Animation Fidelity in Self-Avatars: Impact on User Performance and Sense of Agency

Haoran Yun* Jose Luis Ponton† Carlos Andujar‡ Nuria Pelechano§

Universitat Politècnica de Catalunya, Spain

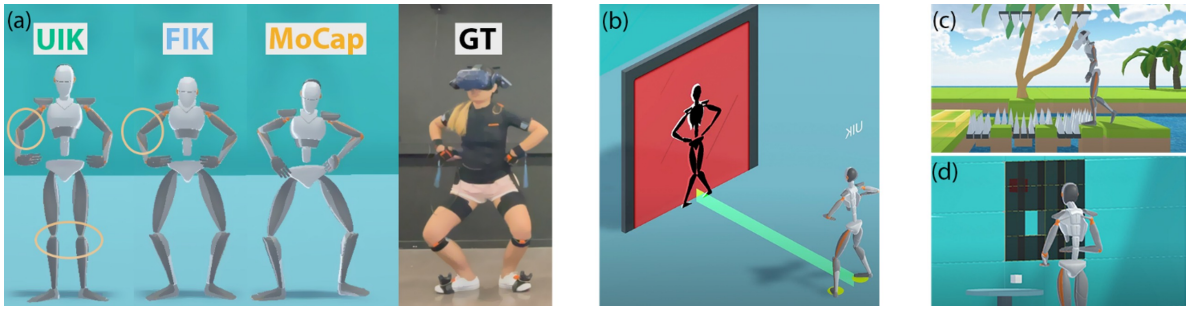

Figure 1: We evaluate self-avatars with three different types of animation fidelity as shown in (a): Unity IK, Final IK and Xsens IMU-based MoCap. We compare them in three tasks involving varied body parts: (b) Copy-pose task, (c) Step-over-spikes task, and (d) Pick-and-place task.

Abstract

The use of self-avatars is gaining popularity thanks to affordable VR headsets. Unfortunately, mainstream VR devices often use a small number of trackers and provide low-accuracy animations. Previous studies have shown that the Sense of Embodiment, and in particular the Sense of Agency, depends on the extent to which the avatar's movements mimic the user's movements. However, few works study such effect for tasks requiring a precise interaction with the environment, i.e., tasks that require accurate manipulation, precise foot stepping, or correct body poses. In these cases, users are likely to notice inconsistencies between their self-avatars and their actual pose. In this paper, we study the impact of the animation fidelity of the user avatar on a variety of tasks that focus on arm movement, leg movement and body posture. We compare three different animation techniques: two of them using Inverse Kinematics to reconstruct the pose from sparse input (6 trackers), and a third one using a professional motion capture system with 17 inertial sensors. We evaluate these animation techniques both quantitatively (completion time, unintentional collisions, pose accuracy) and qualitatively (Sense of Embodiment). Our results show that the animation quality affects the Sense of Embodiment. Inertial-based MoCap performs significantly better in mimicking body poses. Surprisingly, IK-based solutions using fewer sensors outperformed MoCap in tasks requiring accurate positioning, which we attribute to the higher latency and the positional drift that causes errors at the end-effectors, which are more noticeable in contact areas such as the feet.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality,Perception; Computing methodologies—Computer graphics—Animation—Motion capture

*e-mail: haoran.yun@upc.edu
†e-mail:jose.luis.ponton@upc.edu
‡e-mail:carlos.andujar@upc.edu
§e-mail:nuria.pelechano@upc.edu

1 Introduction

Virtual reality headsets allow us to immerse ourselves in highly realistic digital worlds. A fundamental aspect of feeling present in these virtual environments is to have a virtual representation of our own body, known as the user's self-avatar. Ideally, avatars should be animated in a way that allows users to achieve natural behaviors and interactions in the virtual environment, as well as to use non-verbal communication with others. Self-avatar is the virtual representation of oneself and should be distinguished from other people's avatars as they have different requirements. Aspects such as latency or end-effectors and pose accuracy are more crucial for perceiving one's own avatar than for others.

Unfortunately, the limited number of trackers in consumer-grade devices severely restricts the quality of the self-avatar's movements. Most applications limit the representation to floating upper bodies (no legs) with floating hands/globes/tools, sometimes with the arms animated with Inverse Kinematics (IK), by using the tracking data from the HMD and the hand-held controllers. Only a few applications offer a full-body representation. However, due to the lack of trackers, legs are typically animated by playing cyclic time-warped animations. With these solutions, users may notice inconsistencies between their movements perceived via proprioception and those of the self-avatar.

Previous work has demonstrated the importance of having a self-avatar that moves in sync with the user [9, 10, 33]. If we focus on the overall movement without further interaction with the virtual world, current animation techniques based on IK from sparse tracking data may suffice. However, if accurate body poses and positioning of end-effectors matter, then artifacts that affect user performance and the Sense of Agency may pop up. For example, consider the task of building some assembly by holding pieces and putting them in specific locations. In that case, hand-eye coordination is crucial, as is the accuracy of the overall pose, to prevent parts of the arm/body from colliding with other pieces. Another example is moving through a room full of obstacles, where accurate foot positioning is also crucial. Finally, correct body poses also matter in the case of learning to dance or practicing yoga by mimicking an instructor [9].

Given that high-quality motion capture is difficult to achieve with sparse input data, we are interested in studying how animation

2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 979-8-3503-4815-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/VR55154.2023.00044

2642-5254/23/$31.00 ©2023 IEEE
DOI 10.1109/VR55154.2023.00044

fidelity affects user performance and embodiment. By animation fidelity, we refer to the quality of the animations in terms of accurately following the user poses as well as the correct absolute positioning of the body parts. More specifically, we evaluate interactions with the virtual world that need pose and/or positional accuracy. We evaluate embodiment with a perceptual study, in which our main focus is on the Sense of Agency due to its relationship with animation fidelity. Furthermore, we study the effect of the quality of the interaction with the virtual world on user performance by measuring completion time and unintentional collisions. We focus on two popular methods based on Inverse Kinematics from sparse input data (6 trackers): UnityIK[1] and FinalIK[2], and one motion capture system based on a large number (17) of inertial sensors: Xsens Awinda[3].

Our results suggest that animation fidelity affects the Sense of Embodiment and user performance. We found that a straightforward IK solution, such as Unity IK, decreases the Sense of Embodiment when compared to high-quality IK and MoCap solutions. However, when interacting with the environment, having lower latency and minimal end-effector positional error may be more important than synthesizing high-quality poses suffering from positional drift.

The main contributions of this paper are:

- To the best of our knowledge, this is the first study to compare an IMU-based full-body motion capture system to IK solutions for animating self-avatars in VR during tasks that require accurate positioning of end-effectors and body postures.
- We study the relationship between animation fidelity on user task performance and the Sense of Agency to improve future research on VR avatar animation from sparse data.

2 Related Work

2.1 Self-avatars and animation fidelity

A self-avatar is a virtual representation of one's own body from a first-person view of the virtual environment (VE). Previous studies have shown that full-body self-avatars are beneficial in various tasks, such as egocentric distance estimation, spatial reasoning tasks, and collision avoidance [22, 23, 28]. For instance, compared to not having an avatar, users with a full-body realistic avatar collide less frequently with the VE [23]. Similarly, Ogawa et al. [20] demonstrated that users would be less likely to walk through the virtual walls if equipped with a full-body representation compared to a hands-only representation. In social settings, using full-body self-avatars would enhance social presence and communication efficiency [1, 41].

Animation fidelity is a crucial component of self-avatars. Unlike visual fidelity, which addresses the appearance of avatars and has been extensively studied [4, 11, 12, 20], animation fidelity focuses on how accurately and synchronized the self-avatar mimics users' real-world movements. We could use avatars with the highest visual fidelity (with a realistic full-body self-avatar), but low animation fidelity if the body poses are not well mimicked, not in sync with the user, or have errors in the positioning of end-effectors. These avatars are unlikely to induce the user's feeling of owning or being in control of the virtual body [17, 36]. These kinds of problems may be introduced by the tracking system or by the methods used to capture and animate the self-avatar.

