

On the Plausibility of Virtual Body Animation Features in Virtual Reality

Henrique Galvan Debarba, Sylvain Chagué, Caecilia Charbonnier

Abstract—We present two experiments to assess the relative impact of different levels of body animation fidelity on plausibility illusion (Psi). The first experiment presents a virtual character that is not controlled by the user ($n = 13$), while the second experiment presents a user-controlled virtual avatar ($n = 24$, all male). Psi concerns how realistic and coherent the events in a virtual environment look and feel and is part of Slater's proposition of two orthogonal components of presence in virtual reality (VR). In the experiments, the face, hands, upper and lower bodies of the character or self-avatar were manipulated to present different degrees of animation fidelity, such as no animation, procedural animation, and motion captured animation. Participants started the experiment experiencing the best animation configuration. Then, animation features were reduced to limit the amount of captured information made available to the system. Participants had to move from this basic animation configuration towards a more complete one, and declare when the avatar animation realism felt equivalent to the initial and most complete configuration, which could happen before all animation features were maxed out. Participants in the self-avatar experiment were also asked to rate how each animation feature affected their sense of control of the virtual body. We found that a virtual body with upper and lower body animated using eight tracked rigid bodies and inverse kinematics (IK) was often perceived as equivalent to a professional capture pipeline relying on 53 markers. Compared to what standard VR kits in the market are offering, i.e. a tracked headset and two hand controllers, we found that foot tracking, followed by mouth animation and finger tracking, were the features that added the most to the sense of control of a self-representing avatar. In addition, these features were often among the first to be improved in both experiments.

Index Terms—Self-representing avatar animation, character animation, plausibility illusion, presence, sense of control, agency.

1 INTRODUCTION

Motion capture animation is used to amplify the Virtual Reality (VR) experience of users in different manners. On the one hand, it is used to animate a self-representing avatar in the virtual environment and is effective when it moves and behaves in response to the actions of the user [1], [2], particularly when the mapping of user movements to avatar movements is timely, accurate and looks natural [3]. On the other hand, it is also used to represent characters addressing the user, such as pre-recorded animation content, fellow users sharing an experience [4], [5], [6], or live actors that can augment the development of a story line within the virtual environment [7].

Real-time virtual human animation requires the tracking of numerous body parts, often relying on costly specialized hardware. For instance, fingers are normally tracked using gloves with embedded bending and/or inertial sensors, while pelvis, head, and limbs are often tracked using stationary cameras and optical markers attached to the body. It is only recently that the consumer VR industry became invested in addressing the concern of an embodied self-representation in VR. Products such as the HTC Tracker¹, which extends tracking possibilities available to the end user, are an example. Understanding the relative importance of animation features could shed light on the decision making process of setting up a full-body motion capture system. For instance, on

the trade-offs of the subjective perception of an improved VR experience and the monetary cost of adopting certain technologies.

With this problem in mind, we performed two experiments to improve our understanding of the relative impact of different animation features on plausibility illusion (Psi) and the feeling of control of a virtual body. Psi concerns the feeling that events in a virtual environment may be really happening and, as proposed by Slater [8], is one of two orthogonal components of presence. The sense of control relates to the concepts of agency and embodiment, where the perception of sensorimotor contingencies can affect the experience of embodiment that one develops with a virtual representation of oneself [9].

In the first experiment, participants were immersed in a virtual environment using a head mounted display (HMD) and had to repeatedly watch a short animation clip produced with the help of a professional actor. We addressed the question: to what extent does the animation fidelity of a character that is not controlled by the user affect Psi? In the second experiment, male participants were equipped with a motion capture suit, a pair of finger tracking gloves and an HMD. This setup allowed for the interactive control and observation of a self-representing avatar while interacting with a virtual environment. We addressed the question: to what extent does the animation fidelity of a self-avatar affect Psi and the sense of control of that avatar?

We approached these questions with the experimental methodology described by Slater et al. [10], which adapts a classical psychophysics method to presence research. In both experiments, the face, hands and the upper and lower bodies of the animated character (first experiment) or self-avatar (second experiment) could be set to different degrees of animation fidelity, such as no animation (i.e. motionless), procedural animation and motion capture. Participants started by experiencing the most complete

• H. Galvan Debarba, was with Artanim Foundation, Switzerland, and is now with the IT University of Copenhagen, Denmark.

• S. Chagué and C. Charbonnier are with Artanim Foundation, Switzerland.

E-mail: hgdebarba@gmail.com

Manuscript received April 19, 2005; revised August 26, 2015.

1. <https://www.vive.com/us/vive-tracker/>

animation configuration, in which full-body tracking is used to drive the animation. Then, the animation features were reduced or removed by limiting the amount of captured information available for pose reconstruction. Participants had to progress from this basic animation configuration to more complete configurations and declare when the animation realism felt equivalent to the initial and most complete condition. We refer to this perceived equivalency as a configuration *match*.

