleration, braking, and collisions with the mentioned pink walls are provided.

The main goal of the user is to complete a total of 5 laps in the shortest time possible, and thus, keeping the vehicle on the track for as long as possible. The first lap is considered as a practice one, and does not contribute to the calculation of performance metrics, as done in similar works [5], [13]. At an appropriate speed, the time to complete one lap around the circuit is approximately one minute. As mentioned earlier, the circuit unfolds with various elevations and curves, making the main task of the scenario an "Uphill/downhill driving", referring to a driving experience in which the user navigates parts of the track at different levels of elevation [14], [20]. During the race, the user also has the additional goal of breaking as many pink walls as possible by directly hitting them with the vehicle. Finally, he or she must complete the specified laps while minimizing collisions with billboards, trees, and metal grids (the pink walls are excluded), as well as minimizing the time spent off the track.

The usefulness of this scenario lies in its ability to assess the severity of symptoms and the challenges faced in a high-speed driving experience where the users have full control over the vehicle's movement [14], [20]. Furthermore, the need to hit virtual elements with the vehicle allows for evaluating the users' response to impacts and interactions with VE, while moving at relatively high speeds. Finally, the presence of obstacles and bumps capable of altering the users' motion direction allows for assessing the reaction to unexpected changes in movement not initiated by own actions.

4) Roller Coaster (RC): As previously said, roller coasters are among the most popular scenarios in VR and, since they are naturally inducing CS, they have been repeatedly considered for evaluating mitigation techniques [19], [41], [43], [44], [45]. Therefore, they have been chosen to represent the Uncontrolled-motion vehicle-centric scenarios. The RC scenario included in the testbed features an amusement ride characterized by short ascents, steep descents, wide curves, a loop of death, and a series of twists and turns including upside-down sections, leading the user back to the starting point for the next lap. During the virtual experience, the user is kept seated, can look around, and has two goals.

Firstly, in order to make the experience more actively engaging and give the user a reason to observe the moving elements of the VE, it was decided to introduce an observational task, initially proposed in a different context. In particular, the user must press a correct sequence of buttons, different for each lap, displayed one at a time above a number of billboards of different sizes, positions, and orientations distributed along the path. The billboards, initially black, light up one by one, following the sequence, turning blue and indicating the button

to press. A correct press in the sequence causes the associated billboard to light up green and emit a sound identifying its correctness, whereas an incorrect press lights up the billboard in red, indicating an error. The inspiration for this task came from the task proposed in [13], employed in the context of a Racing game scenario, but with a similar purpose. To simplify the experience, the buttons to press consist of the left trigger, identified on the billboards by the letter "L," and the right trigger, identified by the letter "R" (Fig. 2h), on the two hand controllers, making them easily recognizable even by less experienced users. As a result, for a successful run, the user must complete the circuit while making the fewest errors possible in the button-press sequence.

Then, as secondary task, the user must complete the entire path, consisting of three complete laps, as quickly he or she can. This is possible because the cart speed is not entirely fixed, and the user is allowed to increase it from the base value up to doubling it, and vice versa, by operating the touchpad (or the thumbstick) of the dominant hand's controller (up/down). The user is not allowed to decelerate the cart below the base speed and, as a result, cannot bring it to a stop. The base speed value is 15m/s, which can be increased up to 30m/s. Moreover, the actual speed coincides with the target speed only in straight sections. In fact, it can vary along the track depending on the cart's position (going uphill, the cart will slow down below the target speed, whereas going downhill, it will accelerate). The choice to provide this minimum level of control over the motion compared to a completely passive experience was made to obtain a performance indicator related to the execution time. The user is therefore inclined to make the experience more fast-paced (but not less) in order to better perform the required task. At the same time, the onset of CS symptoms can lead to reducing this speed to the base value, impacting task performance.

This scenario aims to stress the conditions that users would face during a vehicle experience in which they have minimal control over movement (e.g., as passengers). The VE includes ornamental elements and details positioned at a moderate distance, and the cart's high-speed movements create significant conflicts between visual and vestibular inputs. Finally, the use of vibrant colors and detailed models aims to intensify symptoms, adding visual stress to the VE [15].

C. Metrics

In this section, all the metrics considered by the testbed will be presented and discussed. In particular, it was decided to use both objective metrics, specifically defined for the considered scenarios and tasks, as well as subjective metrics, based on standard questionnaires from the literature.

1) Objective Metrics: For what it concerns objective metrics, the challenge was to define common indicators for the four scenarios, while ensuring the ability to measure diverse aspects depending on the task context. Consequently, inspiration was drawn from the approach used in [32]. Specifically, for each task, four metrics were defined (three related to task performance, and one related to physiological conditions): *Operation Speed* (OS), *Accuracy* (AC), *Error Proneness* (EP),

and *Delta Hearth Rate* (ΔHR). The three task performance metrics are normalized in a range between 0 (worst) and 1 (best); for AC and EP, the way they are measured varies based on the considered task. ΔHR, in turn, is a non-normalized value, representing the difference between the heart rate measured at the beginning and at the end of the experience.

In the case of the TD scenario, which has a fixed duration, the OS metric loses its meaning, and therefore, it is not considered. The AC metric is defined as:

$$AC_{TD} = \frac{b_h}{b_f} \tag{1}$$

where and b_f and b_h are respectively the number of bullets fired by the user and the numbers of bullets which actually hit an enemy. The EP metric is calculated as:

$$EP_{TD} = 1 - \frac{n_c}{N_t} \tag{2}$$

where n_c is the number of enemies which reached their relative objective (the tower, or the ground), and N_t is the total number of spawned enemies. In case of premature quitting of the experience due to extreme CS, N_t is set to the number of enemies spawned till that time.

Moving to the NS scenario, the OS metric is calculated as:

$$OS_{NS} = 1 - \frac{t_c - T_{min}}{T_{max} - T_{min}} \tag{3}$$

where t_c is the actual completion time for the experience, and T_{min} and T_{max} are calculated as follows. Firstly, an optimal path length is pre-computed, by summing up the 3D distances between each coin and the next one. Then, the optimal path length (in meters) is divided by the minimum walking speed (0.55m/s) for the calculation of T_{max}, and by the maximum sprinting speed (7m/s) for T_{min}. These speeds pertains to the joystick-based locomotion technique; the minimum speed was determined based on the results of the pilot study, whereas the maximum speed was defined in accordance with the implementation used in [32], based on real-world values for humans. If $t_c > T_{max}$, t_c is clamped at T_{max}. The only other performance metric for this scenario is the AC one, defined as:

$$AC_{NS} = \frac{d_{min}}{d_t} \tag{4}$$

where d_t is the total distance traveled by the user, and d_{min} is the optimal path length.

For the TR scenario, the OS, AC, and EP metrics are all defined. OS is calculated like in the NS scenario; in this case, the optimal path is obtained by summing up the distances between the various pink walls. The minimum speed was again adjusted based on the results of the pilot study (20 Km/h), whereas for the maximum speed a value was selected that is realistic for a vehicle of that type (180 KM/h), yet very challenging to sustain throughout the entire circuit. The AC metric is computed as:

$$AC_{TR} = \frac{m_e}{m_h} \tag{5}$$

where and m_e and m_h are respectively the total number of pink walls encountered in the circuit, and the number of pink walls hit by the user. Finally, the EP metric is computed as:

$$EP_{TR} = 1 - \frac{t_e}{t_c} \tag{6}$$

where t_e is the time spent being mistakenly off-track or in contact with obstacles other than the pink walls, and t_c is the total completion time.