Inverse Kinematic (IK) solvers can be used with sparse input from VR devices to calculate joint angles of the articulated human model. Some frameworks are available to animate full-body avatars from six trackers (HMD, two hand-held controllers and three Vive trackers) [21, 26, 42]. Parger et al. [24] proposed an intuitive IK solution for self-avatar's upper-body animation with one HMD and two controllers. Their IK solver outperformed an optical MoCap system in terms of lower latency and accurate pose reconstruction. The reduced Jacobian IK solver proposed by Caserman et al. [3] can smoothly and rapidly animate full-body self-avatars with HTC Vive trackers.

Recently, researchers have shown an increasing interest in data-driven methods to reconstruct full-body animation for avatars from VR devices. For instance, Winker et al. [37] proposed a reinforcement learning framework with physical-based simulation to achieve real-time full-body animation. Ponton et al. [27] combined body orientation prediction, motion matching and IK to synthesize plausible full-body motion with accurate hand placement. Jiang et al. [14] used a transformer model to estimate the full-body motion. Other researchers have looked at using a sparse set of wearable IMUs to estimate full-body motion. These methods could be integrated into self-avatars in VR because of the built-in IMUs on VR headsets, controllers and trackers. For example, Huang et al. [13] used a bi-directional RNN to reconstruct a full-body human pose from six wearable IMUs attached to the head, arms, pelvis, and knees. Yi et al. [40] took the same input, but generated both accurate pose and precise global translation. More recently, Jiang et al. [15] not only accurately estimated the full-body motion but also handled the joint and global position drift that most IMU systems suffer from.

While there is an extensive body of research proposing new animation methods to improve animation fidelity for avatars, little interest has been given to how the animation fidelity of self-avatars impacts user performance, perception and behavior in a VE. Fribourg et al. [9] showed that users preferred to improve animation features when asked to choose among appearance, control (animation) and point of view, to improve the Sense of Embodiment (SoE). In their work, participants preferred mocap based on Xsens over FinalIK. However, their input to the IK system was the joints positions from the Mocap system, and thus the problems with incorrect end-effector positioning and latency were carried on to the IK condition.

Galvan et al. [10] adapted the same methodology to examine the effect of animation fidelity of different body parts. Participants were first exposed to optimal animation fidelity (53-marker optical motion capture). Then, they started with minimal animation fidelity and repeatedly chose one body part to improve until they felt the same level of the SoE as with the optimal configuration. They found users felt the same level of the SoE with an IK solution with eight trackers than with the 53-marker optical motion capture system. Their work also found that the unnatural animation of the full body caused disturbing feelings for users when separately animating the upper body and lower body with different fidelity. Thus, our work focuses on full-body animation instead of body parts to avoid breaking the user's presence. Eubanks et al. [8] explored the impact of the tracking fidelity (number of trackers) on a full-body avatar animated by an IK solver. They found that a high number of trackers could improve the SoE. However, animation fidelity is not only about tracking fidelity, but also about the animation techniques underneath. Our study thus compares not only systems with different numbers of trackers, but also different animation techniques: IK and IMU-based motion capture.

2.2 Sense of Agency

The Sense of Agency (SoA) has been characterized in various ways in different contexts because of its interdisciplinarity property. From the point of view of psychology, the SoA refers to the feeling that *I am the one causing or generating an action* [6]. In the field of VR, the SoA is the feeling of being the agent who conducts the motions of an avatar. It results from synchronizing one's real-world movements with virtual body motions.

The Sense of Agency is a crucial subcomponent of the Sense of Embodiment. According to Kilteni et al. [16], the SoE consists of three subcomponents: the Sense of Agency, the Sense of Self-Location (SoSL), and the Sense of Body Ownership (SoBO). Many

[1] https://docs.unity3d.com/Manual/InverseKinematics.html

[2] http://root-motion.com/

[3] https://www.xsens.com/products/mtw-awinda

studies have studied the impact of single or multiple factors, including avatars' appearance, visibility and tracking fidelity, on the SoE. Fribourg et al. [9] explored the relative contributions of the control factor (i.e. animation fidelity), appearance and point of view that contribute to the SoE. Results showed that control and the point of view were preferred when people had to choose among the three factors to improve their SoE. Recent studies showed that low-quality tracking, which directly impacts the animation of self-avatar, can decrease the embodiment [8, 33]. These findings analyzed the effect of the SoE, which is directly or implicitly related to animation. However, there is still a gap in how the animation fidelity directly impacts the SoE, specifically the subcomponent SoA.

The synchronicity of visuomotor correlation can induce the SoA, while discrepancies can decrease it. Kollias et al. [19] simulated different kinds of motion artifacts that may occur in a real-time motion capture system. They examined the effect of these artifacts on the SoE, specifically on the SoA. Results showed that the artifacts negatively affected the SoA, but not the SoBO.

Studies regarding the SoA mainly focus on subjective perception with questionnaires and objective brain activity measurements such as fMRI and EEG. As suggested by Kilteni et al. [16], the motor performance of VR users should be positively correlated with the SoA, under the assumption that a fine-controlled virtual body performs motor tasks more successfully. Therefore, the users' motor performance in VR could be used to measure the SoA objectively. Our study measured task performance in terms of unintentional collisions between the self-avatar and the virtual obstacles. We believe that the number of collisions and their duration could bring insights into human motor performance in 3D space. High animation fidelity means precise control of self-avatars which can perform better in motor tasks. Therefore, we expected to observe the impact of animation fidelity on collisions, completion time, and the correctness of the body poses.

3 Animation Fidelity Study

This study aims to assess the importance of animation fidelity on the users' performance and the SoE when performing a set of tasks that require careful positioning and/or accurate poses. We want to study the importance of the virtual body correctly mimicking the user's movements as well as the impact of accurate end-effector positioning.

3.1 Experimental conditions

In this study, we adopted a within-subject experiment design with one independent variable: the animation fidelity for the virtual avatar. We designed three conditions for the animation fidelity variable: Unity Inverse Kinematics (UIK), FinalIK (FIK) and motion capture with Xsens (MoCap). These techniques provide different levels of animation quality in terms of end-effector positioning (more accurate in UIK and FIK since hand-held controllers and trackers provide accurate absolute positioning), pose angles (more accurate in MoCap thanks to a larger number of sensors), and latency (higher for MoCap). The first two conditions differ on the IK solvers, while both use sparse tracking data from consumer-level VR devices. The last condition, MoCap, uses tracking data from a professional motion capture system with 17 IMU sensors. Fig. 2 illustrates the equipment used for tracking in the three conditions. The three conditions used have been implemented as follows (see accompanying video):

UIK: This condition uses Unity 2020.3 game engine's built-in IK solver for animating the avatar's limbs (2-segment kinematic chains). It is important to note that it does not consider the full-body pose when solving the IK. Instead, it independently computes each limb's joints based on one target end-effector. To further improve the overall body pose, forward kinematics (FK) is included to animate two joints: head and spine, so that the self-avatar can lean forwards and sideways. IK and FK together generate a full-body animation

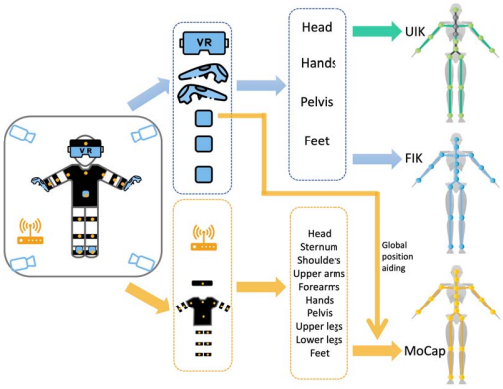

Figure 2: Equipment and conditions. For the experiment, participants were simultaneously equipped with two sets of tracking devices: VR devices (HMD, controllers and three trackers); and the trackers from the Xsens Awinda mocap system. The tracked body parts are shown in the figure. Different IK solvers were applied to animate the avatar using the VR tracking devices.

for the avatar from the tracking data in the HMD, the hand-held controllers and three additional trackers located on the pelvis and the feet.

FIK: This condition uses the VRIK solver from RootMotion's FinalIK package, which combines analytic and heuristic IK solvers for generating the full-body avatar animation. With the same input, FIK produces higher-quality results than UIK because each limb is not solved independently from one end-effector, but rather from an analysis on the user pose from several end-effectors [35]. For instance, the spine is solved considering the position of the HMD and two hand-held controllers, and the elbows use the position of the hands relative to the chest joint to determine the orientation. The only exception are the legs, which are solved independently but using a 3-joint dual-pass trigonometric solver (first solve the knee and then the ankle).