This paper is organized as follows: the next section discusses the constructs of presence and Psi, the sense of control, perception of virtual human animation and related work. Section 3 describes the materials and methods, including the implementation details of the animation configurations available in the experiments. Section 4 presents the results of the experiments, which are further discussed in Section 5. Finally, Section 6 presents our conclusions.

2 BACKGROUND

2.1 Presence, PI and Psi

While the term “presence” is widely used to designate the feeling of “being there” in the virtual environment, [11] and of non-technological-mediation with a virtual world [12], its precise and specific understanding is still debated today (see [13] for a review). On this topic, Slater [8] proposed a framework with two orthogonal components in the scope of presence, place illusion (PI) and plausibility illusion (Psi). While PI encompasses the classical definition of presence (i.e. the feeling of “being there”), Psi relates to how realistic and coherent the features of the virtual environment look and feel, and to what extent participants feel that what happens in the virtual environment conforms to their model of reality.

Experimental procedures to assess the sense of presence have generally relied on questionnaires [14] and physiological [15] or behavioral [1] response of users to virtual events. However, Slater et al. [10] introduced an experimental methodology that builds on the field of psychophysics, more specifically, on the experimental methodology used to assess color perception. The methodology presents participants with a reference configuration and, then, with a variety of different configurations, which participants are supposed to judge for equivalency to that reference configuration, for instance, whether the tested colors match to a reference color from the perspective of human perception. Slater et al. elaborate on how this methodology can be adapted to the study of presence, they build on the argument that “a system (A) is said to be more immersive than another (B) if (A) can be used to simulate an application as if it was running in (B)” [10] to propose the use of a highly immersive system that is capable of simulating a variety of less immersive systems. Then, they run trials in a variety of less immersive (simulated) scenarios and ask participants when and whether their feeling of presence matches that which they felt in the most immersive scenario (i.e. the reference). That is, whether a simpler virtual reality simulation can be perceived to be as efficient, in terms of presence, to the more complete virtual reality simulation.

Slater et al. [10] explored this framework to evaluate the relative importance of virtual reality simulation elements to PI and Psi. The authors concluded that effective PI is mainly related to immersive apparatus and to having and controlling a virtual body (sensorimotor contingencies relating real and virtual body). Thus, being tightly connected to the notion of an efficient feedback loop. On the other hand, effective Psi has been associated to illumination realism and self-representing virtual body animation. Hence, Psi

seems to be especially related to higher order cognitive priors about the elements that are contained in reality, and how one expects these elements to behave and look like. It is argued that both illusions can occur, together or independently, albeit the fact that participants know that the virtual environment is a simulation.

The experimental methodology proposed in [10] has been utilized in PI and Psi research [16], [17], [18], [19], [20]. Azevedo et al. [16] explored the contribution of different sensory channels in an immersive experience, Gao et al. [19] explored the impact of consistent visual appearance, sound and dynamic effects, Bergstrom et al. [17] explored the effect of virtual human behaviors and sound in a virtual environment on Psi, and Skarbez et al. [18] evaluated how the coherence of different elements of the virtual environment — namely the scenario, virtual human, virtual body and physical coherence — affect Psi. Notably, Skarbez et al. [18] concluded that the virtual body, which is in fact one’s own representation in VR, was the most important factor for achieving an elevated illusion of plausibility. They reported that the virtual body was set to the maximum (visible and animated self-representing body) on more than 99% of the trials. However, participants could only enable a single level of virtual body animation. We expand on that and explore animation in depth, considering virtual human animation that is not controlled by the user and is presented as a virtual character as well as animation that is controlled by the user and is presented as a self-avatar. We argue that a better understanding of how users perceive the animation of a virtual body can improve our understanding of the experience of Psi and is crucial in directing development resources for tracking and animation solutions.

2.2 Sense of Control

The feeling of being in control of the actions of a body, the so-called sense of agency [21], is a crucial component of the sense of embodiment. One feels embodied due “to the ensemble of sensations that arise in conjunction with being inside, having, and controlling a body” [9] (p. 374). In this context, Fribourg et al. [22] employed the experimental methodology proposed by Slater et al. [10] to analyze the influence of different factors on the sense of embodiment in a VR experience. Participants were able to set the appearance, the point of view, and the level of control of a self-representing avatar. The level of control could be set to automatic movement (e.g. pre-recorded animation), inverse kinematics (driven by the end-effectors), or full body motion capture. Their results indicate a preference for some degree of control early in the experience, often being the first or second feature to be improved. The work of Fribourg et al. [22] is distinct, but complementary to what we present here. The authors assessed the relative importance of animation control to the sense of embodiment of a virtual body as a whole. Here, we explore the perception of animation features in more depth and with a focus on Psi.

