POLITECNICO DI TORINO
Repository ISTITUZIONALE

An evaluation testbed for locomotion in virtual reality

Original
An evaluation testbed for locomotion in virtual reality / Cannavò, A., Calandra, D., Prattico', F.G., Gatteschi, V., Lamberti, F.. - In: IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS. - ISSN 1077-2626. - STAMPA. - 27:3(2021), pp. 1871-1889. [10.1109/TVCG.2020.3032440]

Availability:
This version is available at: 11583/2848984 since: 2021-01-29T08:17:48Z

Publisher:
IEEE

Published
DOI:10.1109/TVCG.2020.3032440

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

29 November 2026

An Evaluation Testbed for Locomotion in Virtual Reality

Alberto Cannavò, *Student Member, IEEE*, Davide Calandra, F. Gabriele Pratticò, *Student Member, IEEE*, Valentina Gatteschi and Fabrizio Lamberti, *Senior Member, IEEE*

Abstract—A common operation performed in Virtual Reality (VR) environments is locomotion. Although real walking can represent a natural and intuitive way to manage displacements in such environments, its use is generally limited by the size of the area tracked by the VR system (typically, the size of a room) or requires expensive technologies to cover particularly extended settings. A number of approaches have been proposed to enable effective explorations in VR, each characterized by different hardware requirements and costs, and capable to provide different levels of usability and performance. However, the lack of a well-defined methodology for assessing and comparing available approaches makes it difficult to identify, among the various alternatives, the best solutions for selected application domains. To deal with this issue, this paper introduces a novel evaluation testbed which, by building on the outcomes of many separate works reported in the literature, aims to support a comprehensive analysis of the considered design space. An experimental protocol for collecting objective and subjective measures is proposed, together with a scoring system able to rank locomotion approaches based on a weighted set of requirements. Testbed usage is illustrated in a use case requesting to select the technique to adopt in a given application scenario.

Index Terms—Virtual Reality, virtual environments, locomotion, performance, user experience, requirements, evaluation, testbed.

1 Introduction

The continuous grow of VR (Virtual Reality) technology and its applications is posing developers a number of challenges, concerning how to provide users with virtual experiences capable to mimic as much as possible the real ones. When virtual environments are large enough, one of the tasks to be supported is locomotion, which allows users to freely move in the 3D world and explore it [1]. When scenarios are not purely exploratory, locomotion is often performed together with other tasks that may require the users to interact with objects populating the virtual world, e.g., to grab and manipulate them. In these cases, locomotion can be regarded as a *secondary* task, and aspects such as intuitiveness and ease of use become essential for the outcome of the *primary* task [2].

Previous works showed that the most natural locomotion experience can be achieved by providing the users with the possibility to physically walk in the real world while being immersed in VR [3], [4], [5]. Unfortunately, the use of physical, or real, walking is often constrained by the amount of required space and/or the limited area that can be tracked by consumer-level VR products based on *outside-in tracking* technology. Recently, an alternative technology known as *inside-out tracking* has been introduced to tackle the latter limitation, as well as to lower the prices and increase the portability of VR systems. However, devices based on this technology are usually characterized by a reduced tracking accuracy and by a limited region in front of the headset in which the hand controllers can be tracked. Indeed, systems specifically designed to support high accuracy, wide-area tracking also exist, but they are generally very expensive and require large amounts of obstacle-free space [5].

Because of the above scenario, there is a growing body of literature that recognizes the importance of designing *stationary* locomotion techniques able to make the users explore virtual worlds regardless of the amount of space available in the real world [6]. Many alternatives have been proposed, each characterized by different hardware requirements, costs and levels of effectiveness in deceiving the human sensory system, i.e., in making it perceive the virtual movements as realistic. Notwithstanding, finding those which can perform better in a given application domain (or set of domains) is not an easy task. This situation is confirmed by the large number of studies that presented custom evaluation approaches (both in terms of experiments to perform, as well as of data to collect and analyze), often motivated by the need to compare the performance of a newly-proposed technique with related ones [7], [8], [9], [10], etc. As a result, the literature does not offer researchers and practitioners a standard methodology for investigating the problem, but rather provides many fragmented partial solutions, often limited to a particular and non-generalizable use case.

For these reasons, the purpose of the present work is to propose a comprehensive testbed [11] designed to compare VR locomotion techniques. In particular, the contribution of this paper can be split in three parts. A *methodology* supporting the experimental investigation of various techniques from many different perspectives is first proposed. The methodology, grounded on available literature, supports the collection of both objective and subjective measures (metrics) concerning users' performance and experience in the execution of a number of locomotion-related tasks arranged in a set of representative scenarios. This methodology is

• *The authors are with the GRAINS – GRAphics And INtelligent Systems group at the Dipartimento di Automatica e Informatica of Politecnico di Torino, 10129 Torino, Italy. e-mail: (see http://grains.polito.it/people.php).*

Manuscript received XXX XX, XXXX; revised XXX XX, XXXX.

accompanied by a *scoring system*, which is designed to combine collected experimental data with a weighted set of requirements and provide a ranking of selected locomotion techniques based on intended usage conditions. Finally, the paper includes a *use case*, i.e., an example on how the testbed could be exploited to compare several techniques based on the characteristics of a possible VR application they would be used into. The testbed is released as open source at *https://github.com/VRatPolito/LET-VR* with the aim to let other researchers extend it by adding new scenarios/tasks, requirements/metrics and experimental data.

2 Related Work

Locomotion in immersive environments is a well-studied research area, as confirmed by the large number of works in the literature which present surveys and systematic reviews on available techniques [2], [12], [13], [14], [15]. This section firstly introduces locomotion techniques that have been proposed so far. Afterwards, it discusses works that presented methodologies and studies aimed to evaluate these techniques, by considering both the case in which locomotion represents the primary task to perform in the virtual environment, as well as the case in which it is combined with other tasks.

2.1 Locomotion Techniques

As said, real walking represents the most natural technique, since it allows the user to control the position of the avatar in the virtual environment by physically moving in the tracked area. An alternative that is commonly adopted also in desktop or console 3D applications and videogames relies on gamepads and joysticks.

Since these devices may cause disorientation in immersive virtual environments [16], locomotion techniques specifically tailored to VR started to be experimented by the research community (and to be implemented by the industry, in some cases) [2]. These techniques are generally designed to leverage users' spatial orientation abilities and minimize the distance between movements they perform with their body and movements produced in the virtual environment. The taxonomy defined in [17] roughly splits locomotion techniques into "magical" and "mundane". Magical techniques make the users move in a way that is not feasible in the real word, e.g., through so-called *teleportation* [18], using the *world-in-miniature* metaphor [19], or via *hand-based manipulations* of the virtual world [20]. In contrast, mundane techniques rely on real-world metaphors, and can be either vehicle- or body-centric. The former techniques involve the use of virtual vehicles [21] [22], [23], whereas the latter ones are meant to recreate physical walking, running, swimming, etc., by making the user perform the same or similar movements [24]. In particular, body-centric techniques rely on approaches based on *repositioning*, *proxy gestures*, and *redirected walking* [2].

Repositioning systems are solutions in which the users' forward movement due to the physical execution of the walking gesture is counterbalanced by a device that constrains them in a fixed position. Repositioning could rely on active components, like motorized platforms and floor tiles [25], [26], human-size hamster balls [27], etc., or exploit passive elements like low-friction omnidirectional treadmills or slippery shoes-based interfaces able to reduce the friction generated by the users' steps [28], [29], [30]. Compared to active systems, passive systems represent a less expensive and simpler solution. The Cyberith's Virtualizer [31] and KatVR's Kat Walk are examples of commercial systems belonging to the latter category. Solutions based on proxy gestures allow the users to navigate virtual environments through movements performed either with the upper or the lower part of their body. An example of proxy gesture-based solutions are the *leaning* interfaces. They rely on side and forward leaning of the torso for controlling the locomotion, do not require additional hardware, and proved to be characterized by a high usability [32], [33], [34], [35], [36]. Another example is *walking-in-place* (WIP), which has been shown to be slightly more expensive that the above interfaces (since it requires some additional hardware) [37], but also proved to be capable of providing some of the proprioceptive stimuli of real walking. A quite comparable user experience, which is generally perceived also as less fatiguing and comes at no extra cost, is obtained by exploiting upper-body gestures in so-called *arm swinging* (AS). In this case, locomotion is obtained through the rhythmic swinging of the arms, similarly to what occurs in a real walk or in cross country skiing [7], [38]. Lastly, redirected walking includes techniques that influence the user's path through the physical environment by manipulating stimuli offered through/by the virtual environment itself, e.g., changing the user's virtual point of view by applying continuous, hopefully imperceptible, transformations to the mapping between real and virtual movement [39].

2.2 Assessment of Locomotion as Primary Task

Although it is only thanks to rather recent advancements in technology that VR has become largely accessible to various user categories, works that analyze the performance of different locomotion techniques are not new. A first example is represented by the framework proposed more than 20 years ago in [8] to perform a comparison between alternative techniques. The main contributions were the proposal of a taxonomy to categorize available methods and the identification of the most relevant quality factors to evaluate their usability (speed, accuracy, spatial awareness, ease of learning, ease of use, information gathering, and presence) together with relevant scenarios for testing them. Because of the technology available at that time, the framework was limited by the fact that, besides head tracking, interaction could only rely on a spatial input device (like a 3D mouse), and the focus was basically on direction and speed control in absolute and relative displacements.

More recently, a user study was performed [7] with the aim to evaluate the perceived naturalness offered by four different locomotion methods, i.e., keyboard, AS, WIP, and *hip movement* (HM), a technique that lets a user move in a virtual environment by swinging its hip left and right. Experiments also investigated the user's sense of presence and the amount of Unintended Positional Drift (UPD), that is, the physical movement in the forward direction occurring while performing the walking gesture. Another

aspect whose evaluation was proposed in this work is the similarity in terms of physical strain between real walking and the gestures that the considered locomotion techniques use as a proxy for it. Similarly, in [40], AS and a diverse implementation of the WIP method based on a consumer electronics bracelet placed on the user's ankles were compared with real walking in terms of spatial orientation and ability to estimate travelled distances.

Although body-centric techniques were included in the evaluation, the above studies still did not consider the other major approach to locomotion in VR mentioned in the previous section, i.e., repositioning systems. Works like [41] and [42] aimed to tackle this lack, by comparing a passive repositioning system (namely, a slippery shoes-based interface), AS and WIP in terms of usability and motion sickness. Evaluation was performed by asking users to carry out a complex task in a realistic immersive VR scenario simulating an emergency procedure, and by collecting objective and subjective measurements. Unfortunately, since experiments included a single task requesting the users to carry out different actions, authors were not able to isolate the actual contribution of the specific locomotion technique to the success of the task and of individual operations.

A comparable result was achieved by the authors of [43]. In this case, several variations of a single locomotion approach were analyzed. Specifically, the goal was to compare in terms of naturalness, sense of presence and UPD three WIP-based methods, leveraging respectively marching, wiping and tapping gestures. The task used in the evaluation requested the users to walk along a predefined path for a certain amount of time. The simplicity of the task did not allow the authors to find any difference in usability and performance, making them conclude that other tasks had to be designed, e.g., to investigate object avoidance capabilities, analyze the promptness in starting and stopping movements, measure movements accuracy, etc.

An example of accuracy task was presented, e.g., in [9], where the users were requested to be as fast as possible in getting close to different targets positioned on walls, without hitting them. The locomotion techniques used in the experiments encompassed real walking, WIP and a joystick-based method, and combined three possible visual conditions: head-mounted display in a computer-generated environment, unrestricted, as well as field of view (FOV)-restricted natural vision in a corresponding real environment. The combination of the above elements led to the generation of five experimental configurations (some setups were deemed as not relevant). To study the impact of a given technique on performance, the authors used the final distance to target as well as a time-to-collision metric, defined as the target-user distance divided by the user's velocity.

Given the variety of environments and tasks used in these studies, several works focused on the design of systems capable to support the creation of customized VR scenarios where locomotion techniques can be tested into. For instance, in [44], the authors identified a number of key attributes (like environment size, path length, path complexity, presence of obstacles, etc.) that should be controlled in a study on locomotion in VR. Then, they presented a tool that supports the manipulation of these parameters to generate different virtual environments which can stress the techniques of interest in terms of fatigue, motion sickness and usability over short-, medium- and long-distance travels. Although the tool is characterized by a high flexibility in the creation process, it does not come with ways to analyze the performance of the selected techniques, and has not been used yet in comparative studies.