MoCap: The Xsens Awinda system receives acceleration, angular velocity and magnetic field data from 17 body-worn IMUs, processes the data with Strap-down Integration and Kalman filtering, and then outputs the rotations of the joints of the avatar, which are streamed to Unity via UDP; these processing steps increase the latency with respect to the previous conditions. IMUs suffer from a positional drift over time, that might break the Sense of Self-location. To enforce the correct location of the avatar with the user, we use the pelvis tracker to position the avatar in the VE. However, this does not guarantee accurate positioning of the end-effectors and can suffer from foot sliding. Foot lock is applied to reduce the foot sliding of the Xsens pose when the foot touches the floor. Once the foot is locked, we store the position of the HTC tracker, which we will use as a reference to detect whether the user is moving the foot. In the following frames, if the distance between the current HTC tracker and its initial position is larger than a relatively small threshold (1 cm), we unlock the foot; otherwise, it will noticeably modify the leg pose. Note that we are locking the foot at the position given by Xsens (thus, it may contain positional error); we only use the HTC tracker to detect whether the user's foot remains on the ground.

Each participant performed the same three tasks for each condition. Conditions were counterbalanced between participants using a Balanced Latin Square, which ensures each condition precedes and follows every other condition an equal number of times [7].

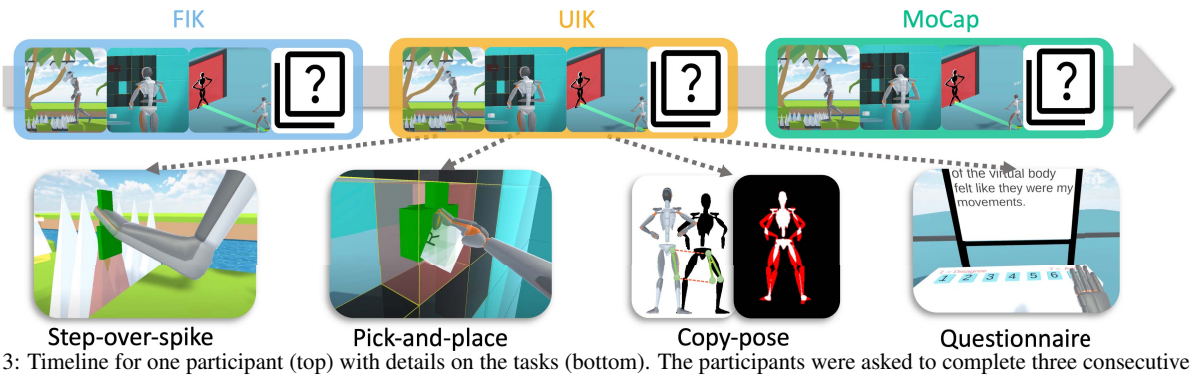

Figure 3: Timeline for one participant (top) with details on the tasks (bottom). The participants were asked to complete three consecutive tasks followed by the questionnaire for the first condition, and then repeat the procedure with the other two conditions. During the first two tasks, the volume of the small colliders were recorded (green), and users had visual feedback from the obstacles (in red) every time a collision occurred. For the copy-pose task, the pose-related metrics were calculated. Questions and buttons for answering were displayed on a whiteboard in VR.

3.2 Tasks

Prior studies have shown that the type of actions users perform in a VE influences users' perception [9, 34]. For instance, when picking up nearby objects, people would pay more attention to the upper body while their ignoring surroundings [5]. Similarly, walking in a room with obstacles on the floor would draw people's attention to objects and lower body parts to plan the future path and avoid collisions [32]. We thus designed three tasks that cover a wide range of common actions in VR games and applications, while each task focused on a different interaction pattern between the virtual body and the VE (see Fig. 3 and accompanying video).

- *Step-over-spikes task* focuses on direct interaction between the lower body and the VE. It consists of walking and lifting the knees to avoid colliding with spike-like obstacles while stepping on small platforms.
- *Pick-and-place task* focuses on direct interaction between the upper body and the VE. It consists of picking up objects and then placing them at specific target locations while avoiding collisions between the arm and nearby obstacles.
- *Copy-pose task* involves only non-direct interactions between the virtual body and the VE. More specifically, we focus on the overall pose of the self-avatar without caring about the exact global positioning of the avatar. For this task, we show a 2D projection of an avatar in a certain pose on a virtual screen, and then the user is asked to mimic the pose as accurately as possible. The design is inspired by OhShape[4].

One task block consisted of the following three sequential tasks, which were presented in the following order: (1) step-over-spikes task; (2) pick-and-place task; (3) copy-pose task. Each task consisted of ten trials separated by five seconds of rest. We decided to use this task order to guarantee that the last task before each SoE questionnaire (see below) equally involved the upper and lower limbs. Participants were requested to complete the entire task block for each of the three conditions on a recurrent basis.

3.3 Apparatus

The experiments were conducted in an acoustically-isolated laboratory room with a 3 m x 6 m space. The VE was developed with Unity 2020.3 LTS and run on a PC equipped with a CPU Intel Core i7-10700K, a GPU Nvidia GeForce RTX 3070 and 32 GB of RAM. We used an HTC Vive Pro HMD with 1440 × 1600 pixels per eye, 110° field of view and 90 Hz refresh rate. Three 6-DoF HTC Vive trackers 3.0 were used for tracking the pelvis and feet. Two HTC Vive controllers were held in both hands of the participants. We installed four SteamVR Base Station 2.0 in each corner of the room to minimize line-of-sight occlusions.

We employed the well-established frame counting approach [3, 31] to determine the latency of the tracking system and the animation techniques used in our experiment. One person was equipped with all tracking devices and repeatedly moved one arm up and down. We used a high-speed 240fps camera to record both the person and a monitor showing the VE. The mean latency from the physical controller to the animated virtual hand was 32 ms for UIK and 33 ms for FIK. These latencies include the SteamVR tracking system, IK solver computation and rendering. For MoCap, the mean latency was 91 ms, which was notably higher than the other two conditions. The MoCap latency includes the IMU-to-Xsens software latency (∼30 ms [5]) [25], motion processing in Xsens software, network communication, motion data unpacking in Unity (∼5 ms), and rendering.

3.4 Procedure

A total of 26 participants took part in the experiment (22 male, 4 female, aged 19-40, M = 22.4, SD = 5.5) but one participant's data was discarded due to a calibration failure.

Upon arriving at the experiment room, participants were instructed to read the information sheet and complete the consent form and a demographic survey detailing their age, gaming and VR experience. We introduced their body measurements to the Xsens software in order to obtain a scaled avatar matching the user's dimensions. Then we placed the 17 wireless trackers on the body of the participant, with the help of a t-shirt and a set of straps. Participants were asked to walk a few meters to calibrate the IMU-based motion capture suit. The calibration was repeated until the Xsens software (MVN Animate Pro) rated it as a "Good" (among "Good", "Acceptable" and "Poor"). The experimenter also validated visually that the subject's pose closely matched that of the avatar. Next, participants were equipped with an HTC Vive HMD, two hand-held controllers and three Vive trackers placed on the pelvis and both feet. They were asked to stand in a T-pose to complete the calibration of the HTC trackers for the IK solvers. During the experiment, participants were equipped with both tracking systems at all times. This ensured that they could not guess what system was being used for each condition. Before each task, participants watched a tutorial video (two minutes in total) that demonstrated how to perform the task.

3.5 Measurements

The step-over-spikes task challenges the participants' lower-body motion so that we can quantitatively assess the effect of animation

[4] https://ohshapevr.com/

[5] https://base.xsens.com/s/article/MVN-Hardware-Overview

Agency *- Scoring: (AG1 + AG2 + AG3 + AG4 + AG5 + AG6 + AG7) / 7*
AG1 The movements of the virtual body felt like they were my movements.
AG2 I felt the virtual arms moved as my own arms.
AG3 I felt the virtual elbows were in the same position as my own elbows.
AG4 I felt the virtual hands were in the same position as my own hands.
AG5 I felt the virtual legs moved as my own legs.
AG6 I felt the virtual knees were in the same position as my own knees.
AG7 I found it easy to control the virtual body pose to complete the exercises.
Ownership *- Scoring: (OW1 + OW2 + OW3) / 3*
OW1 It felt like the virtual body was my body.
OW2 It felt like the virtual body parts were my body parts.
OW3 It felt like the virtual body belonged to me.
Change *- Scoring: (CH1 + CH2) / 2*
CH1 I felt like the form or appearance of my own body had changed.
CH2 I felt like the size (height) of my own body had changed.