Moreover, the self-avatar animation experiment, which is the second experiment in this paper, also explores how each change of the avatar animation configuration affects one’s sense of control over the virtual body. Agency in humans represents an adaptive causal link, that seems to be constantly modeled by action and outcome contingencies developed by repetition [23]. When referring to one’s own body, the sense of agency seems predominantly related to the sensation of motor control over that body. Notably, a variety of experiments have demonstrated how the sense of control and agency can be experimentally manipulated [24], [25], [26]. Moreover, in the context of VR, experimental

results have shown that a virtual body that moves like the user can be felt as the user's own body [1], [2]. Here, participants in the avatar animation experiment were asked to report on how the animation factors that they were able to enable affected their experience of control over the virtual body.

2.3 Perception of Virtual Body Animation

While there is an extensive body of research on the effect of visual fidelity and use of animations, such as render style and shading [27], [28], [29], [30], [31], we found few studies focusing on the perception of virtual animation features.

Kokkinara and McDonnell [3] evaluated the effect of appearance and control of the head pose (position and orientation) to the senses of agency and ownership of a virtual head. Results with respect to the effect of head pose control were null. The authors hypothesized that the fact that facial movement (facial muscles, mouth and eyes) was present across other experimental conditions could result in an elevated sense of agency and ownership regardless of head pose control. Other researchers have looked at the perception of time-warped movements [32] and the distinctiveness and attractiveness of human motion in virtual bodies [33].

To the best of our knowledge, little research has been carried out on the relative contribution of animation features to the perception of animation realism. Most notably, Hodgins et al. [34] examined the salience of animation abnormalities on virtual characters. In the study, participants classified facial abnormalities as more disturbing than the lack of lower body movement or the unrealistic behavior of one of the arms of the character. The authors suggested that these results reflect the aversion to animation conditions that resembled illness or injury. Here, we study the relative importance of character and avatar animation features and relate them with the implementation barriers in the context of real-time motion capture for immersive VR.

3 MATERIALS AND METHODS

This section details the animation features, their implementation, and the experiments. Hereinafter, *character experiment* is used to refer to the evaluation of an animated character (i.e. a virtual character that interacts with the user in VR) while *avatar experiment* is used to refer to the evaluation of a self-avatar (i.e. the virtual representation of the user in VR).

3.1 Experimental Conditions

In the experiments, we manipulated four different *animation fidelity factors*: upper body (*UB*), lower body (*LB*), facial (*FA*) and hand (*HA*) animation. We chose to divide the whole body into these four factors based on typical tracking equipment separation. For instance, hand/fingers and facial animation normally require specialized hardware, while current consumer VR equipment provides partial upper body tracking, but no lower body tracking.

Moreover, by reducing the tracking information available to the rendering computer we are able to simulate a simpler motion capture system, e.g. by using only the poses of the hands and head of the full-body tracking data we can simulate the tracking possible with a VR consumer set. Then, we use the inverse kinematics (IK) solvers and procedural feet animation of the FinalIK library² to generate valid poses for the joints that are omitted and to animate

the character or self-avatar. An overview of the VR IK solver used here is available at [35]. The possible manipulations of the four animation fidelity factors are described below:

Upper Body: upper body animation comprises the animation fidelity of the pelvis, spine, head, shoulders, arms and wrists.

- (*UB* = 0): only hands and head pose tracking information is used for animation. The IK solver generates valid poses for the upper body joints that are not being tracked. The pelvis orientation is defined by the direction that the tracked person is facing.
- (*UB* = 1): in addition to hands and head, pelvis and elbows tracking information is provided. The IK solver generates valid poses for the upper body joints that are not being tracked.
- (*UB* = 2): all available upper body tracking information is used for animation. This includes hands, wrists, forearms, elbows, arms, shoulders, clavicles, pelvis, chest, and head tracking information.

Lower Body: lower body animation comprises the animation fidelity of the hips, knees, feet, and if set at the highest fidelity level, it overlaps with the upper body on the control of the pelvis.

- (*LB* = 0): no tracking information related to the lower body limbs is available. The locomotion functionality from FinalIK is used to procedurally animate the legs based on the pelvis pose. Feet and knee bending direction happens toward the direction that the tracked person is facing.
- (*LB* = 1): foot tracking information is available. The IK solver uses this information to generate valid knee and hip joints rotations. The knee bending axis is perpendicular to the plane defined by the principal component of each foot and the respective ankle to hip vector.
- (*LB* = 2): all available lower body tracking information is used for animation. This comprises feet, ankles, lower legs, knees, upper legs, as well as pelvis pose if not yet available as part of upper body levels 1 and 2.

Face: facial animation comprises the animation fidelity of the eyes and mouth.