Finally, there are also works that tried to define methods for ranking techniques based on their ability to address the fundamental requirements of locomotion in VR. For example, in [45], a ranking tool considering three dimensions, namely, motion sickness, presence, and fatigue was presented. The proposed approach, referred to as a User-Centric Classification (UCC), can be exploited to plot in a 3D space the performance of a given technique, with each axis representing one of the above requirements. Thanks to the adopted visualization, it is quite easy to compare the techniques of interest along the selected dimensions; unfortunately, a considerable number of aspects which, based on previous works, shall be taken into account, are not considered in the devised evaluation scale.

2.3 Assessment of Locomotion as Secondary Task

In the works reviewed so far, the focus was mainly on locomotion itself. However, as said, in many applications like, for instance, videogames, locomotion could be just a secondary task, which might have to be performed together with other (primary) tasks such as searching for an object in the virtual environment, grabbing it, etc.

Thus, in [38], a user study was conducted to evaluate the effectiveness of AS compared to WIP for a grabbing task combined with locomotion. Participants were asked to travel different paths while carrying a virtual object, which had to be released into a basket at the end of the path. Techniques were evaluated in terms of immersion, motion sickness, and physical demand. The mental workload dimension, which was not considered in previously cited works, was additionally investigated by using the NASA-TLX questionnaire [46]. The work in [47] presented a framework to assess naturalness and effectiveness of four techniques including WIP, AS, *Tap* (a metaphorical gesture that allows the users to move by tapping with the index finger in the direction they want to walk) and *Push* (a gesture consisting in closing and opening the hand while dragging it, similar to moving a lever) by considering interaction during locomotion. In particular, during the tests, users were requested to navigate along predefined paths avoiding obstacles and interacting with virtual objects by relocating them.

A further example of user studies investigating the impact of locomotion on interaction with objects is presented in [48]. Authors focused on the effect that different values of translational gain, i.e., the mapping between the physical and the virtual movement, can have on tasks that require picking and placing of virtual objects. Results showed that, although increasing translational gain could represent a seamless way for speeding up locomotion since it does not require any additional hardware, it may negatively affect accuracy, motion sickness, and mental workload for values greater than a given threshold. More recently, another study [49] investigated the relationships between interaction and

locomotion in virtual environments in terms of presence, cybersickness, and usability. A classification of interactions in VR was proposed, based on the following categories: *simple* (interactions performed by manipulating only the three positional degrees of freedom, DOFs) vs *complex* (involving six DOF manipulations), and *sequential* (only one hand involved) vs *parallel* (both hands used simultaneously).

Another frequent task is object search. For instance, in [50], an experiment was designed to compare three bimanual locomotion techniques for desktop virtual walkthroughs. The considered techniques relied on a joystick with different numbers of DOFs which had to be manipulated for controlling an avatar's position in the virtual environment with the dominant hand, combined with a mouse to control the gaze with the other hand. The experiment consisted in a primed search task, where users already knew the position of the target and had to reach it in the shortest time possible. At the end of the experiment, the users were requested to judge ease of use, fatigue, accuracy and speed of the three techniques. In [51], the authors defined a similar task for comparing a mundane technique (AS) with a magical technique (teleporting), whose performance was evaluated in terms of efficacy, effectiveness, motion sickness, user experience and cognitive load.

Works above mostly focused on performance of the studied locomotion techniques and on perceived user experience, but there are also works that analyzed the impact of techniques on psychological and/or cognitive aspects. For instance, in [10], objective and subjective metrics were defined to analyze the influence of a given technique in virtual experiences designed for treating phobias (specifically, the fear of heights), whereas in [52], two user studies were performed to observe the effects of four techniques on the users' ability to gather and remember information from/on the virtual world.

2.4 Considerations

From the above analysis it is rather clear that a number of methods for evaluating and/or comparing locomotion methods were developed, each aimed to stress the features of the particular techniques considered. Experiments designed in the various works generally request the users to perform different tasks, each characterized by diverse levels of complexity and evaluated with a varying set of objective and subjective metrics. This heterogeneity is quite evident also from Table 1, which summarizes the most relevant works discussed in this section from the many considered perspectives (in columns). Only studies on psychological/cognitive aspects are neglected, as not in the scope of this work.

By moving from the above observations and by leveraging findings in the literature, the present work proposes a comprehensive evaluation testbed that can be used to study locomotion techniques from different viewpoints and rank them based on users' preference and performance in the execution of a variety of tasks (commonly requested in typical VR applications and videogames). As it can be seen in Table 1 (last row), the proposed testbed covers all the dimensions addressed in previous works.

3 Evaluation Methodology

The design of the evaluation started with the analysis of experimental studies in the existing literature, which led to the identification of a set of important and recurrent tasks involving locomotion in VR. This set was then enriched with additional tasks, less considered in the literature, but very frequent in common applications. These tasks correspond to functionalities to be possibly supported by a given VR locomotion technique: hence, in the following they will be referred to as *Functional Requirements* (FRs). Tasks were clustered in five scenarios, designed to group tasks/functionalities sharing similar features.

The execution of the tasks in a given scenario is evaluated through a set of metrics, which can be either objective (when based on measurements automatically collected by the testbed application), or subjective (when based on user-provided answers). The metrics, either derived from dimensions explored by previous works or designed ad-hoc, refer to general characteristics that the technique should offer to the users of a given VR application: thus, they will be later referred to as *Non-Functional Requirements* (NFRs).

In the following, FRs and NFRs will be illustrated in detail, by referring to works they have been derived from, when appropriate. Instructions for performing the experiments will be also provided.

3.1 Scenarios and Tasks

The devised scenarios separately address five major aspects that an effective locomotion technique should support: *straight movements*, *direction control*, *decoupled movements*, *agility* and *interaction with objects*. Several screenshots showing operations (requirements) to be performed (supported) in the above scenarios are provided in Fig. 1. Some videos showing tasks execution with different locomotion techniques are available at *http://tiny.cc/8uxlsz*.

3.1.1 S1. Straight movements

The first scenario is designed to assess a given technique under the most simple locomotion conditions, i.e., with movements that do not require directional changes. Hence, it focuses on different tasks one may have to perform while moving on a straight path, from taking a simple stroll, to stopping at exact points, reaching a given target and sprinting. These tasks are quite common in VR simulations where the user impersonates a character and an appropriate level of physicality is required (e.g., [41] and [42]).

In the first task named *T1. Straight line walking* and illustrated in Fig. 1a, the user has to reach a destination in front of it by following the green path on the floor as much as possible (like in [53]). Afterwards, in task *T2. Over/Undershooting*, it has to stop within three green circular areas which are characterized by a decreasing radius; during the virtual experience, the new area appears only when the previous one has been reached (Fig. 1b). This task is a simplified version of the task proposed in [9]. Once the last destination has been reached, the user is asked to follow a moving object, namely a robot, while always remaining within a green area behind it (Fig. 1c). Since the robot periodically changes its speed, this is referred to as task *T3. Chasing*. Finally, the user is requested to cover a long path

TABLE 1
Review of the most relevant related works according to the following dimensions (columns): work presents an evaluation testbed (possibly referred to as framework), i.e., a set of scenarios/tasks including also the evaluation metrics (1); collects objective (2) or subjective measures (3); considers other kinds of interactions during locomotion (4); tackles aspects concerning accuracy, input sensitivity, responsiveness, and/or level of control (5), operation speed (6), error-proneness (7), usability (8), presence and immersion (9), motion sickness (10), physical effort, V/R physical strain similarity, comfort, self-motion compellingness, and/or acclimatisation (11), mental effort/workload, and/or cognitive demand (12), ease of use, intuitiveness, and/or naturalness (13), appropriateness and/or effectiveness (14), learnability (15), enjoyability, satisfaction and/or frustration (16).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
[8] Bowman 1997	✓	✓			✓	✓			✓					✓		
[7] Nilsson 2013	✓	✓	✓						✓		✓		✓			
[40] Wilson 2016		✓			✓		✓		✓							
[41] Calandra 2018		✓	✓	✓	✓	✓	✓	✓	✓	✓						
[42] Calandra 2019		✓	✓	✓	✓	✓	✓	✓	✓	✓						
[43] Nilsson 2013		✓	✓						✓		✓		✓			
[9] Whitton 2005	✓	✓	✓					✓						✓		
[44] Sarapuri 2018										✓	✓					
[45] Albert 2018	✓		✓						✓	✓	✓					
[38] Pai 2017		✓	✓	✓					✓	✓	✓	✓				
[47] Ferracani 2016	✓	✓	✓	✓	✓	✓	✓						✓			
[48] Wilson 2018	✓	✓	✓	✓	✓	✓				✓	✓	✓				
[49] Mayor 2019	✓	✓	✓			✓		✓	✓	✓						
[50] Lapointe 2009	✓	✓	✓		✓	✓					✓		✓			✓
[51] Loup 2018	✓	✓	✓		✓		✓	✓		✓		✓				
[10] Schuemie 2005	✓	✓	✓		✓		✓		✓	✓						
[52] Suma 2010	✓	✓	✓						✓	✓						
This work	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

(shown in Fig. 1d) as quickly as possible (*T4. Sprinting* task). Both following targets at a fixed distance and running are very common operations in VR videogames [54].

3.1.2 S2. Direction control

The second scenario aims to evaluate locomotion techniques in the execution of tasks that involve different ways of performing direction changes, which may be requested by the actual goals of the task being executed, or by constraints set by the simulated environment. The devised tasks require the user to change direction multiple times while proceeding on straight lines, moving backward, following a curved path, walking uphill under different conditions and in presence of a dangerous obstacle (a chasm). The scenario includes aspects that may characterize, for instance, simulations of walkable natural environments, but also of urban settings, in which it could be necessary to navigate multi-story buildings featuring obstacles and stairs.

In the first task (*T1. Multi-straight line walking*), the user has to follow a path made up of a sequence of straight lines (a polyline). This task was inspired by [53], where a user study on locomotion techniques was carried out by simulating a visit to a museum. The user is requested to move between artworks and stand for a while in front of them (until lights switch off) before moving to the next one (as done in [55], though in a different environment). The virtual environment and the complete path to be followed are depicted in Fig. 1e. As said, the emphasis here is on direction control, rather than on speed: hence, a new section of the path appears only when the operation on the current artwork has been completed. In the second task (*T2. Backward walking*), the user is requested to leave the museum room by looking at the last artwork while walking backward to reach the door behind it (Fig. 1f). This is one of the tasks proposed in [47] (the other tasks in that work are taken into account in different scenarios of the proposed testbed). Afterwards, the user has to traverse a tunnel (Fig. 1g, with path shown) that requires it to make continuous direction adjustments while moving (*T3. Curved walking*), like in [43]. The fourth task (*T4. Stairs & ramps*) focuses on direction control from a different perspective, as it first requests the user to climb up a stair and a ramp, then analyzes its preferences by making it choose between a stair or a ramp for a final, long climb (as shown in Fig. 1h). This task has been included to evaluate the disturbance introduced by movements applied to the user's point of view: stairs generate a sussultatory pattern, ramps a more uniform motion, and their effects on the user's experience (e.g., on motion sickness) may vary depending on the locomotion technique being used. Lastly, the degree of confidence in controlling direction with the given technique is investigated in a so-called *T5. Fear* task, in which the user needs to cross a hazardous area on a high building roof, where it risks to fall down in case of wrong movements (Fig. 1i). User's interactions with stairs and hazardous areas were originally investigated in [10].

3.1.3 S3. Decoupled movements

This scenario aims to investigate to what extent a given locomotion technique lets the user decouple the control of walking (and its direction) from gaze and hand movements; this ability is particularly important in scenarios which are not just exploratory, but request the user to perform also gaze or hand interactions. In these cases, allowing the user to move freely while avoiding interference from or onto the interaction technique could prevent a possible deterioration of the user experience.