Table 1: Questionnaire content. The scores are on a 7-Likert scale (1 = completely disagree, 7 = completely agree).

fidelity on the interaction between the lower body and the VE. Similarly, the pick-and-place task is intended to assess the impact of animation fidelity on the interaction between the upper body and the VE. To evaluate these two tasks, we took measurements regarding collisions and completion time. More specifically, we recorded: the total collision volume (V_c), the collision duration (T_c), the number of collisions (N_c), as well as the task completion time (T_{task}). This data was converted into more intuitive metrics as follows:

- Volume per collision $v = V_c/N_c$. It reflects how deep the avatar penetrated the obstacle during each collision, on average.
- Duration per collision $t = T_c/N_c$. It measures the average penetration time of the avatar into obstacles and how quickly participants corrected the collision when it occurred.
- Collision frequency $f = N_c/T_{task}$. It reflects how often the avatar collides with obstacles while performing the task. It is specified as the number of collisions per second.

With these metrics, we investigated the relationship between the animation fidelity and the virtual body-obstacle collisions. To accurately capture the volume and duration of collisions, a set of invisible small cubic colliders ($V_{collider} = 8\,\text{cm}^3$) were used to match the shape of each obstacle.

The goal of the copy-pose task is different from the other two. It evaluates the correctness of the static pose of the avatar when there are no hard constraints for the avatar's end-effectors positions (i.e. no contact points between the avatar and the VE). Thus, three pose-related metrics were used to assess the accuracy of users' poses:

- Jaccard Distance $JD = 1 - \frac{G \cap U}{G \cup U}$ (see Fig. 3). It measures the non-overlap area of the intersection between the 2D projection G of an example avatar over a plane and the 2D projection U of the avatar controlled by the user, divided by the union of the two projections.
- Mean per segment angle error (*MPSAE*) is defined as: $MPSAE = \frac{1}{\|S\|} \sum_{\hat{s}}^{\|S\|} \arccos(\hat{s}^* \cdot \hat{s})$, where S is the set of segments of the skeleton, $\hat{s}$ is the unit vector representing the direction of a segment s, and $\hat{s}^*$ is the direction of the segment in the given pose.
- Mean per part angle error *MPPAE* is like *MPSAE* but only considers one part of the body such as the spine or the limbs corresponding to arms and legs.

Participants could only observe the target poses as a 2D projection on a virtual wall that was located in front of them. Therefore, the metrics used in this task were all calculated based on the same 2D projection in the XY plane. For the Jaccard Distance, the lack of overlap between the two projections must not be a result of the user position being slightly offset with respect to the observed shape. Consequently, we iteratively applied translations in the 2D space to maximize the overlap between the two shapes before computing *JD*. For *MPPAE*, body segments of the avatar were grouped into three body parts: arms, legs and spine. This separation allowed us to study animation fidelity's impact individually on different body parts.

At the end of each block of tasks, participants completed a questionnaire (Table 1) which was adapted from a standard questionnaire from Virtual Embodiment Questionnaire (VEQ) [29]. The embodiment was measured through three main aspects: *agency*, *ownership* and *change*. *Agency* measures the sense of control, *ownership* measures the sense of owning the virtual body as if it is one's own real body, and *change* measures to what extent one feels the virtual body scheme differs in size from one's own real body.

The VEQ does not assess self-location discrepancies since it is not the goal of typical VR applications to produce such effects [29]. In our experiment, the use of the pelvis tracker guaranteed a correct placement of the self-avatar. The appearance and size of the avatar were kept the same through all conditions to guarantee that the only perceived differences would come from the changes in animation fidelity. Questions about *change* in VEQ are typically studied in the context of body swap experiments that manipulate avatars' body size, sex, race, etc. [18,38,39]. However, with the avatar's height and body proportions consistent with the user's physical body, *change* is not expected to be an influencing factor in our study.

The goal of the embodiment questionnaire was to gather the global experience after running the three tasks, so that it would gather both the importance of correct end-effector positioning and the accuracy of the body pose. We decided against asking the 15 questions after each task to avoid doing the experiment too long because it could introduce a biased source.

3.6 Hypotheses

We hypothesize that better animation fidelity would lead to better performance in terms of reducing the number of collisions, and also their volume and duration. Although our conditions had varied trade-offs in terms of the different components of animation fidelity (pose accuracy vs end-effector accuracy, as well as latency), we expected the highest performance for the full-body IMU-based motion capture system, followed by IK methods with VR devices as input. Similarly we would expect the full-body IMU-based motion capture system to outperform the IK solution when copying body poses given its higher number of sensors allowing for a more accurate capturing of the user pose. Finally we expected animation fidelity to affect the SoE of the user. Therefore, our hypotheses are:

H1 Animation fidelity impacts performance of the user in step-over-spikes and pick-and-place (tasks that require precise interaction with the environment), in terms of unintended collisions and completion time.

H2 Animation fidelity impacts performance in copy-pose task, which requires accuracy in the body pose.

H3 Animation fidelity affects the SoE.

4 Results

In this section we summarize the results of our experiment. The complete list of statistical significance and post-hoc tests values can be found in Table 2.

Metric	Test	Post-hoc
Step-over-spike Task		
Volume Per Collision (v)	Friedman test $\chi^2(2) = 11.80, p = .003, W = .235$	Wilcoxon test with Bonferroni adjustment MoCap > UIK ($p = .014, r = .552$), MoCap > FIK ($p = .009, r = .573$)
Duration Per Collision (t)	Friedman test $\chi^2(2) = 4.16, p = .125 (ns), W = .083$	- -
Collision Frequency (f)	Friedman Test $\chi^2(2) = 17.40, p < .001, W = .347$	Wilcoxon test with Bonferroni adjustment UIK > FIK ($p = .002, r = .643$) and MoCap > FIK ($p < .0001, r = .772$)
Completion Time (T)	One-way within-subject ANOVA $F_{2,48} = 4.870, p = .012, \eta^2 = .064$	Pairwise t-test with Bonferroni adjustment MoCap > FIK ($p = .003$)
Pick-and-place Task		
Volume Per Collision (v)	Friedman test $\chi^2(2) = .72, p = .698 (ns), W = .014$	- -
Duration Per Collision (t)	One-way within-subject ANOVA $F_{2,48} = 3.374, p = .043, \eta^2 = .056$	Pairwise t-test with Bonferroni adjustment Non-significant
Collision Frequency (f)	One-way within-subject ANOVA $F_{2,48} = 19.309, p < .0001, \eta^2 = .209$	Pairwise t-test with Bonferroni adjustment UIK > FIK ($p < .0001$) and UIK >MoCap ($p < .001$).
Completion Time (T)	Friedman Test $\chi^2(2) = 6.32, p = .042, W = .126$	Wilcoxon test with Bonferroni adjustment UIK > FIK ($p = .017, r = .541$).
Copy-pose Task		
Jaccard Distance (JD)	Friedman test $\chi^2(2) = 24.60, p < .0001, W = .491$	Wilcoxon test with Bonferroni adjustment. UIK > FIK ($p = .003, r = .632$). UIK > MoCap ($p < .0001, r = .848$). FIK > MoCap ($p = .005, r = .605$).
Mean Per Segment Angle Error ($MPSAE$)	Friedman test $\chi^2(2) = 44.20, p < .0001, W = .885$	Wilcoxon test with Bonferroni adjustment UIK > FIK ($p < .0001, r = .826$). UIK > MoCap ($p < .0001, r = .874$). FIK > MoCap ($p < .0001, r = .864$).
Mean Per Segment Angle Error ($MPPAE$)	Aligned Rank Transform ANOVA Animation Fidelity $F_{2,192} = 179.680, p < .0001, \eta^2 = .652$	Tukey's test UIK > FIK ($p < .0001$). UIK > MoCap ($p < .0001$). FIK > MoCap ($p < .0001$).
	Body Part $F_{2,192} = 244.480, p < .0001, \eta^2 = .718$	Arms > Legs ($p < .0001$) and Arms > Spine ($p < .0001$).
	Animation Fidelity : Body Part $F_{4,192} = 133.460, p < .0001, \eta^2 = .735$	For each Body Part: Arms: UIK > FIK ($p < .0001$) and UIK > MoCap ($p < .0001$) Legs: UIK>FIK ($p < .0001$). UIK>MoCap($p < .0001$). FIK>MoCap ($p = .003$). Spine: FIK > UIK ($p < .0001$) and FIK > MoCap ($p < .0001$) For each Animation Fidelity condition: UIK: Arms>Legs($p < .0001$). Arms>Spine($p < .0001$). Legs>Spine($p < .0001$). FIK: Arms>Legs ($p < .0001$). Arms>Spine ($p < .0001$). Legs>Spine ($p < .0001$). MoCap: Arms>Legs ($p < .0001$) and Arms>Spine ($p < .0001$).
Questionnaire		
Overall Score	One-way within-subject ANOVA $F_{2,48} = 21.033, p < .0001, \eta^2 = .155$	Pairwise t-test with Bonferroni adjustment UIK < FIK ($p < .0001$). UIK < MoCap ($p < .001$).
Agency	One-way within-subject ANOVA $F_{2,48} = 20.888, p < .0001, \eta^2 = .168$	Pairwise t-test with Bonferroni adjustment UIK < FIK ($p < .0001$) and UIK < MoCap ($p < .001$).
Ownership	Friedman test $\chi^2(2) = 14.5, p < .001, W = .290$	Wilcoxon test with Bonferroni adjustment UIK < FIK ($p < .001, r = .771$).
Change	Friedman test $\chi^2(2) = 2.06, p = .358 (ns), W = .041$	Wilcoxon test with Bonferroni adjustment Non-significant