Finally, for what it concerns the RC scenario, the OS metric is calculated like in Eq. 3, but T_{max} and T_{min} are the completion time at the maximum and minimum allowed speed, respectively; these speeds were selected to achieve a reasonable scenario duration ranging from 3 to 6 minutes. The AC metric is computed as:

$$AC_{RC} = \frac{p_c}{C} \tag{7}$$

where p_c is the number of correctly executed button pressures in the displayed sequence of billboards, and C is the total number of billboards in the sequence. In case of premature quitting, C is set to the total number of billboards already encountered along the user's path. In this case, the EP metric is not used, as it would be very similar to the previous one.

To summarize, AC_{TD} measures the accuracy with which the user hits enemies, AC_{NS} assesses how close the user gets to the shortest path in collecting coins, AC_{TR} evaluates accuracy in hitting pink walls on the race track, and AC_{RC} measures the user's ability to maintain focus on billboards along the roller coaster. Additionally, EP_{TD} counts the number of enemies the user fails to hit, EP_{TR} evaluates how well the user stays on the race track route, whereas OS_{NS}, OS_{TR}, and OS_{RC} all measure the execution speed.

The purpose of these metrics is to understand if a high level of CS may have negatively influenced the user's performance or, in the opposite case, if a particularly low level of CS may be associated with a user not actively or correctly performing the task (e.g., for example, closing the eyes to avoid vection without communicating it to the experimenter).

For what it concern the ΔHR metric, it is necessary that the user wears a measuring device that can record heartbeats at the beginning and end of each scenario (the devised protocol does not prescribe the use of a specific device). In [32], the difference between the heartbeat values was used to estimate the amount of physical effort required in the virtual experience. In the context of interest for the present work, it can instead provide useful information about the level of CS, considering that hearth activity and variability showed CS-specific responses [46]. In fact, it is reasonable not to expect differences in terms of pure physical effort between one mitigation technique and another (or in its absence), so any variations in this metric may potentially be attributable to the presence of CS.

2) Subjective Metrics: Subjective evaluation relies on a questionnaire that is grounded on existing literature and organized into several sections (available for download[4]).

---

[4]Questionnaire:http://tiny.cc/u7rdzz

The first section, labeled *General Questionnaire*, aims to collect users' demographic information, as well as their experience in specific areas of interest. Questions require a numerical response in a range between 1 (meaning with "not at all") and 5 (meaning "very much").

The second section, labeled as *MSSQ-Short*, contains the reduced version of the MSSQ [47]. Its aim is to identify and predict, to the greatest extent possible, the individual susceptibility of the users to the psycho-physical stimuli they may encounter while immersed in virtual scenarios. Although initially designed for MS, the MSSQ has been shown to be an excellent predictor of the SSQ Total Score [7].

Then, several sections are dedicated to the evaluation of CS. In particular, it was decided to use the SSQ in its original version [2], since it is still the most widely used tool to date for investigations on CS. The SSQ is administered before the experience (*Pre-SSQ*) to determine whether it is possible to start it or not (based on the severity of possible symptoms), as well as after the experience (*Post-SSQ*) to evaluate the CS actually elicited by the scenario. Afterwards, a section is devoted to recording the user's heart rate before (*Pre-Hearth Rate*) and after (*Post-Hearth Rate*) the experience. The next section is used to record data regarding the Discomfort Scale (DS); hence, it is labeled *Discomfort Scale*. Similarly to many other works, this scale is administered orally every minute [12], [16], in order not to interrupt the execution of the task, and used to identify situations in which the experience needs to be terminated prematurely (if $DS = 10$). The withdrawal caused by CS symptoms can be also indicated in another section, labeled *Withdrawn for Cybersickness*. It should be noted that these two sections are redundant, as the tesbed provides an automatic logging method for the DS and the withdrawal alongside performance metrics, as explained later.

With the aim of assessing the potential impact in terms of presence of the adopted mitigation techniques [48], a specific section was included. The section is composed of the questions marked with [Presence] (along with the question on the overall judgement) within the relative section of the VRUSE usability questionnaire [49]; hence, it is labeled *Presence (VRUSE)*.

Finally, the opportunity to give open feedback is provided, through a section labeled *Comments*.

All the task performance metrics are automatically calculated by the testbed and recorded in CSV log files whose name includes the user ID and a timestamp identifying the application launch. Each row of the log file reports a measurement of all the objective metrics at a specific timestamp, plus the DS value, orally indicated by the user and recorded by the experimenter by entering it in an input field displayed on a PC screen. A new measurement is taken (and inserted as a row) at specific events (start, conclusion of a lap, withdrawal for CS, end), or at regular, one-minute intervals during the experience.

D. Protocol

This section illustrates the suggested experimental protocol for the use of the testbed, which was also applied to the presented use case.

For the SSQ, it is common to use a threshold equal to 20 out of 235 on the Total Score to indicate the beginning of sickness symptoms perception, and a threshold equal to 100 out of 235 to indicate active illness [50]. Hence, the first threshold value was used to filter out, through the *Pre-SSQ*, users entering VR with pre-existing, excessive symptoms.

1) Preparation: A proper experiment preparation is crucial to ensure the accuracy of the results and minimize unnecessary exposure to CS symptoms for the involved users. Following the best practices for this type of investigations, the experiment should be limited to healthy individuals with good balance, and executed in a room characterized by cool temperature, good airflow, and ventilation [51].

Firstly, users are asked to complete the *General Questionnaire*, which can be used to apply an initial filtering on the type of users searched for the experiment. If the user fits the characteristics of the population that the hypothetical adopter of the testbed wants to study (e.g., certain age range, specific gender, particular amount of previous experience), it is possible to proceed with filling in the *MSSQ-Short*. To ensure the experiment's accuracy, users who obtains a *MSSQ RawScore* of ≥ 30.4 (95% percentile [52]) should be excluded, as their susceptibility to CS is deemed too high [53]. This precaution is commonly adopted when conducting investigations on MS [54], but recently it was also utilized for studies on CS [55].

Users must be informed about the purposes of the experiment and the possible consequences in the event of inaccurate responses in the *General Questionnaire* or the *Pre-SSQ*, especially if they are allowed to participate in the experience despite having a high susceptibility to sickness or previous sickness symptoms not properly detected. This can be potentially done through an informed consent form, following the ethical guidelines set by the experimenter's institution.

Then, the *Pre-SSQ* is administered. Results are not recorded, but used to delay or stop the execution of the task in presence of sickness symptoms comparable to CS. In particular, as mentioned above, any score above 20 is considered significant enough to prevent (or possibly postpone) the VR experience.

2) Execution: If the user is unfamiliar with VR technology or with the VR system selected for the experiment, a brief initial explanation is recommended before entering VR. Prior to wearing the HMD, the user's heart rate should be measured and recorded using the method that the experimenter deems most suitable (hearth rate monitors, wearable fitness trackers, smartwatches with built-in hearth rate sensors), and recorded in the *Pre-Hearth Rate* section.

