Keep my head on my shoulders!

Why third-person is bad for navigation in VR

Daniel Medeiros
INESC-ID / Instituto Superior Técnico
Lisbon, Portugal
daniel.medeiros@tecnico.ulisboa.pt

Rafael K. dos Anjos
FCSH-UNL / INESC-ID
Lisbon, Portugal
rkuffner@fcsh.unl.pt

Daniel Mendes
INESC-ID / Instituto Superior Técnico
Lisbon, Portugal
danielmendes@tecnico.ulisboa.pt

João Madeiras Pereira
INESC-ID / Instituto Superior Técnico
Lisbon, Portugal
jap@inesc-id.pt

Alberto Raposo
Tecgraf/PUC-Rio
Rio de Janeiro, Brasil
abraposo@tecgraf.puc-rio.br

Joaquim Jorge
INESC-ID / Instituto Superior Técnico
Lisbon, Portugal
jorgej@acm.org

ABSTRACT

Head-Mounted Displays are useful to place users in virtual reality (VR). They do this by totally occluding the physical world, including users' bodies. This can make self-awareness problematic. Indeed, researchers have shown that users' feeling of presence and spatial awareness are highly influenced by their virtual representations, and that self-embodied representations (avatars) of their anatomy can make the experience more engaging. On the other hand, recent user studies show a penchant towards a third-person view of one's own body to seemingly improve spatial awareness. However, due to its unnaturality, we argue that a third-person perspective is not as effective or convenient as a first-person view for task execution in VR. In this paper, we investigate, through a user evaluation, how these perspectives affect task performance and embodiment, focusing on navigation tasks, namely walking while avoiding obstacles. For each perspective, we also compare three different levels of realism for users' representation, specifically a stylized abstract avatar, a mesh-based generic human, and a real-time point-cloud rendering of the users' own body. Our results show that only when a third-person perspective is coupled with a realistic representation, a similar sense of embodiment and spatial awareness is felt. In all other cases, a first-person perspective is still better suited for navigation tasks, regardless of representation.

CCS CONCEPTS

• **Human-centered computing → User studies**; • **Computing methodologies → Mixed / augmented reality**; **Perception**;

KEYWORDS

Travel, Avatar, Virtual Reality, Full-body tracking, Augmented Reality, Embodiment

VRST '18, November 29-December 1, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6086-9/18/11…$15.00
https://doi.org/10.1145/3281505.3281511

ACM Reference Format:
Daniel Medeiros, Rafael K. dos Anjos, Daniel Mendes, João Madeiras Pereira, Alberto Raposo, and Joaquim Jorge. 2018. Keep my head on my shoulders!: Why third-person is bad for navigation in VR. In *24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), November 29-December 1, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3281505.3281511

1 INTRODUCTION

When using a Head-Mounted Display system (HMD), a person's body is completely occluded, diminishing the sense of presence in the virtual space. The use of a self-embodied representation (avatar) can overcome this issue [32]. Its use is also essential to improve the user's spatial awareness [7] which also affects the way users make distance judgments [23]. Thus, an important aspect of measuring the effectiveness of a Virtual Reality (VR) experience is the sense of embodiment. Kilteni et al. [11] define embodiment as being "the sense that emerges when a virtual body's properties are processed as if they were the properties of one's own biological body" and this sense affects how users interact with virtual objects in Virtual Reality [11]. The sense of embodiment of an avatar is subdivided in three components: (i) the sense of agency, i.e. feeling of motor control over the virtual body; (ii) the sense of body ownership, i.e. feeling that the virtual body is one's own body; and (iii) self-location, i.e. the experienced location of the self.

Some aspects are known to influence a person's sense of embodiment when using an avatar, namely the realism of the representation and the perspective which is viewed. Normally, users are depicted in their own perspective (known as First-Person Perspective or 1PP). Another possibility is Third-Person perspective (3PP) where the virtual camera is positioned behind the person, allowing them to view their own full-virtual body. This representation is widely used in games for improving spatial awareness in conventional displays [6, 29].

In VR, studies also indicate a slight improvement on users' spatial awareness when an avatar is seen in a third-person perspective. Studies using an HMD indicate that an user can feel a high sense of embodiment in these setups, which may indicate that users' real bodies may need to be seen in VR to provide a full-body illusion [24]. Some other authors suggest a similar sense of embodiment in real-world scenarios such as reaching objects [5], but due to the lack of efficiency and its unnaturalness it may not be suited for these tasks. Gorisse et al [9] also suggest a slight improvement in spatial awareness but their results are limited by subjective metrics.

In this paper we further study perspective (1PP and 3PP) and realism of the representation of self-embodied avatars in users' sense of embodiment and spatial awareness factors in Virtual Reality setups. To this end, we use three different representations varying the level of realism of each representation, from an abstract to a realistic humanoid representation. The abstract representation uses spheres and boxes to represent parts of the body. The second is a realistic mesh avatar that is rigged and deformed according to tracking information. The third representation is a low cost point-cloud based avatar, using extracted video information from a person's real body that is mapped into the virtual environment. Studies indicate that the realism of the representation improves embodiment factors, but just slightly, when using a first-person avatar [14]. But, since users' bodies are always seen when using a 3PP avatar, we hypothesize that the realism of the representation highly impacts both spatial awareness and embodiment factors. For assessing both spatial awareness and embodiment factors we use navigation tasks, where users are asked to physically walk while avoiding obstacles in a Virtual Environment. The obstacles are positioned around the user, near users' feet and at head level.

Our results show a clear advantage in navigation tasks with avatars in the first person perspective. Users in this perspective completed the tasks more efficiently and reported a higher sense of embodiment. An exception is the point-cloud representation, which provides users with a similar embodiment and spatial awareness in both 1PP and 3PP conditions. We also prove that in the 3PP condition, the level of realism of the representation highly impacts awareness and embodiment factors.

In the following sections we present related work on user-representations, describe the experiments, report on measures, discuss the results obtained and propose a set of guidelines for Self-Embodied Virtual Reality applications.

2 RELATED WORK

An important part of the VR experience delves into how users are represented in virtual scenes. As opposed to CAVE-like systems, Head-Mounted Displays occlude users' real selves, compromising the overall virtual-reality session. A way to overcome this problem is to use a fully-embodied representation of the person within the virtual environment [31]. Self-embodied avatars are also a key ingredient to the feeling of Presence inside an Immersive Virtual Environment [28]. The use of a virtual body also improves users distance estimation [16] and spatial awareness [7] by providing users with a reference of size and distance of their surroundings [10, 23]. However, the known effect of distance underestimation is still present when a virtual avatar is used [22].

Avatars' level of realism also plays an important part on the VR experience and how it relates to the sense of embodiment of users. A common issue reported in both robotics and animation is the uncanny valley effect [19], which can be less noticeable when characters are animated [21]. This effect is also proven to slightly affect presence and embodiment in 1PP avatars when viewed through a head-mounted display [14].

Another factor that can impact the VR experience is the perspective from which the virtual body is viewed. On the first-person perspective, the virtual camera is in the avatar's eyes, simulating a real-life condition. In the third-person perspective, on the other hand, the virtual camera is placed outside the virtual self (normally behind the avatar's head) giving users an external view of their virtual-self. The use of a third-person perspective may compromise the naturalness of the interaction but can improve awareness of their surroundings [25].