In the first task, shown in Fig. 1j, the user needs to walk forward on a straight path while keeping the gaze fixed on an object (a blue sphere) that moves besides it (*T1. Decoupled gaze*). The ability to look elsewhere while walking is used in many VR experiences [56], e.g., to gather information on the surrounding environment while moving, keep the gaze fixed on a specific spot while waiting for a given event, avoid obstacles coming from aside, etc. In the second task (*T2. Stretched-out hands*), there are two robots to the left and right sides of the user. The user can control

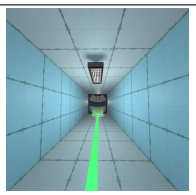
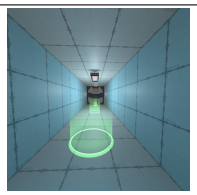
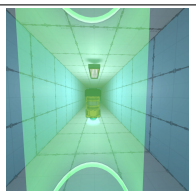
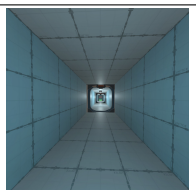
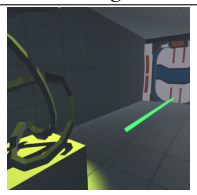
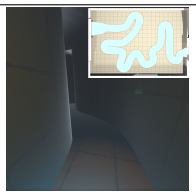
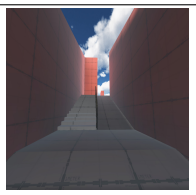
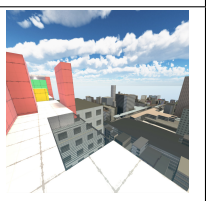
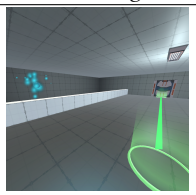
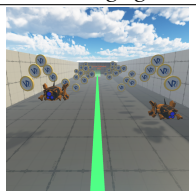
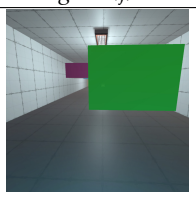
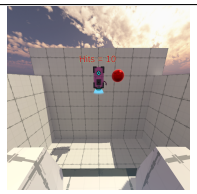
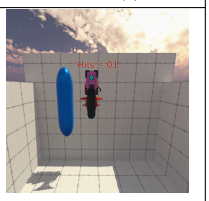
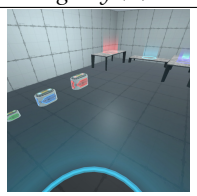
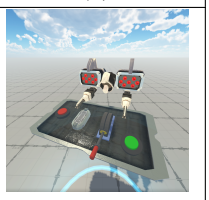
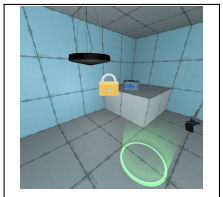
Scenario	Task				
S1. Straight movements	 *T1. Straight line walking (a)*	 *T2. Over/Under-shooting (b)*	 *T3. Chasing (c)*	 *T4. Sprinting (d)*	
S2. Direction control	 *T1. Multi-straight line walking (e)*	 *T2. Backward walking (f)*	 *T3. Curved walking (g)*	 *T4. Stairs & ramps (h)*	 *T5. Fear (i)*
S3. Decoupled movements	 *T1. Decoupled gaze (j)*	 *T2. Stretched-out hands (k)*	 *T3. Decoupled hands (l)*		
S4. Agility	 *T1. Dynamic agility (m)*	 *T2. Stationary agility (n)*	 *T3. Evasion (o)*		
S5. Interaction with objects	 *T1. Grabbing (p)*	 *T2. Manipulation (q)*	 *T3. Interaction in motion (r)*		 *Training (s)*

Fig. 1. Testbed application: a)–r) devised scenarios and tasks (functional requirements of a locomotion technique) considered in the evaluation, and s) training scenario (depicted for sake of completeness). Green circles/cylinders represent destinations to be reached in order to complete the task or parts of it, whereas green lines indicate to the user the path to follow (when needed).

the robots' relative position by moving the hands close/far from its body. The goal here is to make the robots collect a number of floating coins while walking on a straight path (Fig. 1k). The coins are lined up so that the user has to keep its arms completely stretched out to collect them. This task is designed to stress locomotion techniques that rely on hand controllers, which could possibly restrict some arm movements. The last task (*T3. Decoupled hands*) is similar to the previous one, but in this case coins are placed at different heights and distances from the straight path (Fig. 1l). Thus, the user needs to move its hands in the proper way to control the attached robots (which in this case can be moved also vertically) and collect the coins while following the path. This task is meant to further investigate the performance of locomotion techniques that could interfere with arm movements. The ability to effectively maintain a high level of control over the intended motion (preventing undesired/accidental movements) while the hands are busy in other interactions is key in many VR applications and videogames, e.g., for shooting [57] or collecting objects [58].

3.1.4 S4. Agility

The fourth scenario focuses on evaluating the level of virtual and physical agility enabled by the locomotion technique being studied; in particular, aspects pertaining agility while walking freely or being constrained in some of the move-

ments are evaluated. These abilities turn out to be quite important in simulations of chaotic events, as well as in applications requiring an avoidance reaction to sudden threats.

In the first task (*T1. Dynamic agility*), the user has to reach the opposite side of a room by avoiding collisions with blocks of different shapes that are moving towards it by controlling its direction and crouching when needed (Fig. 1m) [59]. In the second task (*T2. Stationary agility*), the user is facing a robot that periodically shoots balls at it with different speeds and directions, following a pseudo-random pattern (Fig. 1n). The user is requested to avoid the balls by only moving the head and hip: in this way, it is possible to evaluate the encumbrance caused by the considered technique to movements that are performed on the spot, i.e., which do not actually involve the use of locomotion, as requested in applications such as, e.g., [60]. The last task, (*T3. Evasion*), is a variation of the previous one, where the user has to dodge a series of larger, human-sized capsule-shaped bullets: in this case, it needs to use the locomotion technique to rapidly move left or right, as bullets can no more be avoided just with on-the-spot movements (Fig. 1o).

3.1.5 S5. Interaction with objects

The last scenario is designed to evaluate the impact of the locomotion technique on precise interaction with objects available in the virtual environment, which is a fundamental feature in applications simulating procedures that involve the use of virtual tools [42]. In particular, the scenario investigates pick-transport-place interactions with objects under various conditions (one or more objects, presence or absence of obstacles, occluded sight), manipulation and positioning of objects requiring precise movements and continuous displacements, and impact of complex interactions performed during locomotion.

In the first task, named *T1. Grabbing*, the user is asked to grab objects (batteries) and place them into several target positions (sockets, to open doors) based on their colors, sizes and shapes, like in [38], [47]. At the beginning of the task, there is only one object to move at a time. Afterwards, two objects need to be grabbed at the same time, one per hand (the size of one of the objects is intentionally larger than the other one, thus occluding the user's sight). In the second part of the task, complexity is further increased by asking the user to carry objects while traversing a maze and avoiding collisions with walls, like in [52] (Fig. 1p). In the second task (*T2. Manipulation*), fine-grained control of locomotion and manipulation is requested to the user in order to reposition objects scattered in the environment at given locations and then assemble a tower by precisely stacking them (Fig. 1q). This kind of interaction is used, for example, in [61]. The last task, *T3. Interaction in motion*, is the most demanding one, as it is designed to stress the interference of the locomotion technique on the way the user performs compulsory interactions with objects while moving. The user is placed in front of a moving element (a robot holding a panel) on which it will have to perform four interactions in a specific order (Fig. 1r). Once the user gets close to the robot, the latter starts to move at a (constant) speed close to the user's walking speed, and continuously adjusts its direction to keep moving away from the user. The user will need to modulate appropriately its speed in order to be able to perform the requested interactions on the robot: if it stays too far, then it will not be allowed to interact; if it gets too close or tries to overtake the robot, the latter will wriggle out of it with a rapid direction change. Regarding interactions, the user will have to press a button, insert a battery in a socket, push a lever, and press another button. The order was chosen so that similar interactions are not placed close to each other.

3.1.6 Training

As it will be discussed in the following, the above scenarios were designed to be modular, so that one could possibly pick just some of them in a given study, depending on the specific aspects to be investigated. Hence, the users need to be trained on the functioning of the particular locomotion technique being investigated before taking the real test. To this purpose, a dedicated scenario was developed and populated with challenges useful to learn how to master the considered technique with the assistance of an external human guidance (see also Section 3.3.2). The training scenario is depicted in Fig. 1s. Several videos showing the execution of the training are also provided at *http://tiny.cc/8uxlsz*.

It is worth noticing that, in order to generalize the testbed scenarios to the characteristics of the user and to the features of the selected locomotion technique, some of the tasks (e.g., *S3.T3*, *S5.T3*, etc.) were designed to adapt to the particular configuration being tested by leveraging a calibration file. The file includes information about the maximum height of the user's head and the maximum distance between its hands, thus allowing the testbed to account both for users stature and for possible constraints set by the locomotion technique (e.g., seated configuration, interface harness, etc.). Information can be collected either manually or automatically (using a procedure included in the training scenario); in the latter case, before leaving the training, the user is invited to reach the calibration spot and keep a particular calibration pose (straighten up as much as possible, with arms laterally outstretched) for a few seconds.

3.2 Metrics

The proposed testbed has been designed to analyze NFRs for a VR locomotion technique (reported in Table 2) on given tasks in objective or subjective terms.

3.2.1 Objective Metrics

Some of the NFRs are analyzed by defining, for each task, several objective metrics. Some of these metrics are taken from previous works; where necessary, new metrics have been defined ad hoc for individual tasks, based on the particular aspects to be investigated. All these "per-task" metrics are reported in Table 3, together with the references to works they have been derived from (when available).

Objective metrics are grouped in one of the following categories, corresponding to three major NFRs possibly relevant for a locomotion technique: the *Operation speed* at which tasks are performed, the *Accuracy* obtained in carrying them out, and the *Error-proneness*, respectively marked as *OS*, *AC* and *EP* in Table 3. The only exception is the *StairsChoice* metric (marked with *OT*, for *Other*, in Table 3); this metric

TABLE 2
Non-functional requirements considered in the testbed: those evaluated via objective metrics are marked with $^+$ if per-tasks, and with $^\#$ if per-scenario; those assessed subjectively are marked with * if assessed per-scenario, with $^\circ$ if assessed on the overall experience.

Accuracy$^+$, abbr. AC	Subjective Units of Discomfort (AQ)*
Operation speed$^+$, abbr. OS	Self-motion compellingness$^\circ$
Error-proneness$^+$, abbr. EP	Acclimatisation$^\circ$
Physical effort$^\#$	Control$^\circ$
Input sensitivity*	Presence$^\circ$
Input responsiveness*	Learnability$^\circ$
Ease of use*	Intuitiveness$^\circ$
Perceived errors*	Comfort$^\circ$
Appropriateness*	Enjoyability$^\circ$
Satisfaction*	Overall system usability$^\circ$
Mental effort*	Mot. sickness: Nausea (SSQ)$^\circ$
Perceived physical effort*	Mot. sickness: Oculomotor (SSQ)$^\circ$
Naturalness*	Mot. sickness: Disorientation (SSQ)$^\circ$
V/R Phys. str. similarity*	Mot. sickness: Total (SSQ)$^\circ$

actually represents a subjective measure which is indirectly obtained through an objective measure, and will contribute in a different way to the evaluation.

Some of the metrics are defined as *elementary*, as they directly represent the corresponding requirement, e.g., for tasks in which the user has a single goal. As a matter of example, in *S1.T1*, spatio-temporal path deviation *STPathDev* is an immediate accuracy measure. This measure was inspired by the metric used in [9], which was slightly modified to better fit the purpose of the evaluation. The metric was originally intended for tasks not considering the possibility for the user to stop or to walk backwards; however, for some of the tasks in the testbed (e.g., *S1.T1*) the act of staying still far from the target path is considered as an error. Hence, *STPathDev* for user *i* was defined as:

$$STPathDev_i = \int_0^{ComplTime_i} PathDev_i(t)dt \qquad (1)$$

where $PathDev_i(t)$ is the deviation from the target path as a function of time. Basically, it corresponds to the area between $PathDev_i(t)$ and the time axis.

Other metrics are named *cumulative*, as the requirement for the task is defined by multiple independent measures like, e.g., in task *S5.T1*, where Error-proneness is given by *NumItemFalls*, *NumBodyColl* and *NumItemColl*.