While the user is in VR, the experimenter conducts the testing activity from in front of the PC monitor where the application is running. At the beginning, a panel containing instructions related to the task to be performed is presented. The start of the experience can be initiated directly by the user (by pressing a button on the controller), but only after the experimenter has "armed" the system by pressing a specific key combination on the keyboard, preventing inadvertent button presses. This should be done only once the user has stated that the rules and objectives of the scenario and its tasks are clear. In the case of the RC and TR scenarios, the user is

expected to be seated on a chair, whereas in the other cases must remain standing.

As mentioned previously, the DS is orally asked at regular intervals (every minute), to avoid interfering with the execution of the task. Generally, it is strongly recommended not to keep users who reach the maximum DS value within the experience.

As soon as the experience is completed and after the removal of the HMD, it is necessary to measure the *Post-Hearth Rate* using the same method chosen previously, and calculate the difference with respect to the *Pre-Hearth Rate*.

Whether the users complete the scenario or quit early due to extreme CS symptoms, they have to fill in the *Post-SSQ* and the *Presence (VRUSE)* sections. If additional information related to the sickness status needs to be provided, the space in the *Comments* section can be used.

The choice of how many and which scenarios to use in the experiment is left to the experimenter, who must identify those relevant to the research objectives (in Section V, the method for accomplishing this goal will be shown). If the experiment involves more than one scenario, the protocol suggests having the same user perform all the scenarios rather than assigning one scenario per user. The use of this approach is facilitated by the limited number of scenarios included in the testbed, and allows for a more comprehensive assessment of CS while reducing the required number of users. If a user is unable to complete all the scenarios planned in the experiment, the experimenter should substitute him or her with another user for the remaining scenarios. Regarding the order of presentation, it is recommended to implement pseudo-randomization (e.g., Latin Square) to prevent the experience in a specific scenario from influencing the next one. If the user needs to execute more than one scenario, it is advised to space out the exposures over time (e.g., by one day) to allow for the possible effects of CS to subside. The procedure is the same for all the scenarios.

E. Pilot Study

As mentioned before, it was necessary to conduct a pilot study to fine-tune the characteristics of the implemented scenarios. For this purpose, 40 volunteers (26 males, 14 females) in the age range 18–64 (M = 29.37, SD = 10.18), with different backgrounds and levels of experience with VR technology were selected, in order to achieve a diverse representation of the target user base. In total, 60 individuals volunteered, but 20 of them achieved a *MSSQ RawScore* above 30.4 [53], and thus were excluded following the devised protocol.

Each participant had the opportunity to try one or more scenario, depending on the time they could dedicate to the study. When participants experienced more than one scenario, the expositions were spaced over time to allow for possible CS symptoms to fade way. A Samsung Odyssey HMD was used as VR kit, whereas the hearth rate was collected using a Xiaomi Mi Band 4 smartband. The MSSQ was employed to exclude participants too prone to CS, whereas the *Pre-SSQ* was used to delay the experience in case of CS symptoms at the beginning of the experiment. Finally, DS was used to interrupt the experience in the case of extreme CS symptoms. A total of 17 trials were performed for each scenario.

For each scenario, an initial configuration was tested: Regarding the TD scenario, the configuration involved two rounds of 5 minutes, separated by a 40-second break between each round. In this case, the experiment did not provide reasons to modify the duration of the rounds, since ∼88% of the participants completed the experience with varying levels of CS. However, it was observed that the excessively prolonged break contributed to reducing CS, and it was consequently shortened to 10 seconds.

For the NS scenario, 10 laps within the VR environment were initially planned, with a total of 530 coins to collect. With this configuration, it was observed that the majority of the participants who terminated the experience prematurely due to extreme CS symptoms (∼88%) had to quit before completing the third lap, whereas the minority of those who completed the entire experience did not show any symptom. Hence, it was possible to identify that the most appropriate number of laps for this scenario could be 2.

In the case of the TR scenario, the initial configuration consisted in 20 laps (19 + 1 for practice). As in the case of the NS scenario, the participants who were able to complete all the laps did not show any CS symptom, whereas those who abandoned the experience (∼76%) quit within the fifth lap, including the practice one. Hence, the number of laps for this scenario was set to 5 (4 + 1 for practice). It was also possible to calibrate the vehicle's speeds more accurately, especially the maximum speed, which was initially too low.

A similar approach was also pursued for the RC scenario, whose initial number of laps was set to 8. Like in the two previous scenarios, the participants who did not complete the experience (∼71%) quit before completing the third lap, whereas the others reached the end without showing any CS symptom. Therefore, 3 laps were chosen as the ideal number.

As said, through the pilot study it was possible to gather information which was key to adjust the testbed scenarios and associated evaluation metrics.

IV. TESTBED UTILIZATION

The testbed is suitable for a wide range of usages within the realm of CS investigation. When focusing solely on mitigation, two primary usages can be identified.

The first usage involves evaluating a new mitigation technique by integrating it into the testbed and conducting experiments according to the provided protocol. This allows comparison with a baseline or other techniques. It is assumed that the testbed has been used previously for assessing various techniques, and those results are available. An example for this usage is detailed in Section V.

Another potential usage for the proposed testbed is that of a VR developer seeking to determine which mitigation techniques to implement for a given application. Also in this case, it is assumed that the testbed has already been used to evaluate a number of techniques, and the results of such studies are available. The testbed could then be used as follows:

1) As a first step, the developer should identify which parts of the application have characteristics that partially or completely overlap with those of the four proposed

scenarios. For instance, if a substantial portion of the VR experience requires the users to navigate the VE, then the NS scenario seems to be the most pertinent. Additionally, if there are segments of the experience that involve road vehicle operation, the parallels are primarily with the TR scenario.

2) Based on this initial analysis, the developer should focus on the mitigation techniques that have been analyzed in the identified scenarios, and see which are those that showed the best results in the testbed. If this has already been done, and the results are available, the developer can employ them to select the techniques that have proven to be the most suitable.

3) Otherwise, it is necessary to conduct new experiments, following the protocol proposed in the testbed, to evaluate a sufficient number of mitigation techniques. This will enable a more informed selection of the best candidates afterward.

V. USE CASE

To demonstrate a possible usage of the testbed and verify its effectiveness, a practical scenario is presented. The use case consists in leveraging the testbed to evaluate a CS mitigation technique. As said, this activity could be performed by a researcher or developer who has designed a new technique and wants to study its effects. The technique selected for this demonstration is the one proposed by the *VR Tunnelling Pro* Unity asset [5], which consists of a dynamic vignette (by default, black) allowing relevant scene objects to be masked (i.e., excluded from the vignetting). This configuration, which from now on will be referred to as *Masked Dynamic FOV (MDF)*, represents one of the most advanced variants of the Dynamic FOV, but has not yet been experimentally investigated. The testbed was used to run a user study and collect the metrics required for the analysis.

A. Use Case: Evaluation of a New Mitigation Technique

Below, the description of the selected technique will be firstly provided. Afterwards, the introduction of the user study will follow. Finally, the results will be discussed.