Additionally, artificial bodies can still provide a high sense of embodiment even when viewed from different points-of-view. A classical extra-corporeal experience is the Rubber Hand Illusion (RHI) [1]. This illusion has a similar counterpart in Virtual Reality, which is called Virtual Hand illusion, and can be induced by visuotactile [30] and visuomotor synchrony [27, 34]. The RHI has also proven to work with a full body [20]. Additional works by Ehrsson et al. [8] and Leggenhager et al. [13] prove that people can still feel embodied using a see-through head-mounted display, and co-located with their virtual body, when they see an image of their own body from a different point-of-view. Boulic et al. [2] also indicate the importance of a third-person avatar for posture validation and collision checking between users' bodies and the virtual environment.

Further work by Salamin et al. [26] used an augmented-reality setup with a displaced camera and an HMD to show that the best perspective depends on the performed action: 1PP can improve object manipulation precision, while 3PP can improve performance in moving actions. Work by the same authors also showed that people preferred the 3PP in comparison to 1PP, and that 3PP avatars required less training in moving actions, such as catching a thrown ball [24, 25]. Kosch et al. [12] found that the preferred viewpoint in a 3PP is behind the person's head, thus providing a real life-like third person experience. 3PP avatars also influence the depth perception of humans, which causes the well-known problem of underestimation [24]. Regarding embodiment, Maselli & Slater [15] state that people cannot have a high sense of ownership over a static body in 3PP, but argue that synchronous visuomotor body feedback may affect this sense and provide full-body illusions when using an avatar in the third-person perspective.

Perspective studies in VR generally focus on how an artificial body affects user's embodiment from a third-person perspective. Although the effects of this perspective in the embodiment are an important aspect of the overall experience, few works focus on how these effects influence classical 3DUI tasks such as navigation, selection, and manipulation. Debarba et al [5] on the other hand, show that users can accomplish reach tasks with a high sense of embodiment using both 1PP and 3PP, but have reduced accuracy in 3PP. Additionally, Monteiro et al. [18] use both 1PP and 3PP avatars and suggest the use of avatars in 3PP in order to reduce cybersickness related side-effects. The 3PP is also found to be safer when compared to 1PP in harmful situations [3, 4]. Regarding graphical fidelity in third-person avatars, studies on this matter are still limited to non-rigged avatars and indicate that an avatar with a realistic human-shaped form increases the sense of body-ownership, producing a full-body illusion [15]. Nonetheless, 1PP is still more efficient than 3PP and considered more natural by users [5, 9].

Although studies show a slight improvement in spatial awareness in displaced see-through systems in 3PP, the use of artificial bodies in VR can produce different results. In VR, the use of a different perspective coupled with a virtual representation that does not match users' bodies may aggravate how people make distance judgments when the avatar is animated [17], compromising users' spatial awareness. Previous work [9] claims improvements in spatial awareness in VR with 3PP avatars over 1PP, but their results are limited to subjective responses and, while showing a slight tendency towards 3PP being better, they have no statistical significance. Further objective metrics are needed to assess not only participants' preferences, but also objective measures of this type of representation. Moreover, since users' bodies are always visible when a third-person perspective is used, we theorize that the realism of the representation have a bigger influence in both the sense of embodiment and spatial awareness. Therefore, the use of a real-time reconstruction of people's real bodies can be an important factor for establishing a high sense of embodiment with a third-person view.

3 USER STUDY

In this work, we assess how the perspective and representation affects efficiency and efficacy in navigation tasks. We consider First- and Third-Person perspectives and, for representation, we utilize three different avatars with increasing levels of graphical fidelity. These range from a stylized box-avatar, a humanoid mesh avatar and a real-time point-cloud avatar, which use depth-cameras to map users' representations inside the virtual environment. All avatars have visuomotor synchronicity in order to provide a more realistic experience. To assess efficiency and efficacy we designed three different tasks that consist in walking while avoiding obstacles, which differ in how the obstacles are positioned in the virtual environment. The first consists in avoiding obstacles that are positioned around the user; on the second task and third tasks, users need to make changes on the vertical plane to surpass them, by going over (Task 2) or below obstacles (Task 3). To maximize our tracking space and provide a more realistic experience, we chose to use a circular path in all three conditions.

In this section we describe the main aspects of designing the test experience regarding user representation and the design of the task. The following subsections present the task concept, the avatar representations used and the setup used on the test task.

3.1 User Representations

We chose three different user representations for our evaluation, which are used in both 1PP and 3PP. Camera positioning in 3PP is based on previous work by Kosch et al. [12], in which the camera is positioned above user's head for improved spatial awareness.

In all the used representations, the depth sensors' joints positions and rotations are mapped directly into the avatars using direct Kinematics. Skeleton tracking was performed using the "Creepy Tracker" toolkit from Sousa et al. [33]. This toolkit provides reliable markerless tracking using Kinect sensors, and enables us to follow users in the area necessary for the study (4 meters by 4 meters). A surrounding bounding box to each joint was used as a basis for collision detection between the users and the obstacles on all tasks. The bounding boxes are procedurally generated, and are used for collision detection and for rendering the abstract avatar's body parts.

3.1.1 Abstract. The first avatar is a simplified avatar representation which is composed by abstract components. Spheres were used for each joint, and boxes for each bone connecting joints and the head. These boxes are scaled according to the user and are also used on the other representations for collision detection. Figures 1A and 1B show this representation in both First and Third Person Perspectives (1PP and 3PP), respectively.

3.1.2 Mesh. The second representation is a realistic mesh avatar resembling a human male from the Mixamo [1] character database, since we used only male participants for the test. This representation did not include animation for individual fingers, since they are not tracked by the Kinect sensor. Figures 1C and 1D show this representation in the First and Third Person Perspectives (1PP and 3PP), respectively.

3.1.3 Point Cloud. This body representation is based on a combination of separate streams of point clouds from Microsoft Kinect sensors, which are transmitted by the "Creepy Tracker" toolkit [33] over the network. These point clouds are in the same coordinate system as the skeleton information, which is still used for collision detection. To avoid visual occlusion, we discarded user's head point-cloud information on the First-Person Perspective.

Figures 1E and 1F show this representation on the first and third-person views.

3.2 Methodology

For assessing the effects of Representation and Perspective, we used a 2x3 factor Within-Subjects Test Design. The test was divided into eight stages: 1) introduction to the study and application of pre-test questionnaire; 2) explanation about the tasks and each of the users representations 3) adjustment of the device for comfort; 4) calibration procedure; 5) training; 6) task execution; 7) application of post-test questionnaire; 8) and a semi-structured interview.

At first, we explained the test objectives. Then, the users completed a pre-use questionnaire to raise the participants profile regarding previous experience with related technologies (HMDs, virtual avatars, etc).

Subsequently, we showed a brief description of the tasks and representations used. Furthermore, we executed the calibration procedure. This procedure was performed to calibrate the tracking system between the HMDs and the depth-sensors. Then, in order to familiarize the user with the procedures, users performed a task in a training scenario, where they could freely explore the virtual environment and familiarize themselves with the setup and each of the representations. Users had a maximum of five minutes for this task.