Finally, metrics are referred to as *compound* when the user's goal is complex and the requirement for the given task needs to be expressed by combining multiple dependent measures. For example, if the user is asked to walk straight while maintaining the gaze on a specific target, like in task *S3.T1*, then *STPathDev* could be used to measure the Accuracy related to the first goal, whereas the percentage of time the user looked at the target could be used for second goal, but the combined metric should be directly proportional to both of them.

The compound metrics that have been defined are reported in the following. In particular, for the Accuracy requirement of tasks *S2.T2*, *S3.T1*, *S3.T2*, and *S3.T3*, a normalized *STPathDev* for user *i* is first defined as:

$$NrSTPathDev_i = \frac{STPathDev_i}{MaxDist \cdot ComplTime_i} \qquad (2)$$

where *MaxDist* is the maximum distance available for the user to the left or to the right of the path. Then, this measure is combined with the percentage of time the user was actually looking at the target behind it while performing the task (given by $LookAtRate_i$):

$$AccuracyBkw_i = LookAtRate_i(1 - NrSTPathDev_i) \qquad (3)$$

Similarly, for *S3.T1*, $NrSTPathDev_i$ is combined with the percentage of time the user was effectively uncoupling the gaze direction from the walk direction (given by $GazeUncRate_i$):

$$AccuracyGazeUnc_i = GazeUncRate_i(1 - NrSTPathDev_i) \qquad (4)$$

For *S3.T2*, the normalized metric is combined with the percentage of time the user kept its hands stretched-out while walking ($StrcRate_i$):

$$AccuracyStrc_i = StrcRate_i(1 - NrSTPathDev_i) \qquad (5)$$

For *S3.T3*, the *NrSTPathDev* is multiplied by the percentage of coins collected by the user ($ScoreRate_i$):

$$AccuracyHandsUnc_i = ScoreRate_i(1 - NrSTPathDev_i) \qquad (6)$$

Finally, in order to describe in quantitative terms a further NFR representing the physical effort associated with the use of a given locomotion technique, the variation in user's heart rate before and after each scenario (to be measured using, e.g., an optical sensor) is accounted by a *Physical effort* metric. Although heart rate variability could be a challenging and noisy measure, it already proved to be a rather accurate estimation of the energy consumption associated with a given physical exercise [38], [62], [63], [64].

3.2.2 Subjective Metrics

The remaining NFRs are analyzed in a subjective way. Subjective evaluation relies on a questionnaire split in three sections. The first section, referred to as *pre-test*, is meant to be delivered before running the experiments. The second section is meant to be delivered after the completion of each scenario (*after-scenario*), with slight differences from one scenario to another. Questions in this section define so-called "per-scenario" metrics. Lastly, the third section is intended as a *post-test* questionnaire, and includes questions contributing to the definition of so-called "overall" metrics. Like for objective metrics, subjective metrics can be *elementary* (when described by a single question), or *cumulative* (when analyzed through multiple questions). Questions (sometimes in the form of statements the user has to express its agreement/disagreement with) are to be scored on a Likert Scale from 1 to 5; for motion sickness symptoms, the scale ranges from 0 (none) to 3 (severe). The complete questionnaire is available at *http://tiny.cc/dzxlsz*. In the following, the three sections are described in detail.

Pre-test section: the section starts with questions aimed to evaluate previous experience with technologies related to the experiments. Afterwards, the section incorporates all the questions from the Simulator Sickness Questionnaire (SSQ) tool [65], which is used to rate the severity of motion sickness symptoms before the experiment (if any).

After-scenario section: at the completion of each scenario, the user is invited to evaluate the Input sensitivity, Input

TABLE 3
Metrics defined to evaluate in objective terms the user's experience with the given locomotion technique in the particular task, classified with respect to the corresponding NFR they contribute to, i.e., *OS* for Operation speed, *AC* for Accuracy, *EP* for Error-proneness and *OT* for Other (for cumulative metrics, numbers are used after the letter identifying the requirement to list contributing elements).

Scenario	Task	Per-task metric
S1. Straight movements	*T1. Straight line walking*	*OS: ComplTime (s)*: task completion time. *AC: STPathDev (m·s)*: spatio-temporal path deviation, calculated as the time integral of the distance from the line to follow as a function of time. *EP: NumWallColl*: num. of collisions with the walls of the corridor.
	T2. Over/Under-shooting (separate values for the three target destinations)	*OS: ComplTime (s)*: see above. *AC: TargetDist (m)*: final distance between the user and the center of the target in which the user stops [9]. *EP: NumExits*: num. of times the user exits the target destination.
	T3. Chasing	*AC1: InsideTargetRate (%)*: percentage of time the user remained inside the reference area while following the moving robot. *AC2: AvgDist (m)*: average distance between the user and the center of the reference area while following the moving robot. *EP: NumInterr*: num. of walking interruptions.
	T4. Sprinting	*OS: ComplTime (s)*: see above. *EP: NumWallColl*: see above.
S2. Direction control	*T1. Multi-straight line walking* (values averaged on the six target destinations)	*OS: ComplTime (s)*: see above. *AC1: InitAngErr (deg)*: initial angular error, i.e., difference in angles between the target and the walking direction after one meter [55]. *AC2: EstPathLen (m)*: distance traveled to reach the target [55]. *AC3: RecallTime (s)*: time spent at determining the new direction and start walking after the lights were turned off [55].
	T2. Backward walking	*OS: ComplTime (s)*: see above. *AC: AccuracyBkw (%)*: path deviation correlated with the percentage of time the user looked at the target in front of it while walking backwards. *EP: NumLookOut*: num. of times the gaze was turned away from target.
	T3. Curved walking	*OS: ComplTime (s)*: see above. *EP: NumInterr*: see above.
	T4. Stairs & ramps	*OT: StairsChoice (0/1)*: user's choice between ramp (0) and stairs (1).
	T5. Fear	*OS: ComplTime (s)*: see above. *AC: Avoidance (m·s)*: inspired by [10], but calculated like *STPathDev*, in this case using the deviation between the walking path and the edge of the large drop. *EP: NumFalls*: num. of falls from the roof.
S3. Decoupled movements	*T1. Decoupled gaze*	*OS: ComplTime (s)*: see above. *AC: AccuracyGazeUnc (%)*: path deviation correlated with the percentage of time the user looked at target on its side while walking. *EP: NumInterr*: see above.
	T2. Stretched-out hands	*OS: ComplTime (s)*: see above. *AC: AccuracyStrc (%)*: path deviation correlated with the percentage of time the user kept the arms stretched while walking. *EP: NumInterr*: see above.
	T3. Decoupled hands	*OS: ComplTime (s)*: see above. *AC: AccuracyHandsUnc (%)*: path deviation correlated with the percentage of coins collected by the user. *EP: NumInterr*: see above.
S4. Agility	*T1. Dynamic agility*	*OS: ComplTime (s)*: see above. *EP: NumObsColl*: num. of collisions with the moving blocks [47].
	T2. Stationary agility	*EP: NumHits*: num. of times the user was hit by a bullet.
	T3. Evasion	*EP: NumHits*: see above.
S5. Interaction with objects	*T1. Grabbing* (separate values for the three parts)	*OS: ComplTime (s)*: see above. *EP1: NumItemFalls*: num. of times grabbed objects fell from user's hands. *EP2: NumBodyColl*: num. of collisions of the user with walls while traveling the maze *EP3: NumItemColl*: num. of collisions of the items grabbed with walls while traveling the maze [47].
	T2. Manipulation	*OS: ComplTime (s)*: see above. *AC1: AvgSetupAcc (%)*: average positional and rotational accuracy of items placed on the table during the setup phase. *AC2: AvgTowerAcc (%)*: average positional and rotational accuracy of items placed on the table during the assembly phase.
	T3. Interaction in Motion	*OS: ComplTime (s)*: see above. *AC: CloseToTargetRage (%)*: percentage of time the user remained close to the moving robot. *EP: NumErrors*: num. of times the user performed an interaction in the wrong order or the grabbed object fell from its hands.

responsiveness, Ease of use, Perceived (Perception of) errors, Appropriateness and Satisfaction of/with the given technique based on statements adapted from the questionnaire proposed in [66]. The user is also asked to judge the level of Mental and Perceived physical effort associated with the use of the particular technique by rating dimensions defined by the ISO 9241-400 standard. Furthermore, the Naturalness of the walking gesture and the Similarity of the real physical strain with the virtual gesture it is serving as a proxy for are assessed based on questions defined in [43]. For task *S2.T5* (*Fear*) the user was also asked to rate the level of disturbance on the *Subjective Units of Discomfort* (SUD) scale (Acrophobia Questionnaire, AQ), as proposed in [10].

Post-test section: this section includes questions that refer to the experience as a whole. In particular, the user is requested to provide its evaluation of the considered technique in terms of Self-motion compellingness (i.e., the perceived physical movement during the experience) and Acclimatisation (i.e., how quickly the user forgot that it was not really walking), as defined in [43]. Moreover, the user needs to judge the level/sense of Control, Presence, Learnability, Intuitiveness, Comfort, Enjoyability and Overall system usability provided by the given technique according to [66]. Lastly, severity of motion sickness symptoms is collected again using the SSQ to study in detail changes from the beginning to the end of the experience.

3.3 Experimental Protocol

This section illustrates how to customize the testbed and how to perform the testing in order to foster reproducibility.

3.3.1 Preparation

As said, the five scenarios were designed to cluster tasks with similar characteristics, in order to maintain the consistency of the per-scenario metrics and to let the testbed users skip the testing of unnecessary aspects. That is, the study participants may be asked to take, e.g., only the first four scenarios, dropping the last one if interaction with objects is not of interest. Reordering the scenarios is possible, though not recommended, as they have been designed to challenge the participants with tasks of increasing complexity.

Differently than scenarios, tasks cannot be skipped, as such changes would influence the evaluation. In fact, metrics have been devised to assess user's performance and experience on the scenario as a whole, with dimensions that in some cases are investigated through more than one task. Moreover, aspects like Motion sickness, Acclimatisation or Physical effort, to name a few, can only be evaluated once the user has spent a sufficient amount of time in the scenario. This is also the reason why it was decided to move the delivery of the questionnaire at the end of each scenario, rather than at the end of each individual task.

Each participant is expected to test the selected scenarios with just one locomotion technique. This is due to the fact that the testbed has been developed around a between-subjects design, in order to enable the reuse of data from previous experiments when new techniques have to be included in a study (as long as the conditions are reasonably comparable to the previous ones, e.g., in terms of demographics information). Hence, questionnaire does not contain any question asking the participants to directly compare two (or more) techniques.

In principle, a within-subjects approach could be used too, but strategies for dealing with learning effects would be hard to design. It is also worth considering that the duration of the complete test experience is about 80 minutes (all scenarios and questionnaire sections included). For a within-subjects study, time requested would explode as the number of techniques tested grows, not to mention that participants could be mentally and physically challenged by requested repetitions. In conclusion, a within-subjects design could probably be feasible for a small number of techniques, possibly limiting the number of tested scenarios.

3.3.2 Execution

Participants must be instructed not to eat or drink anything two hours prior to the experiment and not to consume illicit drugs or caffeine for 12 hours prior to the experiment [67]. Only participants without illnesses or visible altered state of consciousness shall be accepted as testers.

Every participant has to be briefly informed about the purpose of the experience, the structure of the experimental protocol, as well as the operation of the VR hardware and the locomotion technique used for the activity. After that, the participant fills in the pre-test section of the questionnaire and the administrator shall decide whether to proceed with the experiments or discard the participant, considering the match with the target audience set for the study and the SSQ scores.

The participant is then provided with the VR equipment, and let in the training environment. As said, this scenario is meant to let the participant familiarize with the selected locomotion technique, and to deliver all the information that may be necessary to carry out the various tasks (if necessary, further details on the operation of the locomotion technique could be provided also prior to wearing the VR equipment, especially if safety considerations are involved). In particular, participants must be introduced to the concept of "destination", i.e., the visual indicators they will have to reach in order to start, perform or conclude many of the tasks, as well as to the so-called "blocked" state, that is a temporary situation (signaled by a padlock, as in Fig. 1s) in which the locomotion technique is disabled, usually to let the administrator explain the following task before proceeding; this state is activated automatically at the end of each task, but the administrator can also block the participant manually in order to give additional instructions.