- (*FA* = 0): face is not animated, and a static facial expression is used.
- (*FA* = 1): in the *character experiment*, an iPhone X with the ARKit library was used to record the facial expressions of the actor and to animate the face of the character. In the *avatar experiment*, the mouth is animated based on speech captured by the microphone built in the HMD. Lipsync is achieved by translating sounds to visemes. The eyes is animated to make the self-avatar look at itself through a mirror placed in the virtual environment. This self-gaze feature is implemented by rotating the eyes so that the vector from the center of the eyes through the pupils are perpendicular to the mirror plane. The eyes were set to rotate no more than 15 degrees relative to a pre-defined look forward orientation.

Hands: hand animation comprises the animation fidelity of the fingers and thumbs.

- (*HA* = 0): thumbs and fingers are not animated and a predefined pose is used instead.
- (*HA* = 1): procedural animation is used to automatically move thumbs and fingers when close enough to an object. In

the *character experiment*, fingers and thumbs wrap around the virtual objects that the character can interact with, such as documents and a table. In the *avatar experiment*, fingers and thumbs wrap around a specific virtual object that the self-avatar can interact with, and which is also represented by a tracked physical replica.

- ($HA = 2$): the bending sensors integrated in a pair of gloves (Manus VR) are used to animate thumbs and fingers. This animation fidelity level is only used in the *avatar experiment*.

A video demonstrating the visual effect of the different feature levels is provided as supplemental material. In total, the *character experiment* had 36 possible combinations of animation features (3 for Upper Body, 3 for Lower Body, 2 for Face, and 2 for Hands - $3 \times 3 \times 2 \times 2 = 36$), while the *avatar experiment* had 54 possible combinations of animation features (3 for Upper Body, 3 for Lower Body, 2 for Face, and 3 for Hands $3 \times 3 \times 2 \times 3 = 54$). We refer to any specific combination of animation features as an animation configuration and, as in Slater et al. [10], we designate an animation configuration as a vector $c = \{UB, LB, FA, HA\}$.

3.2 Equipment

A Vicon MXT40S with 24 cameras and the Shogun software³ were used for tracking and reconstruction of the participants poses. This tracking system uses the cameras to detect and reconstruct the position and movement of retro-reflective markers in physical space, a protocol with 53 markers was used in this case. Full body motion capture with Shogun requires a two step calibration. In the first step, the application maps the retro-reflective markers to body segments. Then, in the second step, the application estimates the center of rotation of the main joints in the participants body, as defined by a simplified skeleton model. This results in an approximation of the length of the body segments of the participant, which is used to adjust the body segments of the virtual body (i.e. scale). Finally, the software reconstructs the poses of the person based on the tracked markers. The reconstruction includes upper body, lower body and head poses, but does not include finger and facial poses. For the *character experiment*, a recording session was carried out previous to the experiment, where the system was used to estimate and store the actor's movements. The virtual body that represents the actor was created to comply with the hierarchical model used by Shogun, as obtained through the calibration. In the *avatar experiment*, the body size measurements and real-time poses are streamed live to the computer used to render the virtual environment. This information is used to match the size and pose of the avatar to that of the participant at every update (estimated latency from capture to display of 50 ms). The capture rate was set to 120 frames per second, and the capture and display computers were connected in a cabled Ethernet network.

An Oculus Rift head mounted display (HMD) was used in both experiments (1080 × 1200 pixels per eye, 90 frames per second). A dedicated computer (GPU: Nvidia GTX 1080) was used to render the virtual environment in real-time.

A pair of Manus VR gloves were used for finger tracking in the *avatar experiment*. The fingers were tracked based on bending sensors built in the gloves. Although these gloves also include inertial sensors to track wrist and thumb rotations, these sensors were not used in our experiments. The former was not necessary

because the wrist joint rotations could also be obtained using the Vicon system, while the latter was disabled due to the low consistency in the quality of tracking. Consequently, the thumb could only bend around a pre-defined axis relative to the palm.

Finally, facial animation was treated differently in each experiment. In the *character experiment*, an iPhone X with a custom software based on the ARKit library⁴ was used to record the facial expressions of the actor. However, this method was not suitable for controlling the avatar in the second experiment since the HMD occludes the face of the participant. Instead, the Oculus lipsync library⁵ was used for mouth animation in the *avatar experiment*. This library performs the mapping of phonemes into associated visemes, which are units describing the visual appearance of speech, analogous to phonemes. These visemes follow the MPEG-4 Standard⁶, and had to be remapped into the visemes implementation used in the face model of the avatar.

The experiments were developed using the Unity game engine version 2018.2.

3.3 Trial structure

A trial started with one of the basic configurations listed in Table 1. Then, participants had to repeatedly complete a task and perform a configuration transition until their feeling of animation realism matched that of the most complete configuration. To prevent participants from carelessly moving to the most complete configuration, we imposed the following constraints: transitions could only be made in one direction, by increasing the level of an animation feature; a transition increased the feature level by exactly 1, that is, participants could not go from ($UB = 0$) to ($UB = 2$) with a single transition, instead they had to first reach ($UB = 1$) to then transition to ($UB = 2$); participants could perform a transition only after completing a task round. To choose a feature to improve, participants were urged to reflect on the feature that they were missing the most at that moment, and to improve it first. Participants were explicitly told that the order in which they improve the animation features was important for the experiment.