After performing the training task, the users reached a fixed object in the environment and the users performed the test task. Then a questionnaire was given to the users for some user experience issues. These steps were done for each of the combination of the test conditions (perspective and representation) of a total of 12

[1] https://www.mixamo.com/

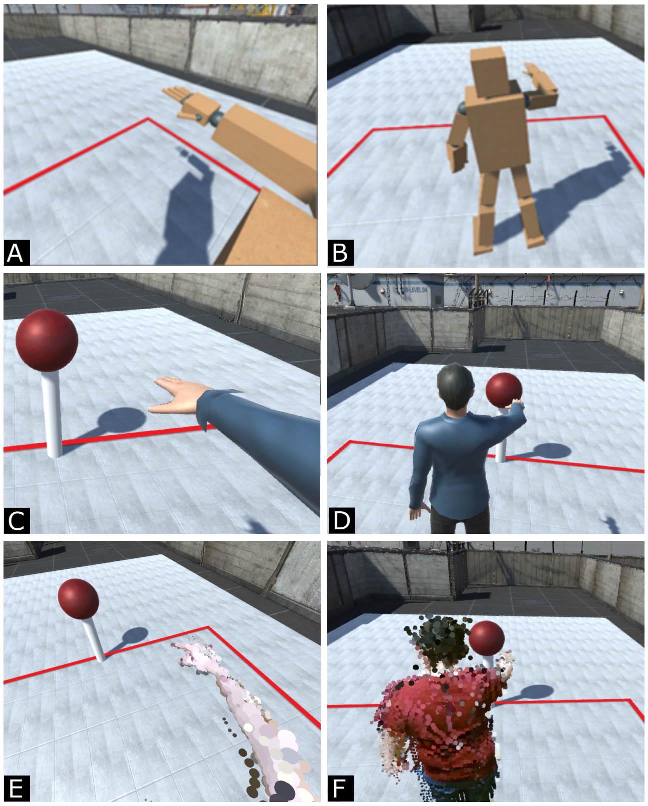

Figure 1: Self-representations used on our study. (A) 1PP Abstract Avatar (B) 3PP Abstract Avatar (C) 1PP Mesh Avatar (D) 3PP Mesh Avatar (E) 1PP Point-Cloud Avatar (F) 3PP Point-Cloud Avatar

permutations. We permuted the order of representations used and the order of perspectives, so if a user performed the order Abstract-Point-Cloud-Mesh in the First-Person Perspective he would do it at the same order on the Third-Person Perspective. The order of the avatar representation was changed in every test, following a Balanced Latin square arrangement, to avoid biased results. After performing each representation-perspective the users filled in a 6-Point Likert Scale Questionnaire to assess embodiment, easiness of completion of each of the tasks and fatigue issues.

3.3 Virtual Environment

The selected environment is based on the Stealth Scene, which was obtained on the Unity Asset Store[2]. This scene was modified to remove visual clutter, to not interfere with the goals of the test by capturing user's attention.

We also included in the environment a representation of the Kinect's tracking limits with a red square, where the user could walk freely.

[2]http://unity3d.com/store

3.4 Tasks Description

In order to isolate different aspects of navigation tasks that we wanted to evaluate, the test was divided into three tasks. For each of the tasks users would go through the test until they reached the starting point again, marked by a green colored sphere, triggering the start of the next task. These were chosen based on natural tasks such as walking while avoiding obstacles based on previous work [24]. To maximize tracking space, we arranged the objects along a circular path, where the user walks anti-clockwise until reaching the initial point. The participants were asked to explicitly avoid the obstacles while performing all three tasks.

In the following subsections we present and explain in further detail each of the proposed tasks.

3.4.1 Task 1 (Barrels Task). In this task, needs to go around the barrels as indicated by the signs on top of them . Figure 2A illustrates the first Task.

3.4.2 Task 2 (Bars Task). In the second task, the user needs to avoid each of the yellow bars by raising their feet (or jumping) until they reach the initial point (Figure 2B).

3.4.3 Task 3 (Tunnels Task). In this test the user needs to go under the two tunnels until they reach the initial point. This tunnel is adjusted according to the user's height, which is estimated using the distance between the head and the toe when the user starts the test. The ceiling of the tunnel is placed 12 centimeters below the user's height (Figure 2C).

3.5 Setup

The physical setup chosen for our study can be seen in figure 3, where a few of the Kinect sensors used for body tracking and point cloud reconstruction can be seen. A wide-baseline setup was used due to two main reasons; firstly the fact that the kinect sensor has a limit on its effective range (0.4m to 4.5m, with skeletons losing reliability starting on 2.5m), and in order to properly evaluate a navigation task, a bigger space was needed. When the user is at the limits of the sensors operating range, the quality of the experience would be compromised, so a wide-baseline setup guarantees the whole body of the user is always visible by at least one camera. Secondly, since a third person perspective is presented as one interaction paradigm, the whole body of the participant must be visible at all times in order to avoid holes in the representation. A narrow baseline or single sensor setup would capture just half of the participant's body, greatly compromising the experience.

Five Kinect sensors were fixated on the walls of the laboratory where the study was being held, covering an area of approximately 4 x 4 meters. The placement of the kinect sensors was chosen in such a way that user's bodies are always visible.

As the visualization platform, we used the Oculus Rift DK2 [3] HMD, which provided the orientation for the camera. The prototype was developed in Unity3D version 5.6.7.

3.6 Questionnaires

To gather users profiles, a pre-test questionnaire was conducted. For assessing human factors such as comfort, sense of embodiment and

[3]https://www.oculus.com/rift

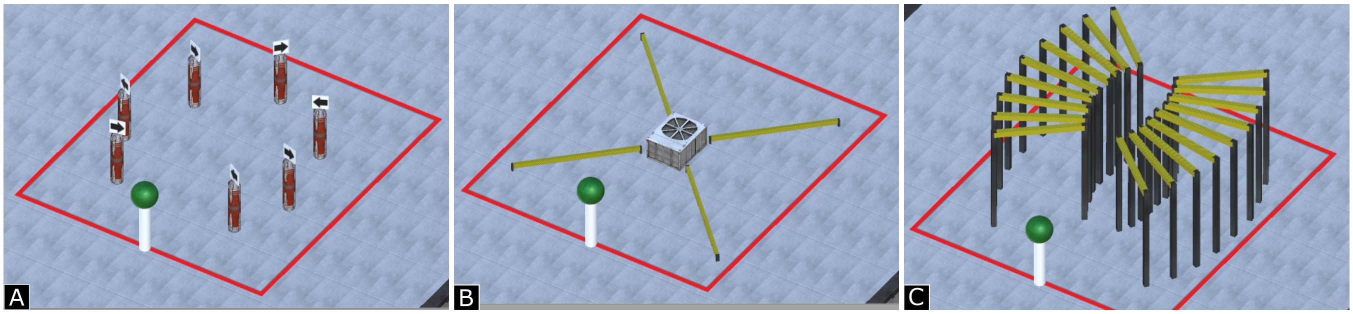

Figure 2: The proposed tasks for our evaluation. The green lollipop marks the initial position of the user.