Table 2: Statistical results. For task performance data, a higher value implies worse performance. For the questionnaire higher score is better. W value: 0.1-0.3 (small effect), 0.3-0.5 (medium effect) and $\geq$ 0.5 (large effect). η^2 value: 0.01-0.06 (small effect), 0.06-0.14 (medium effect), $\geq$ 0.14 (large effect). r value: 0.10 - 0.3(small effect),0.30 - 0.5(moderate effect) and $\geq$ 0.5 (large effect).

	v	t	f	T
Spike-over-spikes Task				
UIK	101.0(61.7)	0.060(0.067)	0.189(0.189)	102.0(25.6)
FIK	**86.2**(**81.3**)	**0.044**(**0.038**)	**0.099**(**0.125**)	**95.3**(**20.7**)
MoCap	126.0(34.4)	0.068(0.036)	0.361(0.377)	110.0(23.6)
Pick-and-place Task				
UIK	187.0(57.2)	0.271(0.090)	0.516(0.305)	101.0(38.0)
FIK	183.0(81.9)	0.280(0.085)	**0.268**(**0.178**)	**78.1**(**27.4**)
MoCap	**171.0**(**65.9**)	**0.231**(**0.091**)	0.283(0.166)	80.6(19.0)

Table 3: Mean and standard deviation for metrics of step-over-spikes task and pick-and-place task.

4.1 User performance on interaction tasks

We first present the results of user performance on the tasks that involved a direct interaction with the VE. Table 3 shows the mean (M) and standard deviation (SD) of all the metrics of the step-over-spikes and pick-and-place tasks. Fig. 4 shows the violin plots of the metrics of both tasks.

Shapiro-Wilk tests showed significant departures from normality for all three measures of the step-over-spikes task. Therefore, non-parametric within-subjects Friedman tests were used and they revealed significant differences for all metrics between animation fidelity conditions. Animation fidelity significantly affected volume per collision and collision frequency but not duration per collision. Table 2 includes a summary of $\chi^2(2)$, p-values and effect sizes calculated for these metrics. Pairwise post-hoc tests (Wilcoxon signed-rank tests) showed that MoCap had significantly higher values than FIK for all metrics except duration per collision, and a significantly higher value than UIK for collision frequency. It also showed UIK had significantly higher collision frequency than FIK.

For the pick-and-place task, Shapiro-Wilk tests showed that volume per collision and completion time data violated the normality assumption (p<.05), while the other two metrics did not. Therefore,

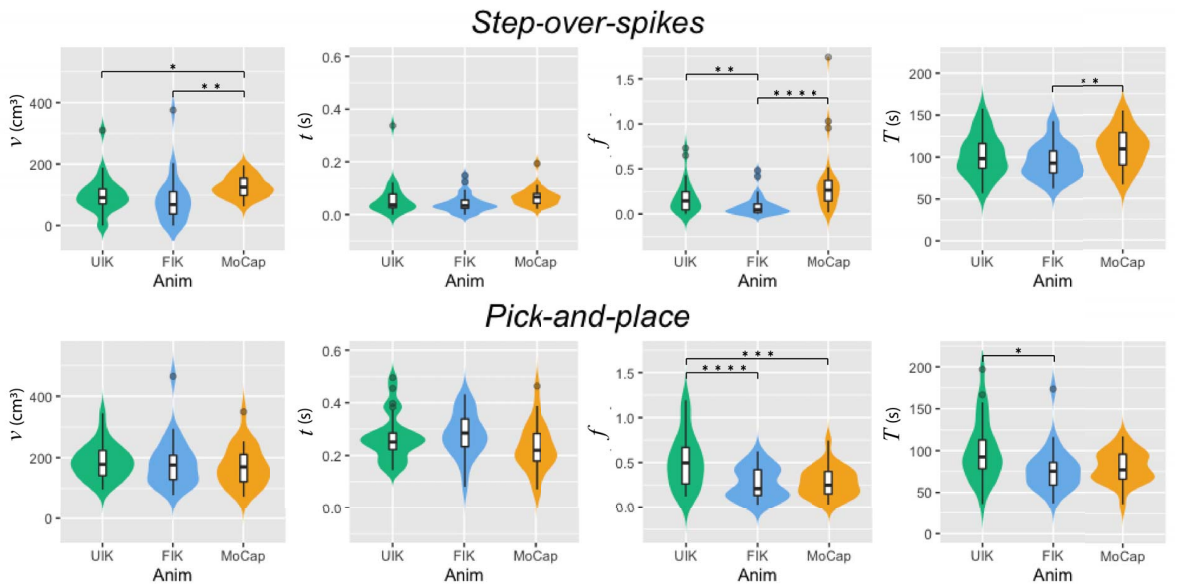

Figure 4: Violin plots for metrics of the step-over-spikes and pick-and-place tasks, showing results for collisions and task completion time. Asterisks represent the significance level: * ($p < .05$), ** ($p < .01$), *** ($p < .001$), **** ($p < .0001$).

Friedman tests and post-hoc Wilcoxon signed-rank tests were conducted for volume per collision and completion time, while one-way within-subject ANOVAs and pairwise t-tests were conducted for the others. The results revealed a significant effect of animation fidelity on duration per collision and collision frequency. Post-hoc tests showed that UIK had significantly higher collision frequency than FIK and MoCap, and a longer completion time than FIK.

Therefore, hypothesis **[H1]** was validated by these results of interactions tasks performed for both the upper body and lower body. We further analyze these results in Section 5.

4.2 User performance on pose-related tasks

We summarize the M and SD for all metrics of the copy-pose task in Table 4 and present the corresponding violin plots in Fig. 5. Shapiro-Wilk tests showed both JD and MPSAE data had a non-significant departures from normality ($p < .05$). Friedman tests were thus conducted for both metrics and revealed significant differences among the three animation fidelity conditions with medium to large effect sizes. Pairwise Wilcoxon tests with Bonferroni p-value adjustment demonstrated significant differences in all pairs of conditions. For both metrics, UIK had significantly higher error values than FIK and MoCap, and FIK had significantly higher errors than MoCap.

For MPPAE, we used a two-way repeated measures Aligned Rank Transform (ART) ANOVA after asserting the normality with a Shapiro-Wilk test ($p < .05$). The result revealed a significant main effect of animation fidelity and body part on MPPAE. It also showed a significant interaction effect between animation fidelity and body part. First, the post-hoc Tukey's tests demonstrated that, for all animation fidelity conditions, MPPAE was significantly higher for arms than for legs and spine. Next, when comparing the MPPAE for each body part, Tukey's tests showed that, for arms, the MPPAE was significantly higher for UIK than for the other conditions. For legs, UIK had significantly higher MPPAE than FIK and MoCap, and FIK had significantly higher MPPAE than MoCap. For the spine, FIK had significantly higher MPPAE than other conditions.