1) Masked Dynamic FOV: As mentioned before, an implementation of this technique is provided as open-source within the free *VR Tunnelling Pro* Unity asset. This package is designed for VR application developers (both commercial and non-commercial) who require a highly configurable, customizable, and plug-and-play version of Dynamic FOV, which is also why it was chosen for this study. In Nie et al. [5], relevant elements were detected at runtime based on their saliency through the application of an AI algorithm, and this information was used to exclude identified elements from the Gaussian blur filter applied to the entire frame. The authors of Ang et al. [4] replicated this behavior without relying on machine learning. They achieved this by excluding areas of the frame characterized by a certain color, markedly different from the rest of the VE (e.g., pink or yellow), and by

coloring the relevant elements accordingly. In the case of *VR Tunnelling Pro*, it is possible to assign a tag to scene objects to mask them from the vignetting; this tagging ensures that vignetting automatically excludes their visual representation, without the need for specific detection algorithms. This allows the exclusion of elements that are known a priori to be relevant for the experience.

Specifically, the elements chosen as relevant (and therefore marked with masking tags) for each of the four testbed scenario are the following:

- For the TD scenario, ground and aerial targets, the tower, the platform, the handgun, and the information panels.
- In the NS scenario, only the coins and the hint arrow.
- For the TR scenario, based on Nie et al. [5], targets (pink walls), road signs, the starting line, and arrows along the circuit's boundary.
- In the RC scenario, the billboards related to the observational task and the starting sign, used to indicate the current lap.

With regard to the configuration of the *VR Tunnelling Pro* asset, the starting point was the default setting, according to which the vignetting is driven only by angular velocity, interpolating between no vignetting at 0°/s, up to 100% (85% frame coverage) at 180°/s and beyond, with a smoothing time of 0.15s. In addition to this, the asset offers the possibility to have a vignetting based on linear velocity too, but it is disabled by default. In order to achieve a behavior similar to state-of-the-art Dynamic FOV implementations based on translational movements, it was decided to activate it and set thresholds at 0m/s for no vignetting and 20m/s (and above) for maximum vignetting, while keeping frame coverage and smoothing time unchanged. This asset configuration ensured effective functioning and consistent behavior across all the four scenarios. The visual effect of the MDF mitigation technique with respect to a baseline scenario with No Technique (NT) applied can be observed in Fig. 2g and 2h.

A developer of the MDF technique could have formulated some hypotheses to guide the design and implementation steps. In particular, he or she could have hypothesized that the combination of vignetting with the masking of relevant elements should minimize information loss, possibly leading to higher task performance while maintaining effectiveness in terms of CS mitigation. To validate these hypotheses through the testbed, a between-subjects user study was conducted. The experimental activity involved 35 volunteers (23 males, 12 females), aged between 18 and 70 (M = 30.62, SD = 11.64), with medium to low prior experience with immersive VR applications. The goal was to involve individuals who represent various demographics that can be considered potential users of consumer VR but who do not frequently use this technology (to avoid possible habituation effects).

In the *General Questionnaire,* 34 out of 35 participants obtained an *MSSQ RawScore* lower than 30.4 [53], and were divided into two groups of 17 individuals, one for each of the tested conditions: the baseline (NT) and the examined MDF technique. Each participant experienced all the four scenarios, but for only one condition. The order was selected following a Latin square design, ensuring a counterbalanced

presentation across participants. The experiments were carried out using a Meta Quest 2 VR system, connected to a VR-ready workstation (Intel i7-8700K, Nvidia GTX 2080 Ti, 32GB of RAM) via Air-Link. To collect the hearth rate, a Xiaomi Mi Band 4 smartband was employed.

Like in the case of the pilot study, in presence of CS symptoms at the end of a scenario, the participant was asked, after the completion of the questionnaire, to wait for the time necessary for the symptoms to disappear; this time could vary from a few minutes to postponing the next scenario to another day, in more critical cases.

2) Results: The subjective results of the experimental activity are reported in Fig. 3 and Table I. The Shapiro-Wilk test was used to analyze data normality. Since the data were found to be normally distributed, the two-tailed unpaired t-test ($p < .050$) was used to unveil statistical differences.

An initial observation that can be made is that the examined technique (MDF) yields a significant reduction in CS symptoms when compared to the baseline condition (NT). This reduction is evident in all the SSQ subcategories, i.e. nausea, oculomotor, and disorientation, as well as in the Total Score, for three out of four scenarios (Fig.3a, 3c, and 3d). For the NS scenario (Fig. 3b), the advantage is visible only in the nausea and oculomotor subcategories, as well as in the Total Score.

Going into the details regarding the SSQ items, it is possible to see that the scenarios that are more demanding in terms of visual stimuli (i.e. the vehicle-based ones) are also those in which the advantages of the MDF technique over the NT condition are more evident. For instance, in the RC scenario, the MDF technique has a benefit on all SSQ items except eyestrain. It is interesting to note that, despite roller coasters being quite common scenarios in investigations related to CS, techniques based on Dynamic FOV have never been tested in these contexts. Hence, it can be inferred that the technique under examination is effective even in particularly extreme scenarios from the perspective of movement. For the TR scenario, the situation is similar, but the advantage is evident for a lower number of items. In this case, the only work that considered a technique based on dynamic vignetting is that of Shi et al. [13]. The authors, though, did not observe any significant difference between that technique and the baseline, deviating from the findings of previous investigations [13], which, incidentally, focused solely on navigational tasks. Consequently, the authors attributed this discrepancy to the presence of high-speed movements and the users' disorientation in perceiving abrupt changes in FOV amplitude, particularly in the case of collisions. It is therefore possible to hypothesize that the implementation of MDF (*VR Tunnelling Pro*), in contrast with the one tested in [13], is characterized by a less annoying behavior, particularly thanks to the smoothing (not employed in [13]) and probably also due to a better choice of activation thresholds. Furthermore, the constant presence of relevant elements may have contributed to lowering disorientation during collisions too.

The situation is different for the two body-centric scenarios, namely TD and NS. For example, in the TD scenario, the use of MDF appears to mitigate symptoms related to fewer SSQ items compared to what observed in the TR scenario,

and to a lesser extent. Specifically, there are no significant differences for headache, sweating, and «fullness of the head», whereas there is an advantage over the NT condition for dizziness with eyes open and eyestrain. However, overall the differences are all less prominent compared to the TR and RC scenarios, and this outcome aligns with the fact that this type of scenarios is less chaotic from the perspective of movements (real or artificial). Nevertheless, as seen before, the advantage is still significant in every SSQ subcategory, and this finding aligns with the previous literature, which shows that similar techniques can be effective in scenarios with prevalence of controller-based rotations [12]. Lastly, it is interesting to note that the NS scenario, one of the most explored in the literature, turns out to be the one in which the advantages in terms of CS brought by the investigated technique are the smallest. While benefits are present in two subcategories and in the Total Score, the analysis of individual symptoms reveals that, for the majority of them, no statistical differences were observed, except for the general discomfort and headache symptoms. This finding could explain why, in some studies with similar techniques and scenarios (e.g., [14]), it was not easy to find differences with respect to the baseline. To better understand this particular situation, however, it is necessary to also consider the results of objective metrics.

Finally, no significant differences were detected in the *Presence (VRUSE)* section in any scenario, suggesting that the technique under examination maintains a level of transparency that does not compromise this aspect of the VR experience compared to the baseline.

Moving to the objective evaluation (Table II), it can be observed that the obtained results exhibit a considerable degree of heterogeneity, especially concerning performance metrics.