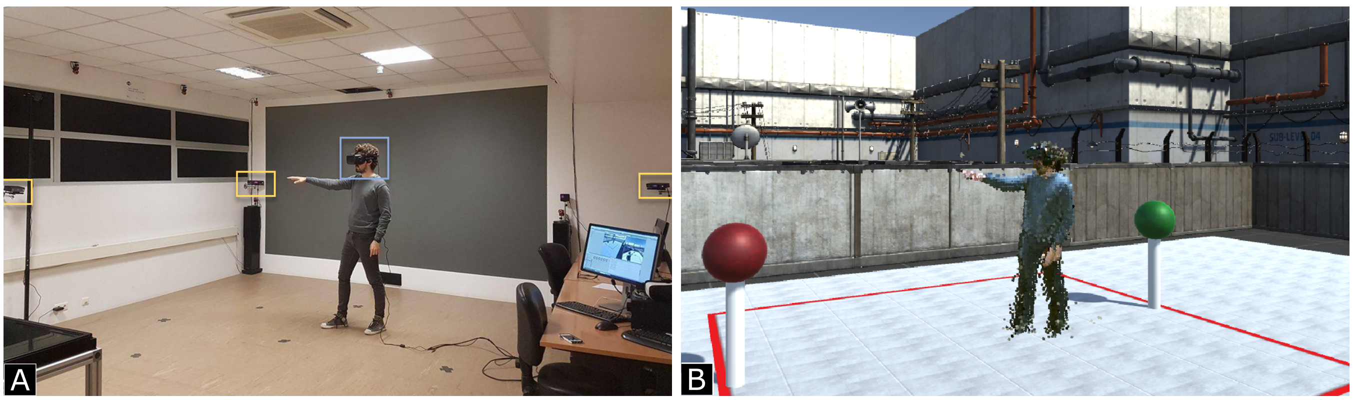

Figure 3: The setup used for our study. Figure A shows the laboratory and one user performing a test, and Figure B the virtual world mapping.

satisfaction, a post-test questionnaire was conducted. The post-test questionnaire was comprised of a list of 11 statements followed by a Likert Scale of 6 values to force users to take a position, where 1 means that the user does not agree completely with statement and 6 means he fully agrees with it, as summarized in Table 1. The first four questions, based on previous work [5], were used to estimate user's sense of embodiment and each of its sub-concepts: sense of agency (Q1), sense of body ownership (Q2), sense of self-location (Q3) and also if they felt that they had more than one body (Q4).The following questions were made to assess the easiness of each of the tasks and fatigue.

In addition to the questionnaire, we conducted a semi-structured interview in order to capture the participants' perceptions about the accomplished tasks, clarify about their answers on the post-test questionnaire and get improvement suggestions.

3.7 Participants

For this test we chose 24 male participants. The ages of the users varied from 21 to 35 years. Regarding experience, the majority of the users had previous experience with 3D applications such as games and modelling systems. The majority of them had previous experience with Head-Mounted Displays (18 participants, or 75%) and 21 with previous experience with Kinect usage.

4 RESULTS

In this section, we present the main observations made during the tests as well as the difficulties and suggestions from users about the test task. To assess the difference between the three user embodied representation both in first and third-person perspective, we collected both objective and subjective data, in the form of logs and inquiries respectively, during the evaluation sessions. For the continuous variable, i.e. time, we used Shapiro-Wilk test to assess data normality. Since all samples were normally distributed, we used the Two-way Repeated Measures ANOVA for finding main effects between the two variables used, namely perspective and representation. When found main effects, additional One-way Repeated Measures ANOVA for multiple comparisons, and the Paired-Samples T-Test test between two samples, to find statistically significant differences. When comparing more than two samples, we applied the Bonferroni correction. Presented significance values are corrected.

In the following subsections we present the analysis made based on the results of the questionnaires and log files data obtained during the test.

4.1 Subjective Responses

As a result of the Two-way Repeated Measures ANOVA we found statistical significance between perspectives on Embodiment factors - Agency ($F(1,26)=7.499$, $p=0.011$), Body Ownership ($F(1,26)=4.489$, $p= 0.044$) and Self Location ($F(1,26) =9.755$, $p=0.004$). Statistical

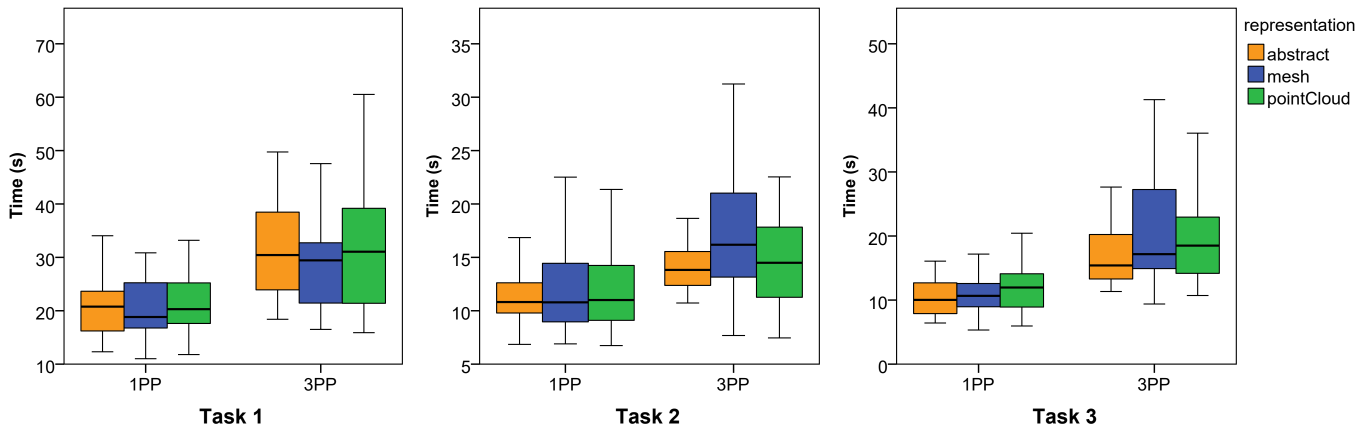

Figure 4: Performance time of Avatars in First-Person Perspective (1PP) and Third-Person Perspective (3PP) divided by task. median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers). Orange represents the Abstract avatar, Blue the Realistic Mesh Avatar and Green, the Point-Cloud Avatar.