To summarize, these results validated our hypothesis **[H2]** in the sense that the pose errors were significantly lower when using MoCap than IK solutions.

	JD	*MPSAE*	*MPPAE*	
			Arms	28.6(3.45)
UIK	0.539(0.035)	13.90(1.51)	Legs	9.09(0.98)
			Spine	5.90(1.43)
			Arms	16.1(3.45)
FIK	0.512(0.040)	11.10(2.10)	Legs	7.22(1.67)
			Spine	10.1(3.26)
			Arms	**13.9**(**2.23**)
MoCap	**0.476**(**0.038**)	**8.03**(**1.19**)	Legs	**5.85**(**1.33**)
			Spine	**5.08**(**1.38**)

Table 4: Mean and standard deviation for metrics of the copy-pose task.

4.3 Sense of Embodiment

Table 5 shows the M and SD of the overall score of the SoE and subcomponent scores for *agency*, *ownership* and *change*. The violin plots for these scores can be found in Fig. 6. A one-way within-subject ANOVA showed a significant effect of animation fidelity on overall score of the SoE. The post-hoc tests (pairwise t-test) showed that the SoE score for UIK was worse than both FIK and MoCap.

We analyzed the average score of *agency* questions, Q1 - Q7, with a one-way within-subject ANOVA (see Fig. 2 for test values). The result showed a significant effect of animation fidelity on *agency* score. The post-hoc tests (pairwise t-test) showed that users reported the SoA in UIK significantly lower than FIK and MoCap.

Since a Shapiro-Wilk test showed a non-significant departure from normality, a Friedman test was conducted for the average score of *ownership* questions, Q8 - Q10. The result showed a significant effect of animation conditions on ownership. The post-hoc test (Wilcoxon test with Bonferroni p-value adjustment) showed UIK had a significantly lower *ownership* score than FIK.

The same set of tests as *ownership* were conducted for the average score of *change* questions, Q11 and Q12. A Friedman test showed no significant effect of animation conditions on *change*. post-hoc tests showed no significant difference on *change* in all condition pairs. Overall, these results validated our hypothesis **[H3]**.

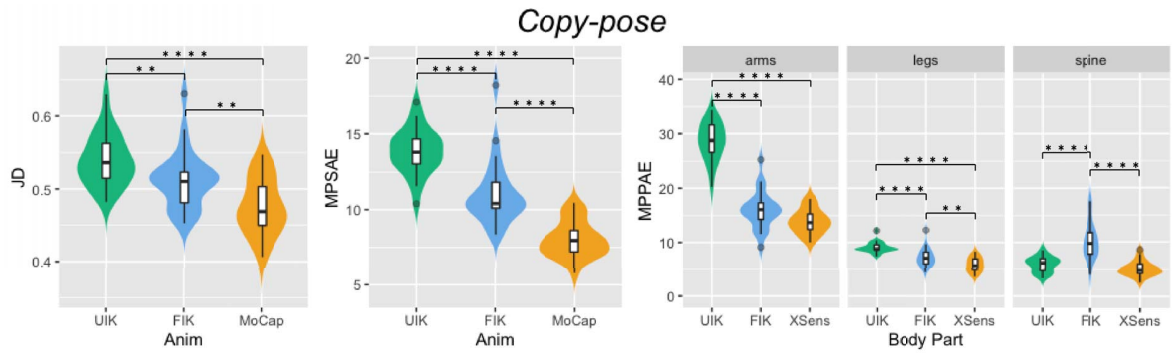

Figure 5: Violin plots of metrics obtained for the copy-pose task. Asterisks represent the significance level: * ($p < .05$), ** ($p < .01$), *** ($p < .001$), **** ($p < .0001$).

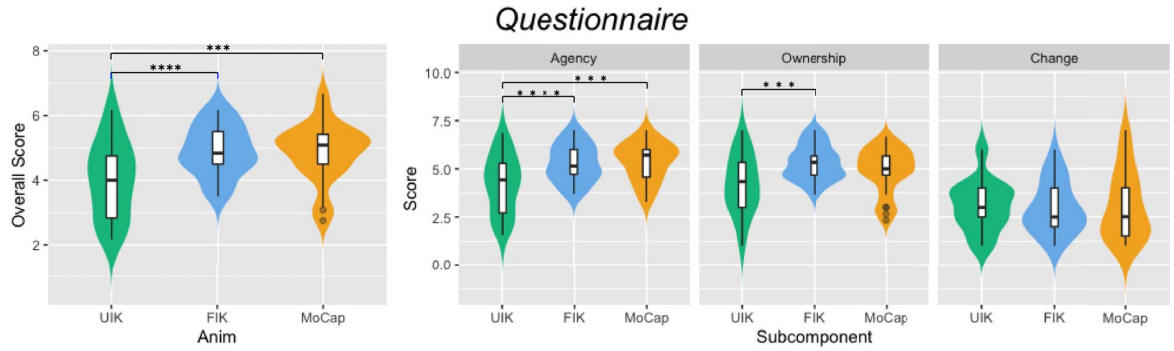

Figure 6: Violin plots for overall score of the SoE and scores for agency, ownership and change individually. Asterisks represent the significance level: * ($p < .05$), ** ($p < .01$), *** ($p < .001$), **** ($p < .0001$).

	Overall	*Agency*	*Ownership*	*Change*
UIK	4.03(1.18)	4.22(1.51)	4.19(1.54)	3.1(1.34)
FIK	**4.91**(**0.774**)	5.29(0.97)	**5.32**(**0.92**)	**2.98**(**1.35**)
MoCap	4.87(0.919)	**5.37**(**0.99**)	4.91(1.20)	3.08(1.79)

Table 5: Mean and standard deviation for the overall score of the SoE and scores of subcomponents.

5 Discussion

5.1 Accuracy of body pose vs. end-effector positioning

As expected, a motion capture suit is able to capture most of the human motion and accurately represent poses, as opposed to applying IK using only a few trackers as end-effectors. We quantitatively assessed this with the copy-pose task and found that the MoCap method performed significantly better than UIK and FIK for all metrics. Poses with MoCap were best aligned with the given poses and also when analyzing each body segment independently.

Therefore, we would expect MoCap to perform better in other tasks due to the high-quality tracking of poses. However, we found that high-quality poses do not improve task performance when tasks are not directly related to the pose, and instead require direct interactions with the VE. One possible explanation is that the positional drift from inertial systems results in the end-effectors moving away from their actual position. When this happens, the user's hands and feet are no longer co-located with their virtual representations, thus introducing inconsistencies (see Fig. 7). The higher latency of MoCap may have also contributed to these performance differences.

In the step-over-spikes task, MoCap was significantly worse than FIK in volume per collision, collision frequency and completion time. MoCap was significantly worse than UIK in volume per collision. We believe that for this task, having an accurate positioning of the feet (no drift) made users feel more confident and reliable when positioning the feet on the ground to avoid spikes. Both FIK and UIK achieved good foot positioning because IK solvers enforced the position of the feet to be the same as the trackers. In contrast, since MoCap is an IMU-based motion capture system, it does not have precise global positioning of the joints.

To lessen the positional drift issue, we moved the MoCap avatar to match the position of the VR pelvis tracker. This improves the overall co-location between the user and its avatar, but it may increase foot sliding. For instance, when one leg is acting as a supporting leg on the ground as the user's pelvis moves, if the pelvis of the MoCap animated avatar is forced to follow the HTC pelvis tracker, it makes the foot slide on the ground and increases the risk of collision with obstacles. To minimize this problem, we implemented a foot lock algorithm. This alleviated foot sliding but not the global position accuracy of the feet.

Overall, if the task requires accurate foot placement, it may be necessary to include foot trackers to position them accurately in the VE, while correctly posing all joints may not be necessary.

5.2 Upper body animation for accurate interactions with the environment

In the pick-and-place task, UIK performed significantly worse than MoCap and FIK in terms of collision frequency. However, we found MoCap and FIK to perform similarly. This is consistent with the results of the MPPAE in the copy-pose task, for which UIK also performed worse than MoCap and FIK due to incorrect elbow positioning. For the pick-and-place task, users had to correctly position the arm to reach the goal without touching the obstacles. The incorrect elbow positioning in UIK made the task more complicated, and thus more prone to collisions. We also found that users took significantly longer to finish the task with UIK than with FIK.