In the TD scenario, the use of MDF does not seem to provide advantages or disadvantages in terms of accuracy compared to the NT, but it significantly improves the EP metric. One could hypothesize that, in the case of NT, the users tend to reduce the frequency of rotations to limit the worsening of CS symptoms, and this inevitably leads to a decrease in the chances of not making errors (i.e. when an enemy is not hit before reaching its target). Hence, it can be asserted that, not only MDF allow for a reduction in CS, but it also enables the users to enhance their task performance, which, in this case, corresponds to better protection of the objectives to be defended (the tower and the ground).

In the case of the NS scenario, where the MDF provided the lowest benefits in terms of CS mitigation among the four scenarios, two interesting significant differences are visible. The first difference concerns the OS metric, which shows that the participants who experienced the MDF were able to perform the task significantly faster than those without mitigation. The second difference regards the percentage of withdrawals for extreme CS symptoms (a unique case among the four scenarios), which is significantly lower in the case of MDF. It can therefore be hypothesized that the lower mitigation could be linked to the fact that, in the case of the baseline (NT), a majority of participants reached intolerable levels of CS much earlier compared to those with MDF, who generally continued the experience for a longer time

(a) Tower Defense (TD)

(b) Navigational Search (NS)
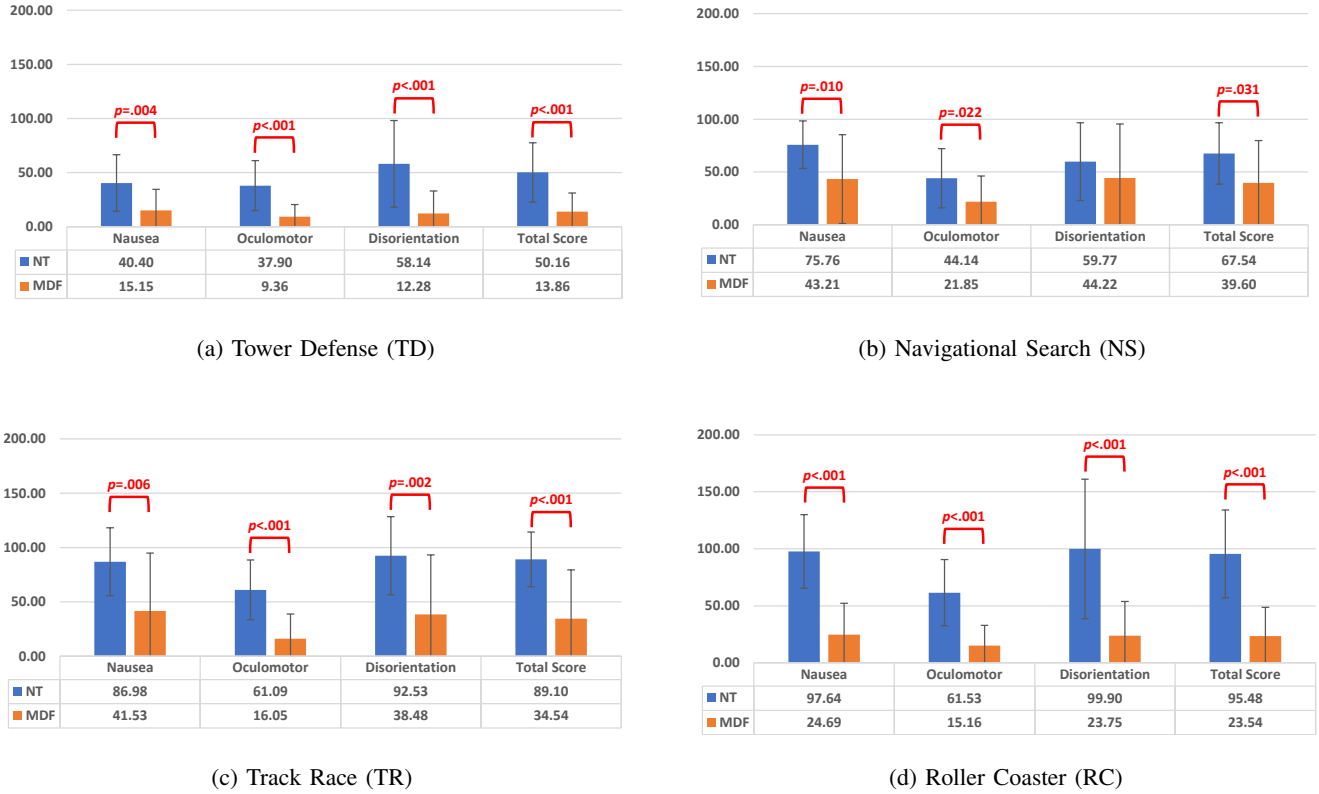

(c) Track Race (TR)

(d) Roller Coaster (RC)

Fig. 3: Subjective results based on the three SSQ [2] sub-scales and the total SSQ score (the lower the better). Baffles are used to report significant differences ($p < .050$). Standard deviation is expressed through error bars. Value ranges for each scale: Nausea [0-200], Oculomotor [0-159], Disorientation [0-292], Total Score [0-235].

TABLE I: Subjective results based on the SSQ items [2] (0: None, 1: Slight, 2: Moderate, 3: Severe). Significant differences ($p < .050$) are highlighted with a bold font, along with the best value between the two conditions.

SSQ Items	TD			NS			TR			RC		
	NT	MDF	*p*-value	NT	MDF	*p*-value	NT	MDF	*p*-value	NT	MDF	*p*-value
General discomfort	0.94	**0.29**	**.008**	1.71	**0.88**	**.029**	1.71	**0.82**	**.018**	2.06	**0.65**	**<.001**
Fatigue	0.76	**0.12**	**.005**	1.35	**0.53**	**.009**	1.47	**0.24**	**<.001**	1.29	**0.12**	**<.001**
Headache	0.35	0.29	.724	0.24	0.24	1	1.06	**0**	**<.001**	0.82	**0.24**	**.027**
Eyestrain	1.06	**0.29**	**.007**	0.88	0.35	.082	1	0.35	.060	1.06	0.47	.072
Difficulty focusing	0.71	**0.12**	**.007**	0.53	0.24	.224	1	**0.24**	**.007**	0.82	**0.18**	**.040**
Increased salivation	0.41	**0.06**	**.036**	0.65	0.18	.064	1.12	**0.41**	**.029**	1.06	**0.18**	**.004**
Sweating	1.06	0.65	.230	1.76	1.06	.068	1.71	**0.82**	**.028**	1.71	**0.47**	**<.001**
Nausea	0.82	**0.35**	**.049**	1.59	1.12	.216	1.76	**0.94**	**.49**	1.82	**0.59**	**.001**
Difficulty concentrating	0.47	0.06	.051	0.71	0.41	.305	0.94	0.35	.060	1.35	**0.29**	**.001**
«Fulness of the head»	0.53	0.18	.057	0.53	0.35	.444	1.06	**0.24**	**.005**	1.06	**0.12**	**.001**
Blurred vision	0.71	**0.06**	**.005**	0.41	0.24	.390	0.88	**0.12**	**<.001**	0.71	**0.06**	**.008**
Dizziness (eyes open)	0.59	**0.12**	**.031**	0.18	0.41	.242	0.53	0.47	.832	0.88	**0.24**	**.014**
Dizziness (eyes closed)	0.65	**0.06**	**.017**	0.59	0.41	.475	0.94	**0.35**	**.041**	0.82	**0.18**	**.032**
Vertigo	0.18	0.00	.073	0.47	0.41	.839	0.47	0.41	.765	1.06	**0.35**	**.006**
Stomach awareness	0.18	0.18	1	0.71	0.59	.693	1.18	0.76	.288	1.29	**0.06**	**.001**
Burping	0.35	0.00	.101	0.82	0.29	.084	0.71	0.24	.051	0.94	**0.06**	**.001**

until completion. However, the prolonged exposure to the VE comes with a cost, i.e. the exacerbation of CS symptoms. Consequently, any disparity in symptoms compared to the baseline may diminish or vanish by the end of the experience.