During the training, the participant shall be asked to perform a set of actions that will be later requested in the execution of the tasks (i.e., walking straight, changing direction, adjusting the speed, walking backwards, etc.), as well as to experiment with hand interactions (by picking up some objects and transporting them to other locations). Operations above need to be repeated until the participant is feeling confident in using the locomotion technique and in control of the various interactions. Finally, calibration data for the participant have to be set, either automatically or manually. Information to be used for the guidance are provided with the testbed in the "administrator script".

After the training, the participant is ready to start the testing procedure, which can be summarized as follows:

1) participant's heart rate is collected just before entering the scenario, with the VR headset on;
2) scenario is started, participant is let in the first task;
3) locomotion is blocked, so that task can be explained by the administrator;
4) locomotion is unblocked, and the participant performs the task;
5) at the end of the task, the system blocks the locomotion again;
6) points 3)–5) are repeated for all the tasks;
7) participant's heart rate is collected again at the end of the scenario;
8) headset is removed, and participant is asked to fill in the after-scenario section of the questionnaire, supervised by the administrator;
9) points 1)–8) are repeated for the scenarios of interest;
10) participant fills in the post-test section of the questionnaire, supervised by the administrator.

In case of interruptions due to internal (e.g., extreme motion sickness) or external reasons (blackout, force majeure, etc.), the activity has to be suspended. It is at the discretion of the administrator to decide whether to let the participant resume the experiment or not, and how to treat the possibly incomplete data, e.g., discarding them, filling missing values with the mean or worst scores, etc.

4 SCORING SYSTEM

As a result of the application of the methodology illustrated in the previous section, raw measurements corresponding to objective and subjective metrics are collected in a so-called *Raw Database* (RDB) for a number of participants with all the locomotion techniques being studied. In order to use these data for comparing the various techniques, a scoring system based on the weighted sum model (WSM) is proposed as part of the testbed. WSM is a multi-criteria decision analysis (MCDA) method which was designed to evaluate a set of alternatives (techniques, in this case) based on a number of decision criteria (requirements) which are assigned a weight indicating the relative importance.

4.1 Metrics Normalization

The contributions of the various metrics to the overall score need to be expressed in the same unit. Hence, a normalization step is required. In the devised system, normalization relies on testing the statistical significance of differences between the mean of the metric for a given technique and for all the other techniques being studied.

A significance threshold equal to 5% is used (the system does not mandate, however, the use of a particular statistic test or threshold). If the test gives a p-value less than or equal to the threshold, one point is assigned to the "best" technique. The proposed normalization approach does not consider the magnitude of the differences. However, the advantage is that it does not require the definition of lower and/or upper bounds for the metrics. As a matter of example, an arbitrary bound on the *ComplTime* of a given task would not allow the testbed to account for future locomotion techniques providing largely different speed performance than those currently available. Similar considerations apply to *NumFalls* or *NumInterr*, which could theoretically grow unconstrained for poorly controllable techniques.

The system requires the testbed user to define, for each metric, a so-called *direction*. Direction can be either positive or negative, and is used to determine whether a technique should be considered as better than another one when the mean value for the metric is greater or smaller than that of the other technique, respectively.

As a matter of example, assume that four techniques T1, T2, T3, and T4 are tested and that, for a given metric m, significant differences are found between T1 and T2, T1 and T3, as well as T3 and T4. The best technique in pairwise comparisons will get 1 point. The selection of the best technique depends on the direction assigned to the metric, which, as said, could be either positive (so that the technique with the higher mean will be selected as the best, and will get the points), or negative (vice versa). Thus, if direction of metric m is positive and the means for T1, T2, T3 and T4 are, e.g., 1.0, 2.0, 2.5, and 1.5, respectively, the score S_m (in points) assigned by the system for that metric will be equal to 0 for T1, 1 for T2, 2 for T3, and 0 for T4.

For a cumulative metric , the score is computed as:

$$S_m = \frac{1}{\hat{N}_e} \sum_{e=1}^{N_e} S_e \tag{7}$$

where S_e is the number of points of the e-th element, N_e is the number of elements and $\hat{N}_e$ is the number of elements which showed at least one statistically significant difference.

4.2 Weights Selection

The scoring system requests the testbed user to assign a weight (in the range [0,1]) to a set of predefined dimensions. Weights are then combined with statistically processed data to compute an overall score for the selected techniques, and rank them based on their suitability to the specific application domain they will be used into.

Dimensions to be weighted include both FRs and NFRs identified in the previous section. As said, FRs correspond to tasks and scenarios illustrated in Fig. 1, whereas NFRs are aligned with the objective and subjective metrics collected through the testbed application and are reported in Table 2. In a basic WSM implementation, weights would be directly combined with requirements. In this case, a two-level combination is used, due to existence of two types of requirements.

For what concerns FRs, weights can be assigned with both a coarse (per-scenario) or a fine (per-task) granularity. If weights are defined per-scenario, the weight of each scenario is applied to all its tasks; if weights are assigned per-task, per-scenario weights are determined automatically by averaging per-task ones. NFR weights, on the contrary, are assigned directly to the corresponding requirements; the only exception pertains Motion sickness, for which weights can be either assigned to major components identified in the SSQ (namely Nausea, Oculomotor symptoms and Disorientation), or to the requirement as a whole (Total).

4.3 Scores Computation

For the computation of the overall score for a given technique, every NFR-related score is first multiplied by the corresponding weight, obtaining a so-called *weighted NFR score*. Then, the per-task weighted NFR scores (related to the first three objective metrics, i.e., Accuracy, Operation speed and Error-proneness), are multiplied by the fine FR weights, whereas the per-scenario weighted NFR scores (related to the questions in the after-scenario questionnaire and to the last objective metric, the *Physical effort*) are multiplied by the coarse FR weights. The weighted NFR scores related to the overall metrics (based on the questions in the post-test questionnaire) will not be influenced by any of the FR weights, as they refer to the experience as a whole. The overall score is obtained by summing up the weighted scores of the various metrics.

The weighted score associated with the objective metrics of task T_i in scenario S_j is calculated as:

$$S_{S_j.T_i} = w_{S_j.T_i} \cdot \\ \cdot (w_{OS} \cdot S_{OS_{S_j.T_i}} + w_{AC} \cdot S_{AC_{S_j.T_i}} + w_{EP} \cdot S_{EP_{S_j.T_i}}) \tag{8}$$

where $w_{S_j.T_i}$ is the weight assigned to the task, w_{OS}, w_{AC} and w_{EP} are the weights assigned to the Operation speed, Accuracy and Error-proneness objective metrics, and $S_{OS_{S_j.T_i}}$, $S_{AC_{S_j.T_i}}$ and $S_{EP_{S_j.T_i}}$ are the scores computed for the three metrics on the given task.

For task *S2.T4*, in order to handle the peculiarity of the *StairsChoice* metric, its contribution is calculated as:

$$\hat{S}_{S_2.T_4} = w_{S_2.T_4}(w_{ST} \cdot S_{ST} + w_{RA} \cdot S_{RA}) \quad (9)$$

where $w_{S_2.T_4}$ is the weight of task *S2.T4*, w_{ST} and w_{RA} are special mutually exclusive weights that can just be 0-0, 0-1 or 1-0, whereas S_{ST} and S_{RA} are the scores computed from the metric, but with opposite directions (stairs or ramps).

Similarly, the score of task *S2.T5* has to be modified to consider the contribution of the *SUD*, which is the only subjective metric related to a single task; thus, the additional score for the task is calculated as:

$$\hat{S}_{S_2.T_5} = w_{S_2.T_5}(w_{SUD} \cdot S_{SUD}) \quad (10)$$

which is basically the contribution of the weighted score for the additional metric combined with the task weight $w_{S_2.T_5}$.

The subjective, per-scenario part of the score for scenario S_j will be calculated as:

$$S_{s_j} = w_{S_j} \sum_{m=1}^{N_s} w_{s_m} \cdot S_{s_{m_j}} \quad (11)$$

where w_{S_j} is the weight of the scenario, N_s is the number of subjective metrics defined for the scenario, w_{s_m} and $S_{s_{m_j}}$ are respectively the weight of a subjective metric *m* and its score for the given scenario.

The overall score for scenario S_j will be computed by combining the (per-scenario) subjective component in (11) with the (per-task) objective component in (8) as:

$$S_{S_j} = S_{s_j} + \sum_{i=1}^{N_t} S_{S_j.T_i} + w_{PE} \cdot S_{PE_j} \quad (12)$$

where N_t is the number of tasks in the scenario; the formulation also considers w_{PE} and S_{PE_j}, which are the weight and the score for the *Physical Effort* objective metric.

For the overall metrics, the score is calculated as:

$$S_o = \sum_{i=1}^{N_o} w_{o_m} \cdot S_{o_m} \quad (13)$$

where N_o is the number of overall metrics, w_{o_i} is the weight of the m-th overall metric and S_{o_m} is its score.

Based on this formulation, the overall score for the given locomotion technique is computed as:

$$S = S_o + \sum_{j=1}^{N_S} S_{S_j} + \hat{S}_{S_2.T_4} + \hat{S}_{S_2.T_5} \quad (14)$$

where N_S is the number of scenarios included in the test.

4.4 Testbed Usage

The results obtained through the application of the scoring system to data in the RDB are recorded in the so-called *Weighted Database* (WDB). The WDB basically contains the values computed for each metric in a given set of experiments, their mean values and significances, as well as the overall scores obtained by applying weights.

The WDB is characterized by a *fixed* part, which is strictly related to the testing activity performed and encompasses:

- the set of techniques included in the experiments;
- the set of scenarios included in the experiments;
- possibly applied demographic constraints.

Then, it contains a *variable* part that needs to be configured by the testbed user in order to compute the overall scores. This part includes:

- the weights assigned to the FR/NFR metrics;
- the direction assigned to each metric;
- the subset of techniques to be actually compared.

The variable part can be adjusted by the testbed user depending on the specific study to be performed, without the need to carry out any further experimental activity.

For the purpose of identifying the locomotion technique that represents the best match for a particular scenario given a set of possible alternatives, the user can either consider an existing WDB (if the fixed configuration is appropriate for the study at hand), or generate a new one (by performing the experiments and applying the statistical analysis). Then, it should adjust the weights for the various requirements depending on the specific application scenario. For instance, should the goal be to select the best technique for a VR application in which the users need to run and complete the assigned tasks in the shortest time possible, then FRs like, e.g., *Sprinting*, and NFRs like, e.g., *Operation speed*, would probably have to be assigned a high weight. The overall scores computed by the system would then provide the ranking of techniques for the specific application.

If some of the techniques in the WDB are not of interest or are not compatible with the given scenario (e.g., the application requires more buttons than those that are available on the controllers when a particular technique is used), then they should be discarded in the RDB, and a new WDB could be generated by repeating the statistical analysis and recomputing weighted scores.

Besides overall scores, the user could also analyze scores computed for individual metrics, with the aim to identify aspects of a given technique that are particularly effective and could be exploited, e.g., in the design of a new technique.

5 Use Case

After having discussed the characteristics of the testbed, a use case is presented to provide potential users with an example of the workflow which is expected to be followed to compare a set of locomotion techniques. Various techniques coming from both industrial and academic research were preliminarily analyzed, by focusing on those capable to address the limited space constraints. Four techniques were ultimately selected for the comparison, chosen among those for which implementation details or ready-to-use packages were available (especially when more variants existed).

In the following, selected techniques and their implementations will be first described. Then, the user study performed to collect experimental data will be introduced, and the use of the relative scoring system will be explained; in particular, it will be shown how to generate the WDB and use it to compare the selected techniques under the particular usage conditions set by a specific VR application.

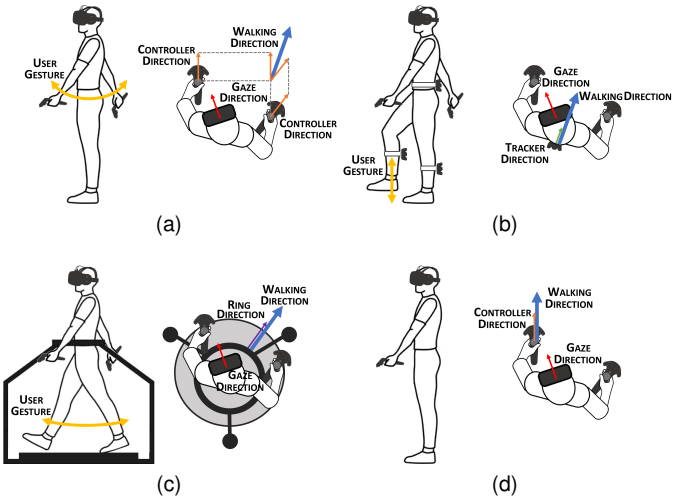


Fig. 2. Locomotion techniques considered in the use case: a) arm swinging (AS), b) walking-in-place (WIP), c) Cyberith's Virtualizer (CV), and d) joystick (JS).