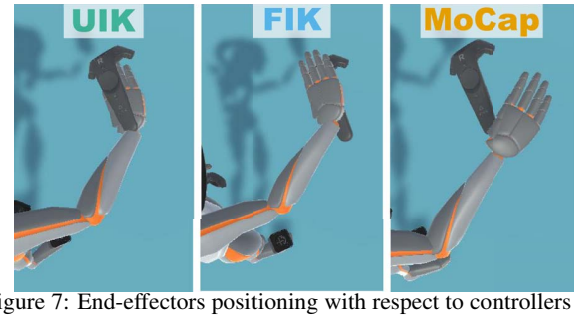

Figure 7: End-effectors positioning with respect to controllers for the different animation conditions.

When comparing FIK and MoCap, our results suggest that the additional tracking data for the elbows in MoCap did not help the participants achieve better performance in terms of collision avoidance in the pick-and-place task, and also in the arm part of pose replication in the copy-pose task. Even though MoCap provides a more accurate elbow position, we believe that the inaccurate end-effector positions lead to more collisions with the obstacles. Another explanation may be due to the latency of the MoCap. A few participants commented that their virtual arms were less responsive when using MoCap while performing the pick-and-place task. As Waltemate et al. [36] stated, when latency increases above 75 ms, user's motor performance in VR tends to decline.

Even if FIK provides less accurate poses for the elbows, its responsiveness and end-effector accuracy compensate for this. Participants can quickly avoid obstacles by adjusting the controllers' position. The result is consistent with the work by Parger et al. [24].

5.3 Performance differences between arms and legs

The results of the MPPAE in the copy-pose task suggest that the arm poses were less precise than the leg poses. The angle error was larger in the arms than in the legs for all conditions. One possible explanation is that the range of movements a human person can do with their upper body is wider than with the lower body. We also studied whether users noticed the tracking inaccuracy by comparing the scores given in questions related to arms (Q2-Q4) and legs (Q5-Q6). The score for arms ($M = 4.60$, $SD = 1.19$) was statistically ($p < .0001$) lower than legs ($M = 5.41$, $SD = 1.05$) when performing a t-test. When performing a two-way ANOVA, adding the animation fidelity as a condition, we found no statistical difference between the scores given to the arms questions between FIK and MoCap.

The participant-reported differences in responsiveness between FIK and MoCap for arm movement were not observed for the legs during the step-over-spikes task.

Based on the result above, we recommend focusing on the upper body when animating a self-avatar since it seems necessary to have higher-quality animations for arms. Lower-quality animation may be enough for the legs. Therefore, as some works have suggested [27], it may not be necessary to include all tracker devices for the lower body when the task does not require accurate foot placement.

5.4 High Sense of Agency can be achieved with a small set of tracking devices

The questionnaire data showed no statistically significant differences between FIK and MoCap. However, as mentioned before, MoCap achieved better results (*JD* and *MPSAE*) than FIK and UIK in the copy-pose task. It suggests that the SoA is not only related to the pose, but also to the interaction with the VE, e.g., we found that in the pick-and-place task MoCap did not achieve the best results.

In other words, our results suggest that one can feel the same level of control over self-avatars animated by a high-end motion capture suit with 17 IMUs or a small set of tracking devices (one HMD, two controllers, and three trackers) and a high-quality IK solution. This finding is consistent with Galvan Debarba et al. [10] that suggested a total of 8 trackers were enough to achieve the same plausibility illusion as an optical-based motion capture system with 53 retro-reflective markers. Goncalves et al. [12] suggested that increasing tracking points, from 3 to 6, does not significantly improve the SoA.

More research is needed to understand how to improve the SoA, given that a higher number of trackers (MoCap) did not always improve the agency scores when compared to a full-body IK such as FIK. Other factors such as end-effectors position accuracy, latency or animation smoothness may affect the users' perception.

It would also have been interesting to randomize the task order so that we could have analyzed whether the results of the SoE were affected by which was the last task being experienced by the participant. However, by looking at the results, we observe that the step-over-spike task (the first task) had FIK giving better quantitative results, the pick-and-place task (the second task) had similar performance for FIK and Mocap, and in the copy-pose task (the last task) MoCap had the best results. Even though the last task had better performance for Mocap, the embodiment questionnaires showed similar results for FIK and MoCap (not statistically significant) which may indicate that the questionnaire did gather the overall experience.

6 Conclusions and Future Work

We conducted a user study to examine the impact of the avatar's animation fidelity on user performance and the SoA. Our results suggest that the IMU-based motion capture system performed better than IK solutions for applications that require pose accuracy. However, IK solutions outperform IMU-based motion capture systems when directly interacting with the VE. In these cases, accurate end-effector placement and low latency may be more critical than exact pose matching due to proprioception. Our study also suggests that a high-end IK solution with sparse input (6 trackers) can achieve similar levels of the SoA as an IMU-based motion capture with dense input (17 trackers). We believe these results give insight into how animation fidelity affects user performance and perception, providing future research directions toward improving self-avatar animation fidelity in VR. Our work also highlights the limitations of current technology to achieve correct self-avatar animation (such as latency, end-effectors and body pose inaccuracy), and thus motivates future research to overcome these issues.

A limitation of our experiment is that the robotic avatar did not accurately match the shape of the participant. Since the avatar's limbs were much thinner than the participants' ones, and because they used hand-held controllers, self-contacts suggested by some copy-pose targets were not reproduced by the avatar (regardless of the condition). In fact, no participant referred to this issue. Further studies are required to study the role of animation fidelity and self-contact [2] when the avatar accurately matches the user's shape.

For future research, we would like to investigate whether participants could perform better using an optical motion capture system, providing both accurate pose and global position. This new condition will allow the decoupling of the positional drift issue from the accuracy of the body pose, allowing for a more in-depth study of the perceptual results. We believe future studies that integrate hand tracking like RotoWrist [30] or data-driven methods for self-avatar animation would be valuable to provide more insight into how animation fidelity impacts the SoE and user performance in VR.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860768 (CLIPE project) and from MCIN/AEI/10.13039/501100011033/FEDER, UE (PID2021-122136OB-C21). Jose Luis Ponton was also funded by the Spanish Ministry of Universities (FPU21/01927).

REFERENCES

[1] S. Aseeri and V. Interrante. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2608–2617, May 2021. doi: 10.1109/TVCG.2021.3067783

[2] S. Bovet, H. G. Debarba, B. Herbelin, E. Molla, and R. Boulic. The critical role of self-contact for embodiment in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1428–1436, Feb. 2018. doi: 10.1109/TVCG.2018.2794658

[3] P. Caserman, A. Garcia-Agundez, R. Konrad, S. Göbel, and R. Steinmetz. Real-time body tracking in virtual reality using a vive tracker. *Virtual Reality*, 23(2):155–168, Nov. 2019. doi: 10.1007/s10055-018-0374-z

[4] Y. Choi, J. Lee, and S.-H. Lee. Effects of locomotion style and body visibility of a telepresence avatar. In *Proc. IEEE VR*, pp. 1–9. IEEE, New York, Mar. 2020. doi: 10.1109/VR46266.2020.00017

[5] S. Cmentowski and J. Krüger. Effects of task type and wall appearance on collision behavior in virtual environments. In *Proc. IEEE CoG*, pp. 1–8. IEEE, New York, Aug. 2021. doi: 10.1109/CoG52621.2021.9619039

[6] N. David, A. Newen, and K. Vogeley. The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and cognition*, 17(2):523–534, 2008. doi: 10.1016/j.concog.2008.03.004

[7] A. L. Edwards. Balanced latin-square designs in psychological research. *The American Journal of Psychology*, 64(4):598–603, 1951.