Moving to the TR scenario, although no significant differences were observed for the OS metric, the participants of the MDF group were significantly more accurate than the baseline group (AC metric). Therefore, the use of the examined

technique allowed the participants to better focus on the task of impacting the various brick walls. Although no significant differences were observed regarding the EP metric, it is notable that the metric values are quite low for both the MDF and NT conditions. This outcome is linked to the fact that the considered metric becomes more useful when the presence of a mitigation technique amplifies user errors and loss of control, such as veering off the road, compared to the baseline.

TABLE II: Objective results. Operation Speed (OS), Accuracy (AC), and Error Proneness (EP) metrics are normalized on a [0-1] scale (the higher the better), the delta hearth rate (ΔHR) values are raw mean values, whereas withdrawals for extreme CS are expressed as percentages. Significant differences ($p < .050$) are highlighted with a bold font, along with the best value between the two conditions.

Objective Metrics	TD			NS			TR			RC		
	NT	MDF	*p*-value	NT	MDF	*p*-value	NT	MDF	*p*-value	NT	MDF	*p*-value
OS [0-1]	-	-	-	0.54	**0.78**	**.047**	0.47	0.44	.809	0.76	**0.90**	**.014**
AC [0-1]	0.39	0.46	.117	0.47	0.58	.074	0.45	**0.79**	**.001**	-	-	-
EP [0-1]	0.56	**0.75**	**<.001**	-	-	-	0.90	0.93	.428	0.76	0.84	.063
ΔHR	2	0.65	.751	1.53	0.82	.880	-1	2.35	.399	0.53	0.24	.949
Withdraw for CS [%]	12%	6%	.559	65%	**29%**	**.040**	25%	24%	1	35%	18%	.257

Regarding the RC scenario, the situation is similar to the TR one for the EP metric (no differences), but there is a significant difference for the OS metric in favour of MDF. Thus, it can be observed that the use of MDF simultaneously reduces CS symptoms and achieves better task performance (i.e., keeping the cart at higher speed without increasing the number of errors) in the potentially most critical scenario of the testbed in terms of CS.

Finally, it can be observed the lack of significant differences regarding the ΔHR metric. This result is likely related to the large variability in physiological responses among the participants. To exclude age and gender as confounding factors, separate analyses were conducted for them, focusing solely on this metric. The distributions of age and gender for the four scenarios were as follows:

- TD: 21 Males, 13 Females; 31 Young Adults ($\leq$ 35 yrs), 1 Adult (> 35 and $\leq$ 60), 3 Elderly (> 60).
- NS: 22 M, 12 F; 31 YA, 1 A and 2 E.
- TR: 22 M, 12 F; 30 YA, 2 A and 2 E.
- RC: 22 M, 12 F; 28 YA, 4 A and 2 E.

Based on these distributions, a gender analysis was performed, analyzing males and females separately, first across the entire sample and then considering only the young adults age group. In both cases, no significant differences were found for this physiological parameter in any of the four scenarios, except for the young female sub-group, where a significant but negligible difference was observed in the RC scenario (NT: 0.53, DF: 0.24, $p = .048$). This outcome suggests that the most useful time to measure this type of bio-signal may not be at the end of the scenario execution, when CS symptoms may have potentially subsided after a peak during the experience. It would therefore be beneficial to also perform real-time monitoring with a specialized device, and then possibly study the correlation with CS progression (measured via DS score).

Observing these results from a developer's perspective, one could assert that the initial hypothesis was correct, and the examined mitigation technique does effectively reduce CS while also improving performance in the considered tasks, albeit to varying degrees in the four scenarios. In particular, the benefits are more evident in the more dynamic scenarios, whereas in the case of the possibly more representative scenario, i.e. NS, the advantage in terms of CS is lost in favour of a longer user dwell time in the VE.

VI. DISCUSSION AND CONCLUSIONS

In this paper, a novel testbed for studying CS in immersive VR is proposed, with a particular focus on the assessment of CS mitigation techniques. The testbed offers four different representative scenarios defined by grounding on previous works on the subject, and integrates, besides CS subjective assessment methods based on standard questionnaires, also automatic measurements aimed to evaluate the impact of CS on user's task performance. The testbed is accompanied by an experiment protocol designed to foster reproducibility, and includes a number of state-of-the-art mitigation techniques (one of which is evaluated experimentally, for demonstration purposes). The tool is designed to be potentially modified and extended, remaining available to the research community to provide a foundation for future investigations on the subject.

The main limitation of the testbed currently lies in the fact that a choice had to be made about the number of scenarios to consider. To maximize coverage of the proposed taxonomy, four scenarios have been included, one per category. This choice, which certainly facilitates the possible execution of the entire testbed by a single user, may have led to the exclusion of some less representative, but still potentially useful, techniques. To overcome this limitation, future developments could move in two directions. The first direction could be to make the presented scenarios flexible and configurable, covering as many variants as possible among those included in the taxonomy. For instance, the TD scenario could be quickly converted to a Follow target one, disabling the stationary shooter logic (weapon and enemies) and introducing an animated flying object rotating around the tower to attract the user's gaze. Similarly, the NS scenario could be easily downgraded to Free exploration (removing the coins), Follow path (replacing the coins with indications of the route to follow), Spatial updating (replacing the coins with pairs of different and less frequent objects), while with more effort, a shooter logic with enemies could be introduced. The TR scenario could be easily converted to a passive driving scenario by replaying a recorded human session. Creating an urban driving scenario, manual or autonomous, would require dedicated developments. The second direction could involve identifying and incorporating relevant yet mostly unexplored scenarios, such as those centered on aerial vehicles (be it fixed-wing or rotary). Additionally, since the scenarios are designed as outdoor settings with distant objects but some practical applications may require consideration of indoor environments or objects close to users, future work may address these

limitations by designing scenarios for such conditions.

Another limitation of the testbed is that it only includes locomotion techniques requiring minimal physical movement within a confined space and relies on hand controllers for the tasks. This constraint limits the application of natural walking and techniques such as redirected walking, which require continuous physical movement. Furthermore, the scenarios are more appropriate for visual effect-based mitigation techniques rather than for techniques involving geometric deformation (e.g., [27]), which are challenging to incorporate into the existing VEs. These issues could be addressed by designing dedicated scenarios for these conditions as well.

Additional future works could encompass the evaluation of the mitigation techniques already included in the testbed, as well as the integration of other techniques that have not been incorporated yet. It might be also worth studying the potential benefits of introducing a scoring system similar to the one proposed in [32], in order to obtain a ranking for the techniques based on specific requirements.