TABLE 1
Start configurations used in both experiments.

Trial	Feature			
	Upper Body	Lower Body	Face	Hands
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0
5	0	0	0	1

3.4 Procedure

Both experiments followed a similar procedure, with differences on participant preparation (i.e. equipment and calibration) and experimental task. Here, we provide an overview of the general procedure, while the particularities of each experiment are presented in dedicated sections below.

4. developer.apple.com/arkit

5. developer.oculus.com/documentation/audiosdk/latest/concepts/audio-overlipsync-overview

6. www.visagetechnologies.com/uploads/2012/08/MPEG-4FBAOverview.pdf

3. www.vicon.com/products/software/shogun

Participants were received by the experimenter and asked to read an information sheet and to sign an informed consent form. Then, participants filled in a characterization form containing questions about their background and physical characteristics. Participants were prepared for the experiment, and the experimenter provided an overview of the experimental task and instructions.

Participants were then exposed to the most complete condition for each experiment. A practice trial was initiated right after, the initial configuration was pooled from one of the possible start conditions presented in Table 1. During the practice trial, the experimenter reviewed the step by step instructions of the trial as participants performed it, and certified that participants understood that they were supposed to inform the experimenter if, at any moment, the animation felt as realistic as in the most complete condition.

The valid trials started immediately after the end of the practice trial. There were a total of five valid trials, each starting with a different initial configuration. One trial started with all animation features at the minimum, while the other 4 trials started with one of the 4 animation features at level 1, as shown in Table 1. The presentation order of the trials was randomized.

Finally, participants filled in the SUS presence questionnaire [14] and participated in a short debriefing session with the experimenter. They were asked whether they were able to describe how each animation feature transition was effectively affecting the animation of the virtual body.

3.5 Character animation experiment

The preparation procedure consisted in sitting the participants in front of a physical table which was spatially aligned to a table in the virtual environment, and equipping them with the HMD and a pair of Oculus Touch motion controllers. In the virtual environment, the point-of-view of the participant was located behind a one-way mirror with a view to an interrogation room. The controllers were represented by virtual replicas, so that the participants could locate their hands. The participants had no other graphical representation of themselves in this experiment (i.e. they did not have a virtual body).

During the experiment, participants had to repeatedly complete a task. The task was divided into two parts, in the first part they had to watch a 34 seconds long animation clip, in the second part they had to answer a question.

The animation clip showed a police inspector (the character) presenting a murder case. In the animation, the inspector walked into the interrogation room, and showed a photo of the victim while describing the case to the participant (Fig. 1). This animation clip was produced with the equipment described in Section 3.2 and the inspector was portrayed by a professional actor.

At the end of the animation clip, the one way mirror turned black and the participant was prompted to select an animation feature option by the question “What would you improve first to make the animation more realistic?” (Fig. 2). Participants could either improve an animation feature, or state that the current animation condition felt equivalent to the most complete configuration. To select an animation improvement, participants had to place the touch controller over the desired option for 2 seconds.

The task was repeated either until the participant stated that the character animation felt equivalent to the most complete configuration or the absorbing animation condition $\{2, 2, 1, 1\}$ was

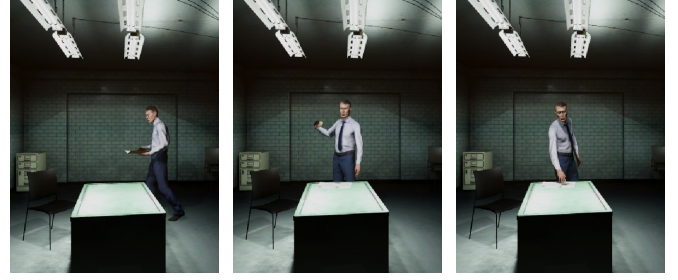


Fig. 1. Overview of the animation clip used in the character animation experiment, the character walked to the table in the interrogation room and described the crime case to the participant while interacting with objects in the environment.

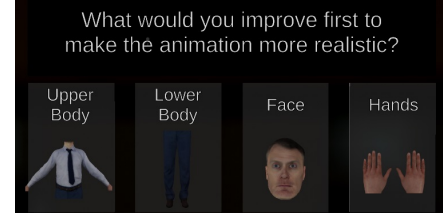


Fig. 2. After watching the character animation, the participant had to decide whether and which animation feature to improve next.

reached. A total of 13 participants took part in the experiment (5 females, mean age and standard deviation: 35 ± 11 years), they were recruited from inside and outside our institution.