Table 1: Summary of the questionnaires. Median and inter-quartile range for users score. (1 - totally disagree, 6 - totally agree)

	1PP			3PP		
	Abstract	Mesh	Point Cloud	Abstract	Mesh	Point Cloud
Q1: It felt like I was in control of the body I was seeing (Agency)	5(1)	5(2)	5(2)	5(2)	4(1)	5(2)
Q2; ...that the virtual body was my own (Body Ownership)	5(1)	5(2)	5(3)	4(3)	4(2)	5(2)
Q3:...as if my body was located where I saw the virtual body to be (Self-Location)	5(1)	5(2)	5(2)	4(4)	4(3)	5(3)
Q4: ... if I had more than one body	2(2)	2(2)	3(3)	3(4)	3(3)	3(3)
Q5: it was easy to walk in the virtual environment	5(2)	5(1)	5(2)	3(2)	4(2)	4(2)
Q6: it was easy to avoid obstacles in the virtual environment (Task 1)	5(2)	5(1)	5(1)	4(2)	4(2)	4(2)
Q7: it was easy to go over the obstacles in the virtual environment (Task 2)	5(2)	5(1)	5(1)	4(3)	4(2)	4(2)
Q8: it was easy to go under the obstacles in the virtual environment (Task 3)	6(1)	6(2)	5(2)	3(2)	4(2)	4(2)
Q9: I felt fatigue	2(3)	2(3)	2(2)	2(3)	2(3)	2(3)

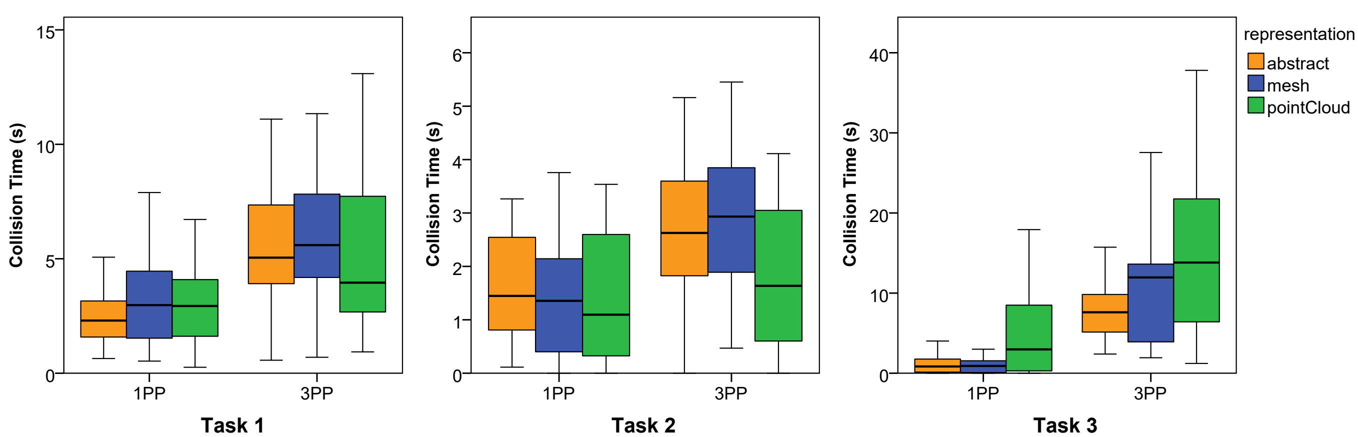

Figure 5: Total Collision time of Avatars in First-Person Perspective (1PP) and Third-Person Perspective (3PP) divided by task. median, first and third interquartile ranges (boxes) and 95% confidence interval (whiskers). Orange represents the Abstract avatar, Blue the Realistic Mesh Avatar and Green, the Point-Cloud Avatar.

significance was also verified for easiness of walking ($F(1,26)=17.827$, $p<0.001$), completing the Barrels Task ($F(1,26) = 0.549$, $p<0.001$), Bars Task ($F(1,26)=4.23$, $p=0.005$) and Tunnels task ($F(1,26)=65.768$, $p<0.001$).

We also found interaction between variables perspective and representation on sense of agency ($F(1.969,51.191)= 3.884$ $p=0.027$), sense of body-ownership ($F(1.558,52)=7.839$, $p=0.001$) and sense of self-location ($F(1.972,51.272)=4.889$, $p=0.011$). Also, the feeling of having two bodies ($F(1.668,51.683)=6.896$ $p=0.002$), easiness of walking ($F(1.971,51.234)=4.086$ $p=0.014$) and completing the tunnels task ($F(1.925,50.057)=12.826$ $p<0.001$). To further investigate this interaction, we made two different comparisons based on the data collected through the questionnaires, between representations on the same perspective and representations between perspectives.

4.1.1 Perspective. When comparing between representations in the First-Person Perspective, we found no statistical differences in any of the questions using a One-way ANOVA. The only two exceptions were found in Q4, the feeling of having more than one body ($F(1.909,49.622)=9.869$ $p<0.001$) and easiness of completing task 3 ($F(1.917,49.835)=3.503$ $p=0.04$) for the 1PP. Post-hoc paired t-tests showed that users felt as if they had more than one body with the Point-Cloud avatar when comparing with both the Abstract ($t(26)=-0.811$ $p<0.001$) and Mesh ($t(26)=0.004$ $p=0.012$) avatars.

When comparing between representations in the Third-Person Perspective, we found a higher number of statistically significant statements. By running the One-way ANOVA we found statistical significance on the 3PP in agency ($F(1.771,46.033)=5.25$ $p=0.008$), body-ownership ($F(1.934,45.531) = 9.314$ $p<0.001$), self-location ($F(2, 51.087)= 3.812$ $p=0.029$) and easiness of completing task 3 ($F(1.925, 49.835)=3.503$ $p=0.001$).

With the results of the post-hoc tests we noticed that users attributed a higher sense of embodiment, specifically on the sense of agency to the Point-Cloud Avatar when comparing to the Abstract Avatar ($t(26)= -2.595$, $p= 0.045$) and when comparing the Point-Cloud with the Mesh Avatar ($t(26)=-2.672$ $p=0.039$). Statistical significance was also found on Sense of Body-Ownership, with Abstract statistically worse than the Point-Cloud ($t(26)=-3.798$ $p=0.003$); a higher sense of self-location was sensed with the Point-cloud in comparison with the Abstract avatar ($t(26)=-2.55$ $p=0.017$). Regarding task 3, we found statistical significance in 3PP ($F(1.879,48.844)$ $p=0.001$), with users also finding easier to execute the Tunnel Task (Task 3) using the Abstract avatar when comparing to the Mesh ($t(26)=-3.349$ $p=0.006$) and Point-Cloud ($t(26)=-3.365$ $p=0.006$) avatars.

4.1.2 Representation. When comparing perspectives between the different representations we found overall better results with the 1PP on all representations. On the Abstract avatars, users felt a stronger sense of embodiment in the First-Person Perspective in all its components: agency ($t(26)=3.514$ $p=0.006$), body-ownership (q2) ($t(26)=3.776$ $p=0.003$) and self-location (q3) ($t(26)=2.848$ $p<0.001$). They also felt that they had more than one body in the 3PP ($t(26)=-2.926$ $p=0.021$), found it easier to walk in the VE ($t(26)=6.176$ $p<0.001$) and to perform the Barrels task($t(26)=5.827$ $p<0.001$) and Tunnels Task ($t(26)=7.963$ $p<0.001$).

Regarding Mesh Avatars, users felt a stronger sense of body-ownership ($t(26)=2.89$ $p=0.024$) and self-location ($t(26)=2.848$ $p=0.024$) with the First-person perspective and also less feeling of having two-bodies ($t(26)=-2.591$ $p=0.045$). Users also found it easier to walk with the 1PP ($t(26)=2.842$ $p=0.027$). About task easiness, they found it easier to avoid obstacles in the Barrels task ($t(26)=2.769$ $p=0.03$) and Tunnel task($t(26)=4.352$ $p<0.001$).