Finally, it would be interesting to use the testbed to explore the benefits of aspects of adaptation and habituation to CS [56]. This would involve designing a kind of training protocol to be conducted over a longer period of time, with the aim of experimentally assessing the possibility for the users to achieve the so-called *VR Legs*, i.e., the capacity to build, by undergoing controlled exposures of increasing intensity and duration to VR, a form of habituation that renders them less responsive to the effects of CS[6].

ACKNOWLEDGMENTS

REFERENCES

[1] K. M. Stanney, R. S. Kennedy, and J. M. Drexler, "Cybersickness is not simulator sickness," *Proc. of the Human Factors and Ergonomics Society Ann. Meeting*, vol. 41, no. 2, pp. 1138–1142, 1997.

[2] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness." *The Int. J. of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[3] H. Kim, D. J. Kim, W. H. Chung, K.-A. Park, J. D. K. Kim, D. Kim, K. Kim, and H. J. Jeon, "Clinical predictors of cybersickness in virtual reality (VR) among highly stressed people," *Scientific Reports*, vol. 11, no. 1, p. 12139, 2021.

[4] S. Ang and J. Quarles, "Gingervr: An open source repository of cybersickness reduction techniques for unity," in *Proc. of 2020 IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 460–463.

[5] G.-Y. Nie, H. B.-L. Duh, Y. Liu, and Y. Wang, "Analysis on mitigation of visually induced motion sickness by applying dynamical blurring on a user's retina," *IEEE Trans. on Visualization and Computer Graphics*, vol. 26, no. 8, pp. 2535–2545, 2020.

[6] C. Groth, J.-P. Tauscher, N. Heesen, S. Castillo, and M. Magnor, "Visual techniques to reduce cybersickness in virtual reality," in *Proc. of 2021 IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 486–487.

[7] N. Tian, P. Lopes, and R. Boulic, "A review of cybersickness in head-mounted displays: raising attention to individual susceptibility," *Virtual Reality*, vol. 26, no. 4, pp. 1409–1441, 2022.

[8] A. V. Milani, N. Tian, and R. Boulic, "Cybersickness assessment framework(csaf): An open source repository for standardized cybersickness experiments," in *Proc. of 2023 IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2023, pp. 477–480.

[9] S. Martirosov, M. Bureš, and T. Zítka, "Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality," *Virtual Reality*, vol. 26, no. 1, pp. 15–32, 2022.

[10] M. L. van Emmerik, S. C. de Vries, and J. E. Bos, "Internal and external fields of view affect cybersickness," *Displays*, vol. 32, no. 4, pp. 169–174, 2011, visual Image Safety.

[11] A. S. Fernandes and S. K. Feiner, "Combating VR sickness through subtle dynamic field-of-view modification," in *Proc. of 2016 IEEE Symp. on 3D User Interfaces (3DUI)*, 2016, pp. 201–210.

[12] N. Norouzi, G. Bruder, and G. Welch, "Assessing vignetting as a means to reduce VR sickness during amplified head rotations," in *Proc. of 15th ACM Symp. on Applied Perception*. ACM, 2018.

[13] R. Shi, H.-N. Liang, Y. Wu, D. Yu, and W. Xu, "Virtual reality sickness mitigation methods: A comparative study in a racing game," *Proc. ACM Comput. Graph. Interact. Tech.*, vol. 4, no. 1, 2021.

[14] H. Buhler, S. Misztal, and J. Schild, "Reducing VR sickness through peripheral visual effects," in *Proc. of 2018 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 517–519.

[15] P. Budhiraja, M. R. Miller, A. K. Modi, and D. Forsyth, "Rotation blurring: Use of artificial blurring to reduce cybersickness in virtual reality first person shooters," 2017, arXiv: 1710.02599 [cs.HC].

[16] Y.-X. Lin, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, W.-C. Lin, and S. V. Babu, "How the presence and size of static peripheral blur affects cybersickness in virtual reality," *ACM Trans. Appl. Percept.*, vol. 17, no. 4, 2020.

[17] C.-Y. Chen, C.-H. Chuang, T.-L. Tsai, H.-W. Chen, and P.-J. Wu, "Reducing cybersickness by implementing texture blur in the virtual reality content," *Virtual Reality*, vol. 26, no. 2, pp. 789–800, 2022.

[18] Z. Cao, J. Jerald, and R. Kopper, "Visually-induced motion sickness reduction via static and dynamic rest frames," in *Proc. of 2018 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 105–112.

[19] D. M. Whittinghill, B. Ziegler, J. Moore, and T. Case, "Nasum Virtualis: A Simple Technique for Reducing Simulator Sickness," Game Developers Conf. (GDC), San Francesco, 2015.

[20] C. Wienrich, C. K. Weidner, C. Schatto, D. Obremski, and J. H. Israel, "A virtual nose as a rest-frame - The impact on simulator sickness and game experience," in *Proc. of 2018 10th Int. Conf. on Virtual Worlds and Games for Serious Applications (VS-Games)*, 2018, pp. 1–8.

[21] A. Kemeny, P. George, F. Merienne, and F. Colombet, "New VR navigation techniques to reduce cybersickness," in *The Engineering Reality of Virtual Reality*. San Francisco, United States: Society for Imaging Science and Technology, 2017, pp. 48–53.

[22] T. Nguyen-Vo, B. E. Riecke, and W. Stuerzlinger, "Simulated reference frame: A cost-effective solution to improve spatial orientation in VR," in *2018 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 415–422.

[23] Y. Farmani and R. J. Teather, "Viewpoint snapping to reduce cybersickness in virtual reality," in *Proc. of Graphics Interface 2018*. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2018, pp. 168–175.

[24] R. Lou and J.-R. Chardonnet, "Reducing cybersickness by geometry deformation," in *Proc. of 2019 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 1058–1059.

[25] R. Lou, R. H. Y. So, and D. Bechmann, "Geometric Deformation for Reducing Optic Flow and Cybersickness Dose Value in VR," in *Eurographics 2022 - Posters*, B. Sauvage and J. Hasic-Telalovic, Eds. The Eurographics Association, 2022, pp. 1–2.

[26] J. Han, A. V. Moere, and A. L. Simeone, "Foldable Spaces: An overt redirection approach for natural walking in virtual reality," in *Proc. of 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 167–175.

[27] T. Nie, I. B. Adhanom, and E. S. Rosenberg, "Like a rolling stone: Effects of space deformation during linear acceleration on slope perception and cybersickness," in *Proc. of 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, 2023, pp. 658–668.

[28] R. Lou, F. Mérienne, R. H. Y. So, T.-T. Chan, and D. Bechmann, "Geometric simplification for reducing optic flow in VR," in *Proc. of 2022 IEEE International Symp. on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2022, pp. 682–685.

[29] K. Carnegie and T. Rhee, "Reducing visual discomfort with hmds using dynamic depth of field," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 34–41, 2015.

---

[6]VR legs - VR AR & XR Wiki: https://xinreality.com/wiki/VR_legs

[30] S. Jung, R. Li, R. McKee, M. C. Whitton, and R. W. Lindeman, "Floor-vibration VR: Mitigating cybersickness using whole-body tactile stimuli in highly realistic vehicle driving experiences," *IEEE Trans. on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2669–2680, 2021.