3.6 Self-avatar Animation Experiment

The preparation process consisted of a few steps. First, participants dressed the motion capture suit, shoes and gloves, and had the retro-reflective markers carefully placed on the suit (Fig. 3a). Then, participants had to wear the HMD and complete the two steps body calibration. Finally, the ManusVR gloves were calibrated by performing a sequence of hand gestures.

The HMD orientation tracking was performed using the built-in inertial sensors and corrected for angular velocity integration drift around the vertical axis using Vicon’s tracking of a set of markers placed on the device. With this approach, absolute tracking (Vicon) information is used to slowly correct the drift error accumulated by the inertial sensors while still taking advantage of the lower update latency of the sensors. HMD position tracking was performed solely based on Vicon’s optical tracking.

During the experiment, participants had to repeatedly complete a task in the virtual environment. The virtual environment contained a mirror (2 x 2 meters), footprints indicating the place that participants should take during parts of the experiment, and an object that participants were able to touch, grab and move. This object was collocated with a physical motion tracked counterpart (Fig. 3b). The task was divided into two sub-tasks and two questions that the participants had to answer.

The first sub-task consisted of stepping onto a pair of footprints on the floor and, while facing the mirror, repeating the phrase “my name is (a), and I am feeling (b)”, where participants were asked to replace (a) with their names, and (b) by expressing how they were feeling about the experience and the virtual avatar (Fig. 4a). The experimenter presented words such as “good”, “bad”, “weird”, “OK”, “worst” and “better” as adequate alternatives but did not constrained participant’s choices to these words. Longer sentence

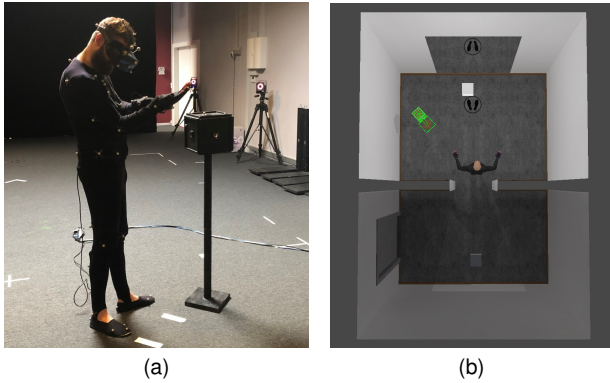


Fig. 3. Participant wearing the HMD, ManusVR gloves and the motion capture suit with retro-reflective markers (a) and an overview of the virtual environment (b) of the avatar animation experiment.

formulations were also allowed. This piece of information was noted by the experimenter and used to understand if something was not working as expected. The second sub-task consisted of grabbing and moving an object from its current location to a new location. The new location was indicated by a spotlight. The participants were asked to avoid grabbing the object with palms facing up since it could produce markers occlusions. Participants were also asked to look at their hands while carrying the object (Fig. 4b).

The first question concerned how their feeling of control over the virtual body had changed when comparing the current experience to the immediately previous one. Participants had to agree or disagree to the statement “I experienced an increased feeling of control over the virtual body” in a 5-point likert scale (Fig. 5a). When starting a new trial, this meant to compare a low fidelity configuration with a higher fidelity configuration from the last trial. Participants were expected to disagree with the statement in this situation, and the answer to this question was used to ensure that participants were paying attention to the experiment. Then, after the first response in a trial, the comparison concerned the difference felt in control due to the most recent animation feature transition. Finally, the participant was prompted to select an animation transition option with the question “What would you improve first to make the animation more realistic?” (Fig. 5b). Participants could either improve an animation feature, or state that the animation realism already felt equivalent to that of the initial configuration.

The task was repeated until participants stated that the animation realism felt equivalent to the initial configuration or the absorbing animation configuration $\{2, 2, 1, 2\}$ was reached.

The virtual avatar was created using the tool Morph 3D⁷. We choose a realistic appearance, of physiologically plausible proportions and a neutral facial expression (Fig. 4). The avatar wore a black suit resembling the motion capture suit. The original skeleton rig was replaced to match the standard humanoid figure provided with the Vicon Shogun software. The same avatar was used for all participants and no personalization, other than matching the virtual body proportions to the participant, was carried. We restricted participation to male individuals to reduce the complexity of setting up and running the experiment. In this case, only a male virtual avatar had to be prepared.

7. <https://forum.unity.com/threads/released-morph-character-system-mcs-male-and-female.355675/> - asset no longer available



Fig. 4. Overview of the two subtasks, in (a) the participant walked to the mirror and repeated a phrase, in (b) the participant had to grab an object and carry it to the place highlighted by a spotlight.