[8] J. C. Eubanks, A. G. Moore, P. A. Fishwick, and R. P. McMahan. The Effects of Body Tracking Fidelity on Embodiment of an Inverse-Kinematic Avatar for Male Participants. In *Proc. ISMAR*, vol. 28, pp. 54–63. IEEE, New York, Nov. 2020. doi: 10.1109/ISMAR50242.2020.00025

[9] R. Fribourg, F. Argelaguet, A. Lecuyer, and L. Hoyet. Avatar and Sense of Embodiment: Studying the Relative Preference Between Appearance, Control and Point of View. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2062–2072, May 2020. doi: 10.1109/TVCG.2020.2973077

[10] H. GalvanDebarba, S. Chague, and C. Charbonnier. On the Plausibility of Virtual Body Animation Features in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1880–1893, 2020. doi: 10.1109/TVCG.2020.3025175

[11] B. Gao, J. Lee, H. Tu, W. Seong, and H. Kim. The Effects of Avatar Visibility on Behavioral Response with or without Mirror-Visual Feedback in Virtual Environments. In *Proc. IEEE VRW*, pp. 780–781. IEEE, New York, Mar. 2020. doi: 10.1109/VRW50115.2020.00241

[12] G. Gonçalves, M. Melo, L. Barbosa, J. Vasconcelos-Raposo, and M. Bessa. Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive VR. *Virtual Reality*, 26(1):1–14, Mar. 2022. doi: 10.1007/s10055-021-00530-5

[13] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll. Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *Proc. SIGGRAPH Asia*, 37(6):1–15, Nov. 2018. doi: 10.1145/3272127.3275108

[14] J. Jiang, P. Streli, H. Qiu, A. Fender, L. Laich, P. Snape, and C. Holz. Avatarposer: Articulated full-body pose tracking from sparse motion sensing, July 2022. doi: 10.48550/ARXIV.2207.13784

[15] Y. Jiang, Y. Ye, D. Gopinath, J. Won, A. W. Winkler, and C. K. Liu. Transformer Inertial Poser: Real-time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation. In *Proc. SIGGRAPH Asia*. ACM, New York, Dec. 2022. doi: 10.1145/3550469.3555428

[16] K. Kilteni, R. Groten, and M. Slater. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, Nov. 2012. doi: 10.1162/PRES_a_00124

[17] K. Kilteni, A. Maselli, K. P. Kording, and M. Slater. Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9, Mar. 2015. doi: 10.3389/fnhum.2015.00141

[18] M. Kocur, F. Habler, V. Schwind, P. W. Woźniak, C. Wolff, and N. Henze. Physiological and perceptual responses to athletic avatars while cycling in virtual reality. ACM, New York, May 2021. doi: 10.1145/3411764.3445160

[19] A. Koilias, C. Mousas, and C.-N. Anagnostopoulos. The Effects of Motion Artifacts on Self-Avatar Agency. *Informatics*, 6(2):18, Apr. 2019. doi: 10.3390/informatics6020018

[20] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose. Do You Feel Like Passing Through Walls?: Effect of Self-Avatar Appearance on Facilitating Realistic Behavior in Virtual Environments. In *Proc. CHI*, pp. 1–14. ACM, Apr. 2020. doi: 10.1145/3313831.3376562

[21] R. Oliva, A. Beacco, X. Navarro, and M. Slater. Quickvr: A standard library for virtual embodiment in unity. *Frontiers in Virtual Reality*, p. 128, 2022. doi: 10.3389/frvir.2022.937191

[22] Y. Pan and A. Steed. Avatar Type Affects Performance of Cognitive Tasks in Virtual Reality. In *Proc. VRST*, pp. 1–4. ACM, New York, Nov. 2019. doi: 10.1145/3359996.3364270

[23] Y. Pan and A. Steed. How Foot Tracking Matters: The Impact of an Animated Self-Avatar on Interaction, Embodiment and Presence in Shared Virtual Environments. *Frontiers in Robotics and AI*, 6, Oct. 2019. doi: 10.3389/frobt.2019.00104

[24] M. Parger, J. H. Mueller, D. Schmalstieg, and M. Steinberger. Human Upper-Body Inverse Kinematics for Increased Embodiment in Consumer-Grade Virtual Reality. In *Proc. VRST*. New York, Nov. 2018. doi: 10.1145/3281505.3281529

[25] M. Paulich, M. Schepers, N. Rudigkeit, and G. Bellusci. Xsens mtw awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3d kinematic applications. *Xsens: Enschede, The Netherlands*, pp. 1–9, 2018.

[26] J. L. Ponton, E. Monclus, and N. Pelechano. AvatarGo: Plug and Play self-avatars for VR. In *Proc. Eurographics*. The Eurographics Association, May 2022. doi: 10.2312/egs.20221037

[27] J. L. Ponton, H. Yun, C. Andujar, and N. Pelechano. Combining Motion Matching and Orientation Prediction to Animate Avatars for Consumer-Grade VR Devices. *Computer Graphics Forum*, 41(8), Sept. 2022. doi: 10.1111/cgf.14628

[28] B. Ries, V. Interrante, M. Kaeding, and L. Phillips. Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. In *Proc. VRST*, pp. 59–66. ACM, New York, Nov. 2009. doi: 10.1145/1643928.1643943

[29] D. Roth and M. E. Latoschik. Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3546–3556, Sept. 2020. doi: 10.1109/TVCG.2020.3023603

[30] F. Salemi Parizi, W. Kienzle, E. Whitmire, A. Gupta, and H. Benko. Rotowrist: Continuous infrared wrist angle tracking using a wristband. In *Proc.VRST*, VRST '21. ACM, New York, Dec. 2021. doi: 10.1145/3489849.3489886

[31] A. Steed. A simple method for estimating the latency of interactive, real-time graphics simulations. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, p. 123–129. ACM, New York, 2008. doi: 10.1145/1450579.1450606

[32] N. Stein, G. Bremer, and M. Lappe. Eye tracking-based lstm for locomotion prediction in vr. In *Proc. IEEE VR*, pp. 493–503. IEEE, New York, Mar. 2022. doi: 10.1109/VR51125.2022.00069

[33] N. Toothman and M. Neff. The Impact of Avatar Tracking Errors on User Experience in VR. In *Proc. IEEE VR*, pp. 756–766. IEEE, New York, Mar. 2019. doi: 10.1109/VR.2019.8798108

[34] E. Tuveri, L. Macis, F. Sorrentino, L. D. Spano, and R. Scateni. Fitmersive games: Fitness gamification through immersive vr. In *Proc. AVI*, p. 212–215. ACM, New York, 2016. doi: 10.1145/2909132.2909287

[35] L. Wagnerberger, D. Runde, M. T. Lafci, D. Przewozny, S. Bosse, and P. Chojecki. Inverse kinematics for full-body self representation in vr-based cognitive rehabilitation. In *Proc. IEEE ISM*, pp. 123–129. IEEE, New York, Nov. 2021. doi: 10.1109/ISM52913.2021.00029

[36] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. VRST'16, p. 27–35. ACM, New York, Nov. 2016. doi: 10.1145/2993369.2993381

[37] A. Winkler, J. Won, and Y. Ye. QuestSim: Human motion tracking from sparse sensors with simulated avatars. In *Proc. SIGGRAPH Asia*. ACM, New York, Dec. 2022. doi: 10.1145/3550469.3555411

[38] M. Wirth, S. Gradl, G. Prosinger, F. Kluge, D. Roth, and B. M. Eskofier. The impact of avatar appearance, perspective and context on gait variability and user experience in virtual reality. In *Proc. IEEE VR*, pp. 326–335. IEEE, New York, Mar. 2021. doi: 10.1109/VR50410.2021.00055

[39] E. Wolf, N. Merdan, N. Dölinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. The embodiment of photorealistic avatars influences female body weight perception in virtual reality. In *Proc. IEEE VR*, pp. 65–74. IEEE, New York, Mar. 2021. doi: 10.1109/VR50410.2021.00027

[40] X. Yi, Y. Zhou, and F. Xu. Transpose: Real-time 3d human translation and pose estimation with six inertial sensors. *ACM Transaction on Graphics*, 40(4), jul 2021. doi: 10.1145/3450626.3459786

[41] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration. In *Proc. IEEE VR*, pp. 547–556. IEEE, New York, Mar. 2019. doi: 10.1109/VR.2019.8797719

[42] Q. Zeng, G. Zheng, and Q. Liu. PE-DLS: A novel method for performing real-time full-body motion reconstruction in VR based on Vive trackers. *Virtual Reality*, pp. 1–17, Mar. 2022. doi: 10.1007/s10055-022-00635-5