[31] S. Weech, T. Wall, and M. Barnett-Cowan, "Reduction of cybersickness during and immediately following noisy galvanic vestibular stimulation," *Experimental Brain Research*, vol. 238, no. 2, pp. 427–437, 2020.

[32] A. Cannavò, D. Calandra, F. G. Pratticò, V. Gatteschi, and F. Lamberti, "An evaluation testbed for locomotion in virtual reality," *IEEE Trans. on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 1871–1889, 2021.

[33] 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.

[34] C. Yildirim, "Cybersickness during vr gaming undermines game enjoyment: A mediation model," *Displays*, vol. 59, pp. 35–43, 2019.

[35] F. Wu and E. Suma Rosenberg, "Adaptive field-of-view restriction: Limiting optical flow to mitigate cybersickness in virtual reality," in *Proc. of the 28th ACM Symp. on Virtual Reality Software and Technology*, ser. VRST '22. ACM, 2022.

[36] T. Weißker, A. Kunert, B. Fröhlich, and A. Kulik, "Spatial updating and simulator sickness during steering and jumping in immersive virtual environments," in *Proc. of 2018 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 97–104.

[37] A. Garcia-Agundez, C. Reuter, P. Caserman, R. Konrad, and S. Göbel, "Identifying cybersickness through heart rate variability alterations," *Int. J. of Virtual Reality*, vol. 19, no. 1, pp. 1—-10, 2019.

[38] M. Bahit, S. Wibirama, H. A. Nugroho, T. Wijayanto, and M. N. Winadi, "Investigation of visual attention in day-night driving simulator during cybersickness occurrence," in *Proc. of 2016 8th Int. Conf. on Information Technology and Electrical Engineering (ICITEE)*, 2016, pp. 1–4.

[39] R. Venkatakrishnan, R. Venkatakrishnan, A. Bhargava, K. Lucaites, H. Solini, M. Volonte, A. Robb, S. V. Babu, W.-C. Lin, and Y.-X. Lin, "Comparative evaluation of the effects of motion control on cybersickness in immersive virtual environments," in *Proc. of 2020 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 672–681.

[40] C. Curry, R. Li, N. Peterson, and T. A. Stoffregen, "Cybersickness in virtual reality head-mounted displays: E xamining the influence of sex differences and vehicle control," *Int. J. of Human–Computer Interaction*, vol. 36, no. 12, pp. 1161–1167, 2020.

[41] S. Davis, K. Nesbitt, and E. Nalivaiko, "A systematic review of cybersickness," in *Proc. of the 2014 Conf. on Interactive Entertainment (IE2014)*. ACM, 2014, p. 1–9.

[42] R. Islam, Y. Lee, M. Jaloli, I. Muhammad, D. Zhu, and J. Quarles, "Automatic detection of cybersickness from physiological signal in a virtual roller coaster simulation," in *Proc. of 2020 IEEE Conf. on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 648–649.

[43] R. Hussain, M. Chessa, and F. Solari, "Mitigating cybersickness in virtual reality systems through foveated depth-of-field blur," *Sensors*, vol. 21, no. 12, 2021.

[44] S. Palmisano and R. Constable, "Reductions in sickness with repeated exposure to hmd-based virtual reality appear to be game-specific," *Virtual Reality*, vol. 26, no. 4, pp. 1373–1389, 2022.

[45] R. Islam, K. Desai, and J. Quarles, "Cybersickness prediction from integrated HMD's sensors: A multimodal deep fusion approach using eye-tracking and head-tracking data," in *Proc. of 2021 IEEE International Symp. on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 31–40.

[46] E. Chang, H. T. Kim, and B. Yoo, "Virtual reality sickness: A review of causes and measurements," *Int. J. of Human–Computer Interaction*, vol. 36, no. 17, pp. 1658–1682, 2020.

[47] J. F. Golding, "Predicting individual differences in motion sickness susceptibility by questionnaire," *Personality and Individual Differences*, vol. 41, no. 2, pp. 237–248, 2006.

[48] S. Weech, S. Kenny, and M. Barnett-Cowan, "Presence and cybersickness in virtual reality are negatively related: A review," *Frontiers in Psychology*, vol. 10, 2019.

[49] R. Kalawsky, *The Science of Virtual Reality and Virtual Environments*, 1st ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1993.

[50] C. Webb, "Simulator sickness in the MH-47G simulator," U.S. Army Aeromedical Research Laboratory, Fort Rucker AL, Tech. Rep. ADA513134, 2010, uSAARL Report No. 2010-11.

[51] A. N. Ramaseri Chandra, F. El Jamiy, and H. Reza, "A systematic survey on cybersickness in virtual environments," *Computers*, vol. 11, no. 4, pp. 1–15, 2022.

[52] A. M. Bronstein, J. F. Golding, and M. A. Gresty, "Vertigo and dizziness from environmental motion: Visual vertigo, motion sickness, and drivers' disorientation," *Semin Neurol*, vol. 33, no. 03, pp. 219–230, 2013.

[53] J. Li, A. Reda, and A. Butz, "Queasy Rider: How head movements influence motion sickness in passenger use of head-mounted displays," in *13th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '21. ACM, 2021, pp. 28—38.

[54] C. Schartmüller and A. Riener, "Sick of Scents: Investigating non-invasive olfactory motion sickness mitigation in automated driving," in *12th Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '20. ACM, 2020, pp. 30—39.

[55] I. M. Arafat, S. M. Shahnewaz Ferdous, and J. Quarles, "Cybersickness-provoking virtual reality alters brain signals of persons with multiple sclerosis," in *Proc. of 2018 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 1–120.

[56] K. Stanney, B. D. Lawson, B. Rokers, M. Dennison, C. Fidopiastis, T. Stoffregen, S. Weech, and J. M. Fulvio, "Identifying causes of and solutions for cybersickness in immersive technology: Reformulation of a research and development agenda," *Int. J. of Human–Computer Interaction*, vol. 36, no. 19, pp. 1783–1803, 2020.

Davide Calandra received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from Politecnico di Torino university, Italy, in 2014, 2017, and 2023. He is a research fellow at the Department of Control and Computer Engineering of Politecnico di Torino, with interests in interactive graphics applications, eXtended Reality and human-machine interaction. He is a student member of IEEE and IEEE Consumer Technology Society. Contact him at davide.calandra@polito.it.

Fabrizio Lamberti received the M.Sc. and the Ph.D. degrees in computer engineering from Politecnico di Torino university, Italy, in 2000 and 2005, respectively. Currently, he is a Full Professor at the Department of Control and Computer Engineering of Politecnico di Torino, where he leads the VR@POLITO lab. His interests are in the areas of computer graphics, human-machine interaction and intelligent systems. He serves as an Associate Editor for several journals, including IEEE Transactions on Computers, IEEE Transactions on Consumer Electronics, and International Journal of Human-Computer Studies. He is a Senior Member of IEEE. He is a Member of the Board of Governors of IEEE Consumer Technology Society, for which he is serving as Vice President of Technical Activities. He is the Chair or the Digital Reality Technical Community (DRTC). He is the Vice-Chair of IEEE Computer Society Italy Chapter. Contact him at fabrizio.lamberti@polito.it.