5.1 Considered Techniques

As said, four techniques were selected for the comparison: AS, WIP, Cyberith's Virtualizer (CV), and joystick (JS). The VR system used in the experiments was the HTC Vive.

5.1.1 Arm Swinging

According to the experiments performed in [51], AS seems to leverage the most natural gesture among the approaches not involving the use of feet. To generate movement, the user must hold a button on the hand controllers (the grip, in the considered implementation), then swing the arms back and forth for walking/running. For this technique (as well as for the others), the trigger buttons on the hand controllers are used to interact with objects. In this work, the publicly available AS method in [68] is used, in which the direction of movement is determined by averaging the orientation of the two controllers (Fig. 2a), whereas the stride length is directly mapped on the arm swinging. This choice allows the user to decouple the walk direction from the gaze orientation.

5.1.2 Walking-in-Place

The second technique included in the study is WIP, which is another common locomotion method proposed in various versions and implementations. To generate movement, the user needs to perform a particular gesture with its legs while remaining in place. The direction of movement can be obtained from the head (by tying it to the gaze direction) or from other devices (when the gesture is recognized, e.g., using wearable sensors). In the use case, the LLCM-WIP variant proposed in [37] was chosen; according to this approach, the movement is generated through a direct mapping between the space covered by two sensors attached to the user's leg and the speed of the user in the virtual environment, whereas the direction is obtained through a third sensor placed on the user's back. In the exploited implementation, three Vive Trackers were used, two attached on the user's calves through custom 3D-printed supports, one tied to the back through a belt (Fig. 2b). With this solution it was possible to "align" this technique to the other ones (by acting on the configuration parameters), and to decouple the head/gaze orientation from the direction of movement. The gesture used is the marching one, which is considered as a kind of standard for this technique [43]. A filter on horizontal leg movements was added, in order to mitigate unwanted motion when the user turns around.

5.1.3 Cyberith's Virtualizer

The third technique that was considered in the comparison is the CV, whose functioning is illustrated in Fig. 2c. This is one of the few commercial solutions for locomotion in VR that falls under the category of passive repositioning systems (thanks to the use of a low-friction walking surface and of a rotating containment ring, which prevents the user from displacing in the physical space). In this case, the walking direction depends on the orientation of the ring. This slippery shoes-based interface, which was originally presented in [31], has been used already in previous research works and compared with other techniques, though only in broad terms [2], [41].

5.1.4 Joystick

For the JS, the implementation developed in [69] was used. In this implementation, movement is activated by pressing the pad button of any of the hand controllers and modulating the speed by moving up and down the thumb over the touch pad: the upper bound of the pad generates the maximum speed, which decreases linearly by moving down the finger towards the lower bound (zero speed). The direction is given by the controller whose touch pad is pressed to activate the motion. In order to transition from walking to running, the user needs to press the grip button (on the same controller) while performing the previous actions; this way, speed will be set to a fixed value, higher than the one reachable through the previously described modulation.

5.1.5 Configuration

During the experiments, the user is requested to stay within a predefined working area (since the testbed supports the study of a specific locomotion technique at a time, room-scale movements are disabled, although they could be studied as a separate technique). The working area is signaled in the virtual environment through a semi-transparent cyan cylinder with a 65cm radius around the initial user's position, which at the user's feet level is displayed as a brighter circle.

When, due to unintentional movements, the user enters a warning region that ranges from 90% to 100% of the working area radius, the controllers generate a haptic feedback in order to help it to remain at the center of the working area. When the user exits the working area, movement in the virtual environment is disabled, so that the only action allowed to the user is stepping back into it. The reason for this choice was to allow the user to crouch and perform on-the-spot head and hip movements while avoiding motion sickness and disorientation that would be introduced, e.g., by simply limiting the headset movements to three DOFs. In order to standardize the experience across the four techniques, maximum speed was set to 7m/s [70]. For JS, the maximum value was used for running (when the

grip button is pressed), whereas for walking the speed can be modulated between 0 and 3.5 m/s (slightly higher than common values for real walking, in the 1.5–2.6m/s range [71]). For the other techniques, the user can modulate the speed between zero and the maximum value.

Regarding interaction with objects, the implementation adopted in [41] and [42] was exploited, which proved to be characterized by a high usability regardless of the particular locomotion technique used in the experiments. In particular, the user needs to touch the object with the tip of the hand controller and, once the contact is confirmed by a visual (outline) and tactile (vibration) feedback, it can push and hold the grab button (the trigger in this case, visually signaled by a blinking outline) to bind the object to its hand and move it. The object can be dropped by releasing the trigger button, or passed to the other hand by grabbing it with the corresponding controller.

5.2 Collecting Experimental Data

In order to produce the RDB, an experimental activity was carried out by involving 48 volunteers from the student and academic population at the authors' university (37 males and 11 females, aged between 19 and 37). Each participant was assigned to one of the four locomotion techniques.

According to information collected through the pre-test section of the questionnaire, participants had a very limited experience with VR and related concepts, since 12.5% of them were regular users of the technology, whereas 31.25% asserted to play 3D videogames often or very often. Only 8.33% of the participants had used sometimes the assigned locomotion technique before the experiment, 20.83% had a little knowledge of it, and 70.83% had never used it. Regarding motion sickness, none of the participants reported particularly high symptoms before the experiments. Like in [72], a statistical analysis was performed to identify possible differences among the groups with respect to the pre-test conditions (precisely, Kruskal-Wallis and Dunn's post-hoc tests were used, due to the observed data distribution); no statistically significant difference was found.

Participants who had never experienced a VR system were given time to familiarize with this technology (headset, controllers, etc.). They could then initiate the training. Afterwards, they were asked to complete the experimental protocol. All the scenarios were tested, in the default order. All the participants were able to complete the experiment.

5.3 Computing Normalized Scores

Data in the RDB were processed using the devised scoring system. Possible outliers (identified by using the Z-score) were removed first. Afterwards, normality distribution of the data was checked by applying the Shapiro-Wilk test on each metric.

Since participants experimented all the scenarios using only one interface, it could be assumed that collected data are independent. Hence, in order to compute the p-values required by the scoring system, the following approach was pursued. For data presenting non-normal distributions, statistical significance of the differences was studied using the Kruskal-Wallis test, followed by the Dunn's post-hoc test for pairwise comparisons. For normally distributed data, statistical significance was tested using ANOVA, followed by one-to-one comparisons with the Tukey's HSD range test.

To implement the procedure illustrated in Section 4 and generate the WDB for the four techniques, an Excel spreadsheet was developed: pasting the experimental data and choosing the weights, the proper statistical tests can be applied to compute the scores (for all the metrics and the selected techniques). The spreadsheet including both raw data and weighted scores is available at *http://tiny.cc/hzxlsz*.

Fig. 3 plots the contribution of individual metrics for each technique with all weights set to 1 (default). Different colors are used to indicate the task/scenario in which points were obtained, i.e., where differences were statistically significant and, on average, metric values were better than those of the other technique in pairwise comparisons. Due to the different granularity of the metrics, plots are organized as follow: Fig. 3b–3c report per-task scores, Fig. 3d–3n per-scenario scores, and Fig. 3o–3y overall scores. With these weights, i.e., assuming that requirements have all the same importance for the application the techniques would be used into, the following ranking is obtained: JS (54.5pts), AS (53.0pts), CV (23.6pts), and WIP (17.9pts).

It is worth remarking that both per-metric and overall scores shall not be interpreted in absolute terms: in fact, they indicate to what extent a given technique matches the weighted set of requirements compared to the other techniques included in the current evaluation. As said, by varying the subset of techniques being compared or the weights assigned to requirements, scores and ranking could change. Moreover, it is also worth observing that, for the sake of clarity, in Fig 3 the results of the pairwise comparisons had to be omitted: thus, it is not always possible to determine which was the technique that scored better in a specific comparison. For instance, considering Fig. 3a it is possible to observe that, overall, JS was statistically faster than other techniques in a larger number of tasks compared to the counterparts; however, from the scores it is not possible to determine, e.g., whether JS was better than another specific technique in a given task. When the number of points for a metric in a given scenario or task (depending on the metric type) is equal to the number of considered techniques minus one, then the technique that scored significantly better than any other one can be identified: this is the case of CV for what it concerns mental effort in scenario *S3* (Fig. 1k). To avoid ambiguities, average values and statistical significance are reported in the Appendix at *http://tiny.cc/quxlsz* as well as in the provided RDB and WDB (spreadsheet above).

5.4 Assigning Weights and Making the Selection

In order to show how to use the generated WDB for selecting the technique maximizing the weighted requirements for a given scenario, a VR game named *VIVECraft* [73] was considered. This is a modified version of the popular game *Minecraft* (a survival game with "pixelated" graphics) with support for VR. Reasons for choosing it are manifold: it is conveniently available, it is quite complex from the point of view of stressed FRs and NFRS and, more importantly, it natively supports more than one locomotion technique; hence, it would be reasonable to study which of the currently supported or of the existing techniques score better.

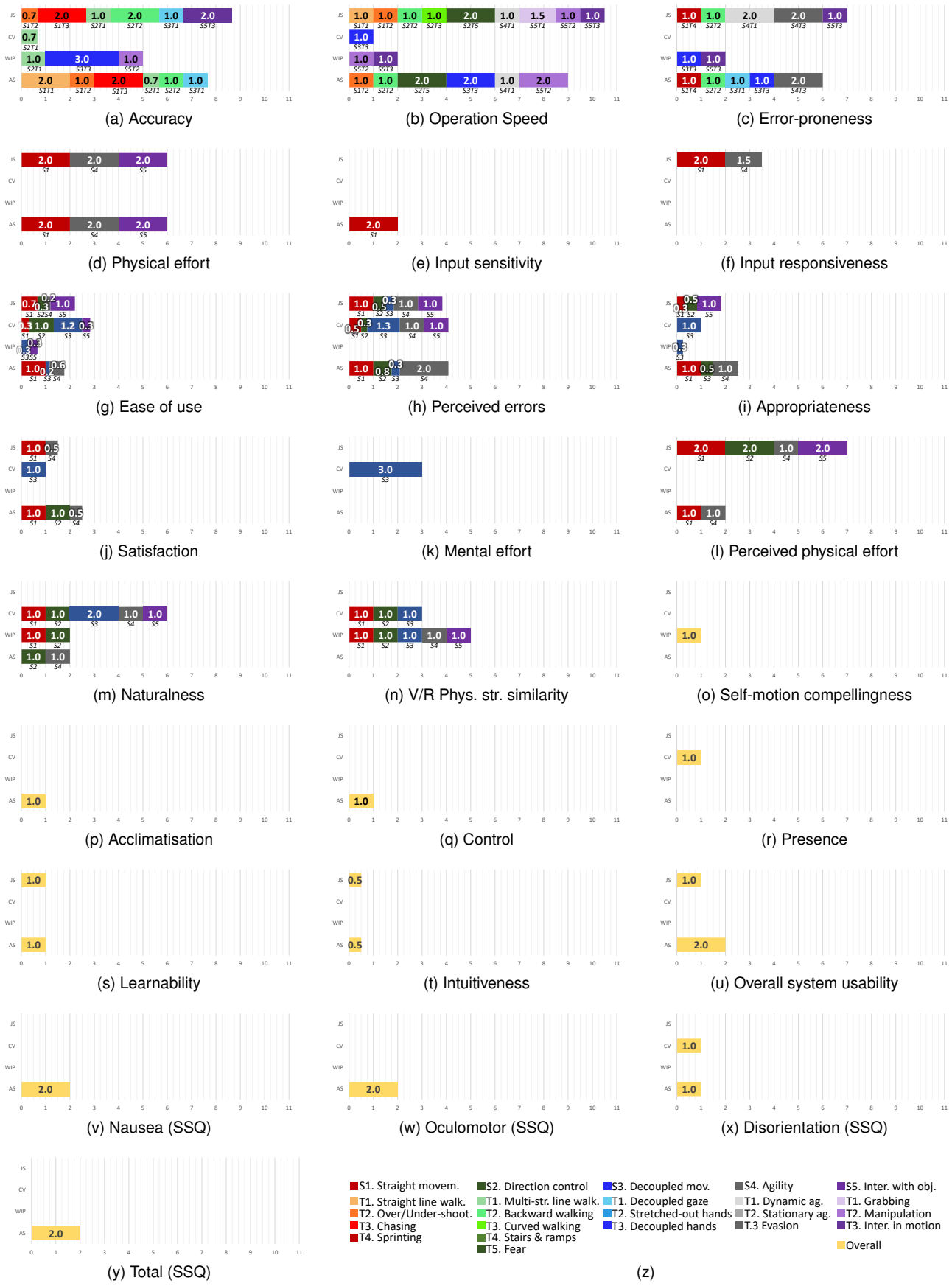


Fig. 3. Scores obtained by the four techniques for all the metrics (weights set to 1). The same scale (0–11) was used in order to make the scores visually comparable. Values correspond to points obtained in pairwise comparisons. Missing values correspond to situations in which the given technique was never statistically better than any of the other techniques.The color coding is the same adopted in Fig. 1 and Table 3.