Lastly, on Point-cloud avatars, no difference was found regarding sense of embodiment and its sub-components. About task easiness, users only found it easier to perform the Tunnel task with the First-Person perspective ($t(26)=2.69$ $p=0.036$).

Table 2: Obstacles hit per task. Median number of obstacles hit (inter-quartile range).

	1PP			3PP		
	Abstract	Mesh	Point Cloud	Abstract	Mesh	Point Cloud
Task 1	4(2)	4(2)	4(2)	5(1)	5.5(2)	5(2)
Task 2	4(0)	4(1)	4(2)	4(1)	4(0)	4(2)
Task 3	5(7)	5(9)	10.5(12)	12(6)	12(8)	16(7)
Task 3 (Just Bars)	0(1)	0(2)	0.5(3)	2(8)	1(3)	7(16)

4.2 Task performance

In this subsection we present the analysis of results collected from users during the evaluation session. For assessing task performance of the users between the different representations we collected data through logs. We counted the time to assess the efficiency of the representation, the number of obstacles hit and the collision time to evaluate spatial awareness. Figures 4 and 5 show the total and collision time for each task in both perspectives and representations, respectively. The number of obstacles hit can be found on Table 2.

In the following sub-sections we present the results obtained for each of the metrics used (time, number of obstacles hit and collision time) for each of the sub-tasks.

4.2.1 Barrels Task.

Number of collisions: We found statistical significance on Barrels Task regarding number of objects collided on the Perspective factor($F(1,23)=24.636$ $p<0.001$, with the 1PP having a smaller number of objects hit. This behaviour was observed both with the Abstract ($t(23)=-2.497$ $p=0.06$) and Mesh ($t(23)=-3.657$ $p=0.009$) Avatars.

Collision time: When running a two-way repeated measures ANOVA, we found statistical significance on the perspective factor ($F(1,23)=26.592$ $p<0.001$), with better results on the 1PP in all representations (Abstract: $t(26)=-4.201$ $p<0.001$; Mesh:$t(23)=-4.35$ $p<0.001$; Point-cloud: $t(23)=-3.6$ $p=0.002$).

Completion time: We only found statistical significance in the perspective factor ($F(1,20)=76.686$ $p<0.001$), with the 1PP being more efficient than 3PP in all cases (Abstract: $t(21)=-7.818$ $p<0.001$; Mesh: $t(21)=-6.555$ $p<0.001$; Point-cloud: $t(23)=-6.336$ $p<0.001$).

4.2.2 Bars Task.

Number of collisions: We did not find any statistical significance for the number of collision in the Bars Task.

Collision time: A two-way ANOVA pointed statistical significance in both representation (F(1.879,41.34)= 3.456 p=0.04) and perspective (F(1,22)=17.574) p<0.001 factors, but with no interaction between variables. When grouping representations by perspective, we found statistical significance on the 3rd Person Perspective (F(1.655,38.072)=3.7 p=0.042, post-hoc tests indicated less collision time with the Point-Cloud representation (t(23)=3.022 p=0.018). Comparing the perspectives in each representation, we found better results in the 1PP in the Point-Cloud avatar (t(23)=-3.136 p=0.015).

Completion time: The First-Person Perspective was also the most efficient on this Task (F(1,16)=49.364 p<0.001), in all cases (Abstract: t(16)=-5.898 p<0.001; Mesh:t(22)=-4.260 p<0.001; Point-cloud: t(22)=-3.779 p=0.003).

4.2.3 Tunnel Task.

Number of collisions: For this task, we considered two possibilities: the number of objects collided and the number of horizontal bars collided.

Regarding number of objects collided, we found statistical significance on both representation (F(2,43.077)=7.832 p=0.001) and Perspective (F(1,23)=39.606 p<0.001). We also found statistical significance between representations on the First-Person Perspective (F(2,22)=7.150 p=0.004). Post-hoc test indicated that fewer objects collided with the Abstract in comparison with the Mesh Avatar (t(23)=-3.761 p=0.003).

When considering just the collision with the tunnels, we found statistical significance on representation (F(2,46)=15.858 p<0.001), perspective (F(1,23)=16.935 p<0.001) and interaction between factors (F(2,46)=4.591 p=0.015). Comparing between representations on both perspectives we found statistical significance on both 1PP (F(2,46)=6.124 p=0.012) and 3PP(2,46)=12.306 p<0.001. In the 1PP, users collided less with the tunnels using the Abstract avatar in comparison with the Point-Cloud avatar (t(23)=-2.802 p=0.03). In the 3PP, the Mesh had better results in comparison with both Abstract (t(23)=-2.890 p=0.024) and Point-Cloud Avatars (t(23)=-4.831 p<0.001). The comparison between the different representations in each perspective showed less collisions in the 3PP both on Abstract (t(23)=-3.744 p=0.003) and Point-cloud avatars.

Collision time: When running a two-way repeated measures ANOVA we found statistical significance in the perspective factor in favor of the 1PP (F(1,23)=36.841 p<0.001). Post-hoc tests indicated better results in all representations (Abstract: t(23)=-4.333 p<0.001; Mesh: t(23)=-3.858 p=0.003;t(23)=-4.871 p<0.001).

Completion time: The First-Person Perspective was also the most efficient on this Task (F(1,16)=49.364 p<0.001), in all cases (Abstract: t(23)=-4.856 p<0.001; Mesh:t(23)=-6.305 p<0.001; Point-cloud: t(23)=-7.563 p<0.001).

5 DISCUSSION

From an overall analysis of the results, we can verify that the first person perspective was found to be more suited for navigation tasks. The performance results were significantly better for all representations (time, collision, collision time), showing that it not only allowed users to perform the tasks faster, but with higher precision. Users felt a higher sense of embodiment in the abstract and mesh representations, except for sense of agency in the latter. The only exception was the point cloud representation where no difference in embodiment was found between perspectives. This can be explained by the fact that this representation uses a reconstruction of users' bodies inside the virtual environment, increasing the sense of embodiment. Additionally, 1PP is indeed a more natural point of view to which they are used to. Self-location in the 3PP was found to be significantly harder.

However, this confirmation does not correspond to earlier work. Debarba et al. [5] have reported a similar sense of embodiment in the 3PP when compared to the 1PP. In their work, the majority of the interaction time is limited to reaching tasks. The navigation phase of the test is limited to reach the test area. In the remaining time, the majority of users' bodies remain stationary. This may explain the difference in embodiment factors. In our study, users stay in movement most of the time, so the relationship between people and the environment is always in motion. We verified that this aspect affects embodiment and all of its components, particularly in the sense of self-location.