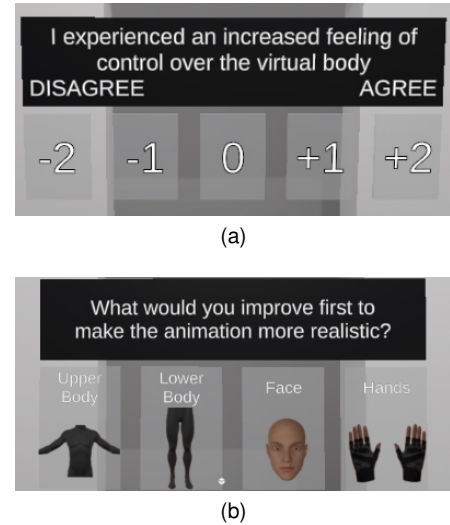


Fig. 5. After completing the task, the participant had to report whether the latest change has increased the feeling of control over the virtual body (a), and decide whether and which feature to improve next (b).

A total of 24 male participants took part in the experiment (mean age and standard deviation: 36 ± 10 years), they were recruited from inside and outside our institution, and 12 of them practice a professional activity in the VR industry.

3.7 Response Variables

We recorded three response variables: the configuration in which participants declared a match of animation realism with the most complete configuration, i.e. the matching configuration; the sequence of transitions that participants took to reach it; and, in the *avatar experiment*, the scores to the sense of control statement presented after each animation configuration transition in 5-point Likert scale (Fig. 5a).

With the matching configurations, we investigated which animation features were judged to be unnecessary or went unnoticed by participants, i.e. the animation features that were unlikely to be active in the matching configuration. With the transitions, we investigated the order in which participants judged to be optimal in improving the experience of animation realism. We assumed that the animation features adding the most to the participant’s plausibility illusion (Psi) would be selected early in the trials. Finally, the sense of control statement was used to evaluate the influence of each self-avatar animation feature to one of the aspects related to the sense of embodiment.

Lastly, we also used the answers to the post-experiment SUS presence questionnaire [14] to investigate the relationship between overall place illusion (PI), as acquired with the SUS questionnaire, and the *matching configurations* and number of *transitions*.

4 RESULTS

4.1 Analysis

To analyze the matching configurations and the transitions made to reach these configurations, we make the simplifying assumption that the five trials performed by each participant are statistically independent. However, these trials cannot be truly independent as they were performed by the same participant, who is subject to a learning effect and individual preferences. This limitation is partially addressed in our experiment design by starting each of the five experimental trials with different basic configuration (Table 1). Participants were informed that the initial configuration was not going to be the same across all trials and, therefore, that they should not expect a sequence of animation feature selections to result in the same animation configuration from trial to trial. Instead, we emphasized that they should decide based on their current observations of the character or self-avatar animation.

Moreover, in the original study [10], authors also carried a correlation test to verify if characteristics of the trials performed by the same participants are correlated to each other. This can be achieved by constructing a table containing the number of transitions, arranged with participants as the rows (1 to 13 in the *character experiment*, and 1 to 24 in the *avatar experiment*) and trial number as the columns (1 to 5). If columns are correlated, it indicates that participants have presented a consistent behavior across trials and, thus, to some extent, the behavior across trials depended on the participant. For the *character experiment*, we found one statistically significant correlation (i.e. $p < .05$) out of the ten correlation tests. For the *avatar experiment*, we found that all correlations were significant (i.e. $p < .05$) and, thus, that participants performed a consistent number of transitions across trials. These results are discussed further in Section 5.8.

Lastly, the response to the sense of control statement (*avatar experiment* only) was evaluated per participant instead of per trial and, thus, is not concerned by the discussion above.

4.2 Matching Configurations

As originally described in [10], we use the matching configurations UB, LB, FA, HA obtained in the experiments to estimate the joint probability distribution $P(u = UB, l = LB, f = FA, h = HA | match)$, where $P(u, l, f, h | match)$ represents the probability of configuration UB, LB, FA, HA given a *Psi match*, for all supported combinations of animation features. That is, we estimate the probability of having a given animation configuration when a match is declared. Then, using Bayes' theorem, we estimate the probabilities $P(match | u, l, f, h)$ that participants will declare a match when experiencing any given configuration UB, LB, FA, HA . Lastly, we used $P(u, l, f, h | match)$ to assess the marginal probability that a given animation feature will be active in the matching configuration. For instance, $P(u = 2 | match) = 0.6$ describes that, given a *Psi match*, there is a probability of 60% that the upper body animation feature is at level ($UB = 2$).

Fig. 6 presents the probabilities for $P(match | u, l, f, h)$ and $P(u, l, f, h | match)$ as estimated in the character and self-avatar animation experiments, respectively. Only configurations with a

TABLE 2
Probability of any given animation feature level in the matching configuration.

Experiment Level	Character experiment			Avatar experiment		
	0	1	2	0	1	2
Upper Body (UB)	.077	.477	.446	.117	.283	.600
Lower Body (LB)	.015	.815	.169	0	.575	.425
Face (FA)	.046	.954	-	.133	.867	-
Hands (HA)	.185	.815	-	.025	.05	.925

probability $P(u, l, f, h | match)$ above 0.02 and with 6 or more occurrences in the *character experiment* and 10 or more occurrences in *avatar experiment* are presented.