In order to set the FR weights, a testbed user should first consider that, in the game, it is crucial to move along lines and stop precisely at given points, e.g., to safely cross bridges, etc. (hence, it could set both $w_{S1.T1}$ and $w_{S1.T2}$ to 1). Control on direction change is also important, though not as much as previous requirements ($w_{S2.T1} = 0.5$). Properly managing direction on curved paths is a marginal plus ($w_{S2.T3} = 0.1$). On the contrary, backward walking is a great feature to have, especially during combat phases ($w_{S2.T2} = 1$), as much as being able to run ($w_{S1.T4} = 0.75$). The ability to react to dangerous situations like enemies appearance and to effectively engage in melee fighting ($w_{S2.T5} = 0.5$, $w_{S4.T1} = 0.5$), as well as to move nimbly and dodge attacks ($w_{S4.T2} = 1$, $w_{S4.T3} = 1$) are particularly relevant. Due to the specific implementation, decoupling gaze from movement is not mandatory, but decoupling hands can have a positive impact ($w_{S3.T1} = 0$, $w_{S3.T3} = 0.25$). Interaction is required for crafting and throwing objects, as well as for managing menus ($w_{S5.T1} = 0.5$, $w_{S5.T2} = 0.5$). Interaction in motion is a clear advantage, though not strictly mandatory ($w_{S5.T3} = 0.75$). Due the specific implementation of up/down movements which are based on stairs, corresponding weights could be set accordingly ($w_{S2.T4} = 1$, $w_{ST} = 1$, $w_{RA} = 0$). Concerning NFRs, based on the above discussion Input Responsiveness is very important ($w = 1$), whereas Sensitivity is somehow less relevant ($w = 0.25$). Operation speed ($w = 0.5$) is less crucial than Accuracy and Error-proneness ($w = 1$ and $w = 0.75$, respectively). Since *VIVECraft* is a playful application, it is preferable to have low fatigue so that users can engage in longer sessions ($w = 1$ and direction set to negative for Physical effort). Moreover, Enjoyability, Naturaless, Comfort, Self-motion compellingness, Presence are all considered equally important ($w = 1$). Similar considerations apply to motion sickness and discomfort ($w = 1$ for both Motion sickness total score and SUD). The weights of the Ease of use, Mental effort, Appropriateness, Acclimatisation, Learnability, Intuitiveness and Overall system usability were set to 0.5, being all of them rather important for an application not targeted to expert VR users. Remaining requirements were set to 0.

With this arbitrary configuration, the generated WDB would produce the following ranking: JS (28.4pts), AS (25.6pts), CV (5.68pts), and WIP (4.4pts). For sake of completeness, it is worth noticing that actual implementations of the JS and AS techniques in the game slightly differ from those considered in the WDB. Hence, a new WDB shall be created with those implementations, or matching implementations must be introduced in the game. Integrations to the WDB would be requested also to include in the comparison, e.g., AS-seated in case AS-standing is already available. As anticipated, changing the number of techniques in the comparison, scores (and ranking) could vary. For example, if CV is left out from the comparison, ranking changes as follows: AS (18.8pts), JS (18.1pts), WIP (3.7pts). This variation indicates that JS was ranked first due to points earned in pairwise comparisons with CV.

6 Conclusions and Future Work

In this paper, a testbed supporting the comparison of techniques for locomotion in large-scale virtual environments from many different perspectives is proposed.

Compared to works like, e.g., [38], [40], [41], [43], which present the results of ad-hoc evaluations, the devised testbed integrates and complements a heterogeneous collection of analysis methods and tools reported in the literature by proposing a comprehensive set of both objective and subjective locomotion-related metrics to be evaluated in the execution of a variety of representative tasks. Moreover, differently than in previous works where other evaluation approaches were defined (e.g., [8], [9], [45], [47]), in this paper the methodology underlying the proposed testbed is accompanied by a scoring system that allows potential users to identify the locomotion techniques that best fit a set of user-weighted requirements. A use case showing how to leverage the testbed for choosing, among four known alternatives, the best techniques given a well-defined application scenario is also illustrated. The procedure followed for collecting and processing experimental data is discussed in detail in order to foster the reproducibility of results.

The testbed is released as open source, and comes with all the results obtained applying the devised methodology. This way, comparisons could be performed among the four techniques considered so far by setting weights that match the requirements of a particular application; new experiments could be performed with the same techniques to improve statistical significance of collected measurements; alternative ways for testing significance could be also exploited; finally, new techniques could be integrated, and new tasks (with corresponding metrics) could be added as well in order to investigate other FRs (NFRs).

A present limitation of the scoring system lays in the method used to account for the contribution of the various metrics and solve the multi-criteria decision-making problem. In fact, with the selected method, adding and removing a technique to/from the analysis could lead to a different ranking. Moreover, the approach adopted to normalize the above contributions does not account for the actual magnitude of the differences between the (mean) values of a given metric for the various techniques, but it just considers statistical significance: that is, it values the fact that a technique is better than another one, but it does not consider how much it is better (for a given metric). Finally, overall results depend on the number of statistical differences found; thus, even though one may decide to simply discard metrics for which a significance is found only for a few techniques, it is indeed necessary to work towards extending the availability of experimental data in order to improve the robustness of obtained scores.

In the future, it is planned to address the above issues by defining alternative approaches to normalization (using different significance thresholds, extending the interval of points that can be assigned, etc.) and to scoring, in general; to this aim, the possibility to collect also data about real walking and to use them in the normalization step will be investigated. Further experiments with these and other techniques will be performed, since more data are needed to extend the representativeness and significance of the publicly provided dataset. Moreover, although the testbed already supports a fair number of FRs, there are still some others missing (e.g., regarding movement in the third-dimension) which could be added. Action-specific tasks

could be included as well, to better focus on interactions that are typical of specific applications (e.g., first-person shooter games). Other scenarios supporting the study of jumping, swimming, climbing and even flying configurations could be considered too. Similar considerations apply to NFRs, e.g., pertaining psychological and/or cognitive aspects, such as spatial awareness, sense of direction, navigation abilities, etc. To this purpose, a study aimed to identify possibly missing requirements and to characterize the testbed discrimination capabilities could be performed.

ACKNOWLEDGMENTS

This work has been supported by VR@POLITO initiative.

REFERENCES

[1] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev, *3D User interfaces: Theory and practice*, 2004.
[2] N. C. Nilsson, S. Serafin, F. Steinicke, and R. Nordahl, "Natural walking in virtual reality: A review," *Computers in Entert.*, vol. 16, pp. 8:1–8:22, 2018.
[3] R. A. Ruddle and S. Lessels, "For efficient navigational search, humans require full physical movement, but not a rich visual scene," *Psychological Science*, vol. 17, pp. 460–465, 2006.
[4] D. Waller and E. Hodgson, "Sensory contributions to spatial knowledge of real and virtual environments," in *Human Walking in Virtual Environments*, 2013, pp. 3–26.
[5] E. A. Suma, S. Babu, and L. F. Hodges, "Comparison of travel techniques in a complex, multi-level 3D environment," in *Proc. IEEE Symp. on 3D UI*, 2007, pp. 147–153.
[6] A. Garg, J. A. Fisher, W. Wang, and K. P. Singh, "ARES: An application of impossible spaces for natural locomotion in VR," in *Proc. CHI Conf.*, 2017, pp. 218–221.
[7] N. C. Nilsson, S. Serafin, and R. Nordahl, "The perceived naturalness of virtual locomotion methods devoid of explicit leg movements," in *Proc. Motion in Games*, 2013, pp. 155–164.
[8] D. A. Bowman, D. Koller, and L. F. Hodges, "Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques," in *Proc. IEEE Virtual Reality*, 1997, pp. 45–52.
[9] M. C. Whitton, J. V. Cohn, J. Feasel, P. Zimmons, S. Razzaque, S. J. Poulton, B. McLeod, and F. P. Brooks, "Comparing VE locomotion interfaces," in *Proc. IEEE Virtual Reality*, 2005, pp. 123–130.
[10] M. Schuemie, B. Abel, C. van der Mast, M. Krijn, and P. Emmelkamp, "The effect of locomotion technique on presence, fear and usability in a virtual environment," in *Proc. Europ. Conf. on Media, Communication & Film*, 2005, pp. 129–135.
[11] D. A. Bowman, D. B. Johnson, and L. F. Hodges, "Testbed evaluation of virtual environment interaction techniques," *Presence: Teleoperators & Virtual Environments*, vol. 10, pp. 75–95, 2001.
[12] C. Boletsis, "The new era of virtual reality locomotion: a systematic literature review of techniques and a proposed typology," *Multimodal Technologies and Interaction*, vol. 1, p. 24, 2017.
[13] M. Al Zayer, P. MacNeilage, and e. folmer, "Virtual locomotion: A survey," *IEEE Trans. on Visualization & Computer Graphics*, 2018.
[14] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller, "State of the art of virtual reality technology," in *2016 IEEE Aerospace Conf.*, 2016, pp. 1–19.
[15] J. C. Cardoso and A. Perrotta, "A survey of real locomotion techniques for immersive virtual reality applications on head-mounted displays," *Computers & Graphics*, vol. 85, pp. 55–73, 2019.
[16] W. B. Lathrop and M. K. Kaiser, "Perceived orientation in physical and virtual environments: Changes in perceived orientation as a function of idiothetic information available," *Presence: Teleoperators & Virtual Environments*, vol. 11, pp. 19–32, 2002.
[17] J. N. Templeman, P. S. Denbrook, and L. E. Sibert, "Virtual locomotion: Walking in place through virtual environments," *Presence*, vol. 8, pp. 598–617, 1999.
[18] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey, "Point & teleport locomotion technique for virtual reality," in *Proc. Symp. on Computer-Human Interaction in Play*, 2016, pp. 205–216.
[19] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual reality on a WIM: Interactive worlds in miniature," in *Proc. SIGCHI Conf. on Human factors in Computing Systems*, 1995, pp. 265–272.
[20] S. L. Stoev, D. Schmalstieg, and W. Straßer, "Two-handed through-the-lens-techniques for navigation in virtual environments," in *Immersive Projection Tech. and Virtual Env.*, 2001, pp. 51–60.
[21] L. P. Fiore, E. Coben, S. Merritt, P. Liu, and V. Interrante, "Towards enabling more effective locomotion in VR using a wheelchair-based motion platform," in *Proc. 5th Joint VR Conf.*, 2013, pp. 83–90.
[22] J. Wang and R. W. Lindeman, "Comparing isometric and elastic surfboard interfaces for leaning–based travel in 3D virtual environments," in *IEEE Symp. on 3D UI*, 2012, pp. 31–38.
[23] S. Beckhaus, K. J. Blom, and M. Haringer, "ChairIO–the chair-based interface," *Concepts and Tech. for Pervasive Games: A Reader for Pervasive Gaming Research*, vol. 1, pp. 231–264, 2007.
[24] S. Fels, Y. Kinoshita, T.-P. G. Chen, Y. Takama, S. Yohanan, A. Gadd, S. Takahashi, and K. Funahashi, "Swimming across the pacific: A VR swimming interface," *IEEE Computer Graphics & Applications*, vol. 25, pp. 24–31, 2005.
[25] J. L. Souman, P. R. Giordano, M. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. D. Luca, H. H. Bülthoff, and M. O. Ernst, "Cyberwalk: Enabling unconstrained omnidirectional walking through virtual environments," *ACM TAP*, vol. 8, pp. 25:1–25:22, 2011.
[26] H. Iwata, H. Yano, H. Fukushima, and H. Noma, "Circulafloor [locomotion interface]," *IEEE Computer Graphics & Applications*, vol. 25, pp. 64–67, 2005.
[27] E. Medina, R. Fruland, and S. Weghorst, "Virtusphere: Walking in a human size VR 'hamster ball'," in *Proc. Human Factors and Ergonomics Society Annual Meeting*, 2008, pp. 2102–2106.
[28] L. Avila and M. Bailey, "Virtual reality for the masses," *IEEE Computer Graphics & Applications*, vol. 34, pp. 103–104, 2014.
[29] D. Swapp, J. Williams, and A. Steed, "The implementation of a novel walking interface within an immersive display," in *Proc. IEEE Symp. on 3D UI*, 2010, pp. 71–74.
[30] H. Iwata and T. Fujii, "Virtual perambulator: A novel interface device for locomotion in virtual environment," in *Proc. IEEE Virtual Reality Ann. Int. Symposium*, 1996, pp. 60–65.
[31] T. Cakmak and H. Hager, "Cyberith virtualizer: A locomotion device for virtual reality," in *Proc. ACM SIGGRAPH*, 2014, p. 6.
[32] E. Guy, P. Punpongsanon, D. Iwai, K. Sato, and T. Boubekeur, "LazyNav: 3D ground navigation with non–critical body parts," in *IEEE Symp. on 3D UI*, 2015, pp. 43–50.
[33] A. Kitson, A. M. Hashemian, E. R. Stepanova, E. Kruijff, and B. E. Riecke, "Comparing leaning–based motion cueing interfaces for virtual reality locomotion," in *IEEE Symp. 3D UI*, 2017, pp. 73–82.
[34] E. Kruijff, A. Marquardt, C. Trepkowski, R. W. Lindeman, A. Hinkenjann, J. Maiero, and B. E. Riecke, "On your feet!: Enhancing vection in leaning-based interfaces through multisensory stimuli," in *Proc. Symp. on Spatial User Interact.*, 2016, pp. 149–158.
[35] G. de Haan, E. J. Griffith, and F. H. Post, "Using the wii balance board as a low-cost vr interaction device," in *Proc. ACM Symp. on Virtual Reality Software and Technology*, 2008, pp. 289–290.
[36] A. Harris, K. Nguyen, P. T. Wilson, M. Jackoski, and B. Williams, "Human joystick: Wii-leaning to translate in large virtual environments," in *Proc. ACM SIGGRAPH Int. Conf. on Virtual Reality Continuum and its Applications in Industry*, 2014, pp. 231–234.
[37] J. Feasel, M. C. Whitton, and J. D. Wendt, "LLCM-WIP: Low-latency, continuous-motion walking-in-place," in *Proc. IEEE Symp. on 3D UI*, 2008, pp. 97–104.
[38] Y. S. Pai and K. Kunze, "Armswing: Using arm swings for accessible and immersive navigation in AR/VR spaces," in *Proc. Int. Conf. on Mobile and Ubiquitous Multimedia*, 2017, pp. 189–198.
[39] E. A. Suma, Z. Lipps, S. Finkelstein, D. M. Krum, and M. Bolas, "Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture," *IEEE Trans. on Visualization & Computer Graphics*, vol. 18, no. 4, pp. 555–564, 2012.
[40] P. T. Wilson, W. Kalescky, A. MacLaughlin, and B. Williams, "VR locomotion: walking> walking in place> arm swinging," in *Proc. 15th ACM SIGGRAPH Conf. on VR Continuum and Its Applications in Industry*, 2016, pp. 243–249.
[41] D. Calandra, M. Billi, F. Lamberti, A. Sanna, and R. Borchiellini, "Arm swinging vs treadmill: A comparison between two techniques for locomotion in virtual reality," in *Proc. Eurographics*, 2018, pp. 53–56.
[42] D. Calandra, F. Lamberti, and M. Migliorini, "On the usability of consumer locomotion techniques in serious games: Comparing arm swinging, treadmills and walk-in-place," in *IEEE Int. Conf. on Consumer Electronics - Berlin*, 2019, pp. 1–5.
[43] N. C. Nilsson, S. Serafin, M. H. Laursen, K. S. Pedersen, E. Sikstrom, and R. Nordahl, "Tapping-in-place: Increasing the natu-