Earlier work also suggested an improvement in spatial awareness with the use of third-person avatars [9, 18, 24]. Some differences in these papers explain the different results from our work. In both Gorisse et al. [9] and Monteiro et al. [18] the third-person was used to expand user's view and be able to see further parts of the virtual environment. For example, in Monteiro et al. [18] the task consisted in controlling a vehicle and in the Third-Person perspective people were able to see further details in the road and respond faster when further actions such as turning were needed. Gorisse et al. [9] on the other hand, reported an improvement in users' spatial awareness when reacting to objects being thrown at them, something that was already confirmed by Salamin et al. [24]. For the navigation phase, the study [9] used solely subjective metrics to assess the spatial presence. Another point to be considered is that the obstacles presented in the virtual environment were only used to affect the performance. One example can be seen in the video provided where users were seen stepping out of the limits of the virtual environment, something that was not analyzed in that study. In our case, our study focused on the relations between users and the virtual environment. This is highly influenced by how an user makes distance (or spatial) judgments, which indeed affects the feeling of spatial awareness. This difference can clearly be seen in Figure 6, where the paths taken in the 1PP for the first task are more fluid to the expected trajectory to avoid the proposed obstacles.

As seen, the use of an avatar in the first-person perspective is indeed the most efficient and effective in navigation tasks. The only exception was found when the Point-Cloud avatar was used, where they felt an equivalent sense of embodiment with both 3PP and 1PP. This equivalence was also found in spatial awareness factors (collision number and collision time). This representation had a smaller collision time in the second task when compared to its alternatives in the 3PP. By analyzing users' path we also noticed a more fluid path with the Point-cloud avatar in the 3PP when compared to the other representations, which indicates an improved perception of their surroundings and how they make spatial judgments (Figure 6). Also, when comparing between the representations in the 3PP we also found a higher feeling of embodiment in all of

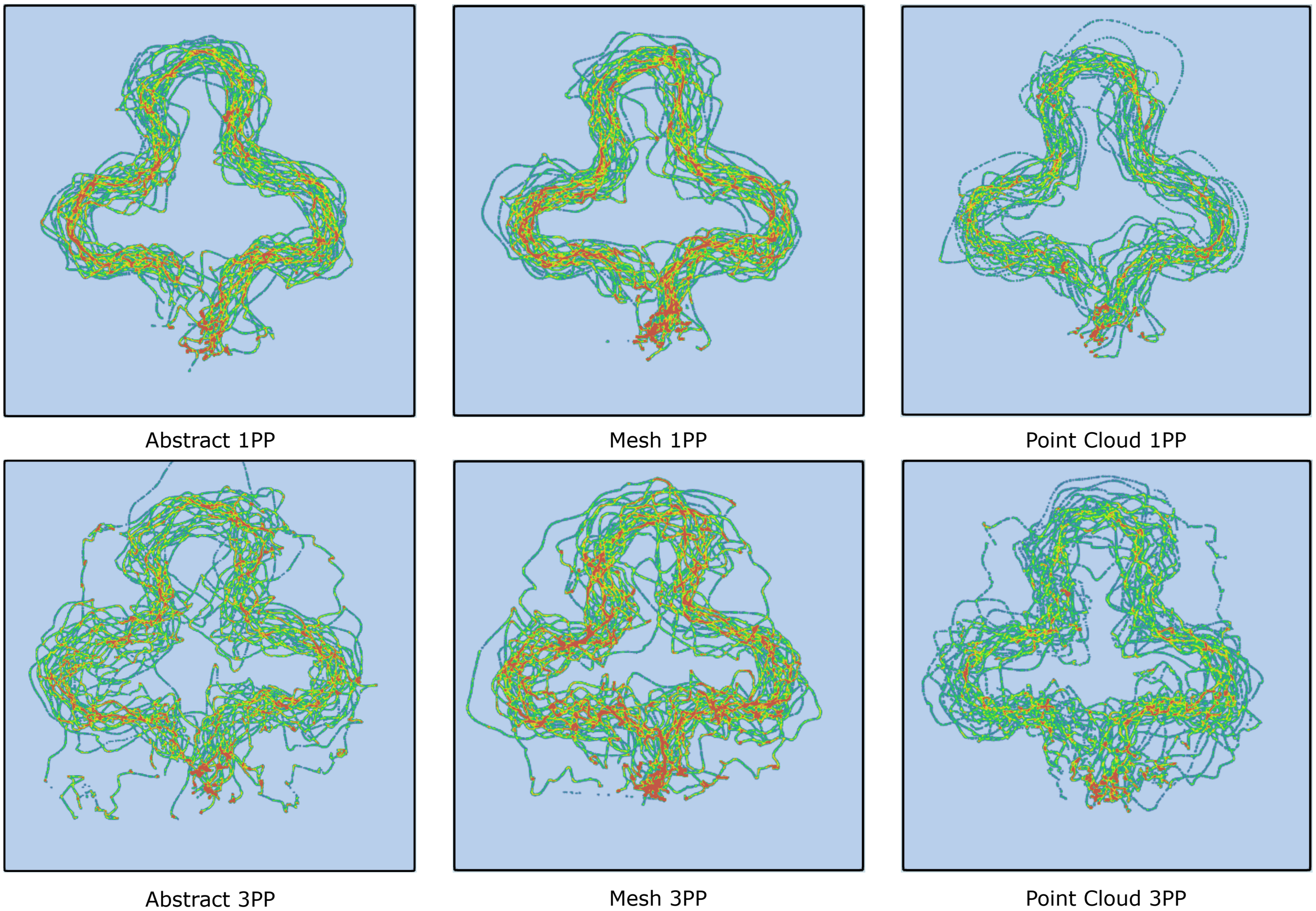

Figure 6: Heatmaps representing of users' paths for Task 1 separated by representation and perspective.

three sub-components of the sense of embodiment: agency, body-ownership and self-location. This corroborates our hypothesis that the realism of the representation impacts mostly the 3PP, since the avatar's body is always seen in this perspective.

However, some particularities were found in Task 3, regarding spatial awareness and task easiness. On the Tunnels task, we noticed a higher amount of collided time and objects collided with the Point-cloud (average=10.46s) in comparison with the abstract (average=5s) and mesh (avg=7.67s). This may indicate a higher effect on the perceived distance compression provided by head-mounted display, related to users' height. With a highly detailed avatar, people tend to make distance judgments in a similar way to how they make in real-life and due to the perceived compressed space, make more errors in the vertical axis.

6 CONCLUSIONS AND FUTURE WORK

In immersive VR, Head-Mounted Displays occlude the user's self, causing a decrease in the feeling of presence by people and a degraded user experience. Self-embodied avatars can help overcome this problem by improving presence and overall distance estimation by people in VR setups. In the present work we study factors that influence the sense of embodiment including both realism and the perspective from which the avatar is viewed (either First or Third-Person). To this end, we used three different representations using varying degrees of realism in both perspectives, ranging from Abstract to a Realistic Point-Cloud Representation. To assess each representation-perspective combination, we conducted a comprehensive user study featuring a real-walking navigation task while avoiding obstacles.

From a detailed statistical analysis, discussion and evaluation of the results, we propose specific guidelines regarding body representation and camera perspective for developing embodied virtual reality applications. Among the most salient findings, we can say that the realism of the representations with a 1PP avatar did not seem to interfere with both efficiency and spatial awareness of the user. Although, the increase of graphical fidelity avatar highly affects these factors in 3PP avatars. We also noticed an inverse relation between the distance estimation on the vertical axis and the graphical fidelity in 1PP avatars.

The sense of embodiment is seen to be influenced by the perspective in which the avatar is viewed, except for the Point-Cloud representation, which maintained a high level of embodiment in both situations. A more realistic representation can also improve the sense of awareness when using a Third-Person embodied avatar. But still, the 1PP remains the most efficient and effective perspective for navigation tasks. This result differs from previous research that states that 3PP improves spatial awareness for navigation tasks. This is only true when the enhanced field of view of a 3PP increases the chance of success of the users. In navigation tasks where the obstacles are visible, 1PP is still the best suited perspective.

Since embodiment factors are closely related to the task being performed, a similar study needs to be performed for different use-case scenarios. Additionally, we plan to assess how different conditions should be considered for the same combinations of perspectives and realism in avatars, e.g. in collaborative settings, and social environments where communicative tasks engage different users to accomplish success.

ACKNOWLEDGMENTS

This work was partially supported by: Fundação para a Ciência e a Tecnologia (FCT), through grants IT-MEDEX PTDC/EEISII/6038/2014, UID/CEC/50021/2013 and SFRH/BD/91372/2012, CAPES-Brasil Foundation, Ministry of Education of Brazil through the scholarship grant Ref. 9040/13-7 and the European Research Council under the project Ref. 336200.

REFERENCES

[1] Matthew Botvinick, Jonathan Cohen, et al. 1998. Rubber hands' feel'touch that eyes see. *Nature* 391, 6669 (1998), 756–756.

[2] Ronan Boulic, Damien Maupu, and Daniel Thalmann. 2009. On scaling strategies for the full-body postural control of virtual mannequins. *Interacting with Computers* 21, 1 (2009), 11–25.

[3] Pierre Bourdin, Itxaso Barberia, Ramon Oliva, and Mel Slater. 2017. A virtual out-of-body experience reduces fear of death. *PloS one* 12, 1 (2017), e0169343.

[4] Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. 2017. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one* 12, 12 (2017), e0190109.

[5] Henrique G Debarba, Eray Molla, Bruno Herbelin, and Ronan Boulic. 2015. Characterizing embodied interaction in First and Third Person Perspective viewpoints. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on.* IEEE, 67–72.

[6] Alena Denisova and Paul Cairns. 2015. First person vs. third person perspective in digital games: do player preferences affect immersion?. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 145–148.

[7] Mark Draper. 1995. *Exploring the influence of a virtual body on spatial awareness.* Master's thesis. University of Washington.

[8] H Henrik Ehrsson. 2007. The experimental induction of out-of-body experiences. *Science* 317, 5841 (2007), 1048–1048.

[9] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. 2017. First-and Third-Person Perspectives in immersive Virtual environments: Presence and Performance analysis of embodied Users. *Frontiers in Robotics and AI* 4 (2017), 33.

[10] Victoria Interrante, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *IEEE Virtual Reality Conference (VR 2006).* IEEE, 3–10.

[11] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387.

[12] Thomas Kosch, Robin Boldt, Matthias Hoppe, Pascal Knierim, and Markus Funk. 2016. Exploring the Optimal Point of View in Third Person Out-of-Body Experiences. In *Proccedings of 10th International Conference on Pervasive Technologies Related to Assistive Environments (PETRA 2016).* ACM.

[13] Bigna Lenggenhager, Michael Mouthon, and Olaf Blanke. 2009. Spatial aspects of bodily self-consciousness. *Consciousness and cognition* 18, 1 (2009), 110–117.

[14] J.-L. Lugrin, J. Latt, and M. E. Latoschik. 2015. Anthropomorphism and Illusion of Virtual Body Ownership. In *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments (ICAT - EGVE '15).* Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 1–8. https://doi.org/10.2312/egve.20151303

[15] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Front Hum Neurosci* 7 (03 2013). https://doi.org/10.3389/fnhum.2013.00083

[16] Erin A McManus, Bobby Bodenheimer, Stephan Streuber, Stephan De La Rosa, Heinrich H Bülthoff, and Betty J Mohler. 2011. The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on applied perception in graphics and visualization.* ACM, 37–44.

[17] Betty J Mohler, Sarah H Creem-Regehr, William B Thompson, and Heinrich H Bülthoff. 2010. The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators and Virtual Environments* 19, 3 (2010), 230–242.

[18] Diego Monteiro, Hai-Ning Liang, Wenge Xu, Marvin Brucker, Vijayakumar Nanjappan, and Yong Yue. [n. d.]. Evaluating enjoyment, presence, and emulator sickness in VR games based on first-and third-person viewing perspectives. *Computer Animation and Virtual Worlds* ([n. d.]), e1830.

[19] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[20] Valeria Ivanova Petkova, Mehrnoush Khoshnevis, and H Henrik Ehrsson. 2011. The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Frontiers in psychology* 2 (2011), 35.

[21] Lukasz Piwek, Lawrie S McKay, and Frank E Pollick. 2014. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition* 130, 3 (2014), 271–277.

[22] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. 2013. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 23.

[23] Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. 2008. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology.* ACM, 167–170.

[24] Patrick Salamin, Tej Tadi, Olaf Blanke, Frederic Vexo, and Daniel Thalmann. 2010. Quantifying effects of exposure to the third and first-person perspectives in virtual-reality-based training. *IEEE Transactions on Learning Technologies* 3, 3 (2010), 272–276.

[25] Patrick Salamin, Daniel Thalmann, and Frédéric Vexo. 2006. The benefits of third-person perspective in virtual and augmented reality?. In *Proceedings of the ACM symposium on Virtual reality software and technology.* ACM, 27–30.

[26] Patrick Salamin, Daniel Thalmann, and Frédéric Vexo. 2008. Improved Third-Person Perspective: a solution reducing occlusion of the 3PP?. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry.* ACM, 30.

[27] Maria V Sanchez-Vives, Bernhard Spanlang, Antonio Frisoli, Massimo Bergamasco, and Mel Slater. 2010. Virtual hand illusion induced by visuomotor correlations. *PloS one* 5, 4 (2010), e10381.

[28] Ulrike Schultze. 2010. Embodiment and presence in virtual worlds: a review. *Journal of Information Technology* 25, 4 (2010), 434–449.

[29] Ellen L Schuurink and Alexander Toet. 2010. Effects of third person perspective on affective appraisal and engagement: Findings from SECOND LIFE. *Simulation & Gaming* 41, 5 (2010), 724–742.

[30] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. 2008. Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience* 2 (2008), 6.

[31] Mel Slater, Bernhard Spanlang, and David Corominas. 2010. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 92.

[32] Mel Slater and Martin Usoh. 1994. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality* 1 (1994), 125–148.

[33] Maurício Sousa, Daniel Mendes, Rafael Kuffner Dos Anjos, Daniel Medeiros, Alfredo Ferreira, Alberto Raposo, João Madeiras Pereira, and Joaquim Jorge. 2017. Creepy Tracker Toolkit for Context-aware Interfaces. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17).* ACM, New York, NY, USA, 191–200. https://doi.org/10.1145/3132272.3134113

[34] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *2010 IEEE Virtual Reality Conference (VR).* IEEE, 95–102.