ralness of immersive walking-in-place locomotion through novel gestural input," in *Proc. IEEE Symp. on 3D UI*, 2013, pp. 31–38.
[44] B. Sarupuri, S. Hoermann, M. C. Whitton, and R. W. Lindeman, "Lute: A locomotion usability test environment for virtual reality," in *Int. Conf. on Virtual Worlds and Games for Serious Applications*, 2018, pp. 1–4.
[45] J. Albert and K. Sung, "User-centric classification of virtual reality locomotion," in *Proc. ACM Symp. VR Soft. and Tech.*, 2018, p. 127.
[46] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, 1988, vol. 52, pp. 139 – 183.
[47] A. Ferracani, D. Pezzatini, J. Bianchini, G. Biscini, and A. Del Bimbo, "Locomotion by natural gestures for immersive virtual environments," in *Proc. 1st Int. Workshop on Multimedia Alt. Realities*, 2016, pp. 21–24.
[48] G. Wilson, M. McGill, M. Jamieson, J. R. Williamson, and S. A. Brewster, "Object manipulation in virtual reality under increasing levels of translational gain," in *Proc. CHI Conf. on Human Factors in Computing Systems*, 2018, p. 99.
[49] J. Mayor, L. Raya, and A. Sanchez, "A comparative study of virtual reality methods of interaction and locomotion based on presence, cybersickness and usability," *IEEE Trans. on Emerging Topics in Computing*, 2019.
[50] J.-F. Lapointe and P. Savard, "A comparative study of three bimanual travel techniques for desktop virtual walkthroughs," in *Proc. IEEE Work. on Haptic Audio Visual Env.& Games*, 2009, pp. 182–185.
[51] G. Loup and E. Loup-Escande, "Effects of travel modes on performances and user comfort: A comparison between ArmSwinger and Teleporting," *Int. Jr. of Human-Computer Inter.*, pp. 1–9, 2018.
[52] E. Suma, S. Finkelstein, M. Reid, S. Babu, A. Ulinski, and L. F. Hodges, "Evaluation of the cognitive effects of travel technique in complex real and virtual environments," *IEEE Trans. on Visualization & Computer Graphics*, vol. 16, pp. 690–702, 2010.
[53] M. Nabiyouni, A. Saktheeswaran, D. A. Bowman, and A. Karanth, "Comparing the performance of natural, semi-natural, and non-natural locomotion techniques in virtual reality," in *Proc. IEEE Symp. on 3D UI*, 2015, pp. 3–10.
[54] M. Asnabrygg, "Unbreakable VR runner," 2016. [Online]. Available: https://store.steampowered.com/app/494310/Unbreakable_Vr_Runner
[55] R. Paris, M. Joshi, Q. He, G. Narasimham, T. P. McNamara, and B. Bodenheimer, "Acquisition of survey knowledge using walking in place and resetting methods in immersive virtual environments," in *Proc. ACM Symposium on Applied Perception*, 2017, p. 7.
[56] Bethesda Game Studios, "Fallout 4 VR," 2017. [Online]. Available: https://store.steampowered.com/app/611660/Fallout_4_VR
[57] Echo Combat. [Online]. Available: https://www.oculus.com/experiences/event/1713610428731487/
[58] Metricminds GmbH and Co KG, "Catch and Release," 2018. [Online]. Available: https://store.steampowered.com/app/679750/Catch__Release/
[59] Beat-Games, "Beat Saber," 2018. [Online]. Available: https://store.steampowered.com/app/620980/Beat_Saber/
[60] SUPERHOT-Team, "Super Hot VR," 2017. [Online]. Available: https://www.playstation.com/en-us/games/superhot-vr-ps4/
[61] Owlchemy Labs, "Job simulator," 2016. [Online]. Available: https://store.steampowered.com/app/448280/Job_Simulator
[62] S. L. Fischer, P. B. Watts, R. L. Jensen, and J. Nelson, "Energy expenditure, heart rate response, and metabolic equivalents (mets) of adults taking part in children's games," *The Jr. of Sports Medicine and Physical Fitness*, vol. 44, p. 398, 2004.
[63] K. R. Westerterp, "Assessment of physical activity: A critical appraisal," *Europ. Jr. of Applied Physiology*, vol. 105, pp. 823–828, 2009.
[64] A. P. Hills, N. Mokhtar, and N. M. Byrne, "Assessment of physical activity and energy expenditure: An overview of objective measures," *Frontiers in Nutrition*, vol. 1, p. 5, 2014.
[65] 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. Jr. of Aviation Psychology*, vol. 3, pp. 203–220, 1993.
[66] R. S. Kalawsky, "VRUSE–A computerised diagnostic tool: For usability evaluation of virtual/synthetic environment systems," *Applied ergonomics*, vol. 30, pp. 11–25, 1999.
[67] S. Bruck and P. A. Watters, "Estimating cybersickness of simulated motion using the simulator sickness questionnaire (SSQ): A controlled study," in *Proc. 6th Int. Conf. on Computer Graphics, Imaging and Visualization*, 2009, pp. 486–488.
[68] ElectricNightOwl, "Arm Swinger." [Online]. Available: https://github.com/ElectricNightOwl/ArmSwinger
[69] C. Boletsis and J. E. Cedergren, "VR locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques," *Advances in Human-Computer Interaction*, vol. 2019, 2019.
[70] J. B. Cronin and K. T. Hansen, "Strength and power predictors of sports speed," *Jr. Strength & Cond. Res.*, vol. 19, pp. 349–357, 2005.
[71] R. W. Bohannon, "Comfortable and maximum walking speed of adults aged 2079 years: reference values and determinants," *Age and Ageing*, vol. 26, pp. 15–19, 1997.
[72] B. K. Jaeger and R. R. Mourant, "Comparison of simulator sickness using static and dynamic walking simulators," in *Proc. of Human Factors & Ergonomics Society A.M.*, vol. 45, 2001, pp. 1896–1900.
[73] "VIVECraft," 2019. [Online]. Available: http://www.vivecraft.org/how-to-play/



Alberto Cannavò received his B.Sc. degree from University of Messina, Italy, in 2013. He received his M.Sc. and PhD. degrees in computer engineering from Politecnico di Torino, Italy, in 2015 and 2020. He now holds an assistant researcher position at the Dipartimento di Automatica e Informatica of Politecnico di Torino. His fields of interest include computer graphics and human-machine interaction.



Davide Calandra received his B.Sc. and M.Sc. degrees in computer engineering from Politecnico di Torino, Italy, in 2014 and 2017. Currently, he is a Ph.D. student at Politecnico di Torino, with interests in interactive graphics applications, virtual and augmented reality, and human-machine interaction.



Filippo Gabriele Prattichò received his B.Sc. and M.Sc. degrees in computer engineering from Politecnico di Torino, Italy, in 2014 and 2017. Currently, he is a Ph.D. student at Politecnico di Torino, where he carries out research in the areas of virtual, augmented and mixed reality, human-computer and human-robot interaction, serious games, and user experience design.



Valentina Gatteschi received her B.Sc. and M.Sc. degrees in management engineering and her Ph.D. degree in computer engineering from Politecnico di Torino, Italy, in 2005, 2008 and 2013. She is now an assistant professor at the Dipartimento di Automatica e Informatica of Politecnico di Torino. Her interests include intelligent systems and their application to computer graphics.



Fabrizio Lamberti received his M.Sc. and Ph.D. degrees in computer engineering from Politecnico di Torino, Italy, in 2000 and 2005. He is now a full professor at at the Dipartimento di Automatica e Informatica of Politecnico di Torino. His interests pertain computer graphics, human-machine interaction and computational intelligence. He serves as an Associate Editor for several journals, including the IEEE Transactions on Computers and the IEEE Transactions on Learning Technologies.