In the character animation experiment, three configurations preceding the absorbing condition $\{2, 2, 1, 1\}$ achieved a probability above 50% of being accepted as a matching configuration, $P(match | u = 1, l = 1, f = 1, h = 1) = .73$, $P(match | u = 2, l = 1, f = 1, h = 0) = .61$ and $P(match | u = 2, l = 1, f = 1, h = 1) = .8$ (Fig. 6a). Moreover, the marginal probabilities that a given character animation feature level will be active at the matching configuration are presented in Table 2. Animation level 1 had the highest marginal probability for all four animation features. The feature to be accepted without improvements the most often was the hand, with 18.5% of the trials finishing with $HA = 0$, the fingers of the character do not move in this configuration.

In the self-avatar animation experiment, two configurations preceding the absorbing configuration $\{2, 2, 1, 2\}$ achieved a probability equal or above 50% of being accepted as a *Psi match*, $P(match | u = 0, l = 2, f = 1, h = 2) = .5$ and $P(match | u = 2, l = 1, f = 1, h = 2) = .66$. In addition, two configurations achieved match probabilities close to 50%, $P(match | u = 1, l = 1, f = 1, h = 2) = .49$ and $P(match | u = 2, l = 1, f = 0, h = 2) = .46$ (Fig. 6b). Moreover, the marginal probabilities that any given self-avatar animation feature level will be active at the matching configuration are presented in Table 2. Lower body and face animation were most often set to level 1, while upper body and hands were most often set to level 2. Face was the feature with the highest probability of being accepted at the most basic level ($FA = 0$), with 13.3% of the trials finishing without facial animation. We should note that, since each feature started at level 1 (instead of level 0) in one of the five trials (Table 1), it was not always possible to complete a trial with a particular feature set to level 0.

4.3 Transitions

Participants were told that the transition order was important and that, when a transition was desirable, they should favor the feature that they miss the most. Using the record of transitions between configurations we can reconstruct the path that participants took to go from an initial configuration to a matching configuration in any particular trial. Then, considering all experimental trials, we can estimate the probability of performing any specific transition among the available options.

The diagrams in Fig. 7 present the configuration transitions recorded in each of the experiments. The graphs describe all recorded transitions, from nodes in the left to nodes in the right, until a match was declared or the absorbing condition (rightmost nodes) was reached and transitions were no longer possible. Gray shaded nodes are configurations that were not visited in any trial of each particular experiment. The width and shading of the lines represent the proportion of particular transition paths taken by

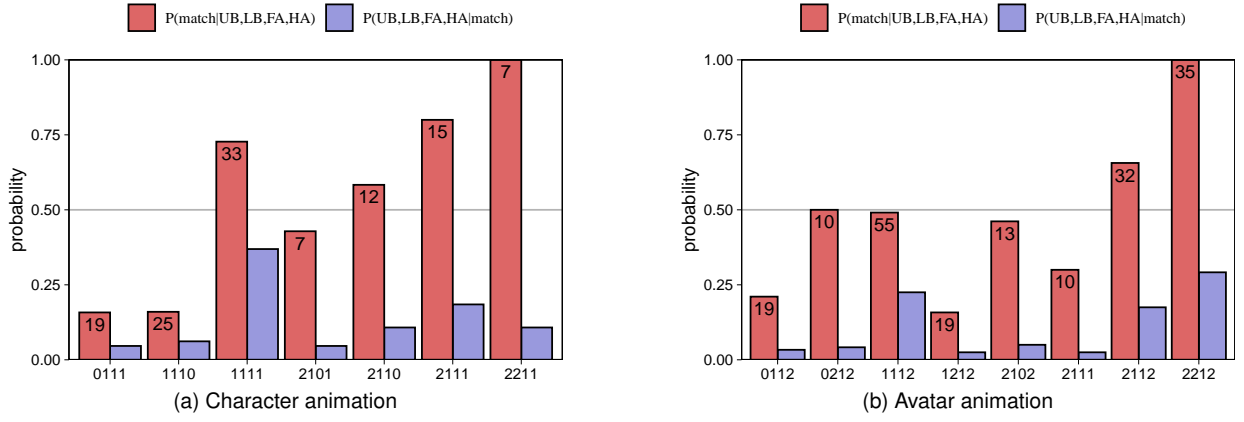


Fig. 6. Matched configurations and related probabilities for the character animation (a) and the self-avatar animation (b) experiments. The probabilities of a Psi match for different configurations, or $P(\text{match}|u, l, f, h)$, are presented in red (numbers in the scale represent n). The probabilities of specific configurations at a Psi match, or $P(u, l, f, h|\text{match})$, are presented in blue.

participants relative to the totality of transitions in the experiment. As a consequence, the wider and darker a connection is, the more times that specific transition was observed in the experiment.

For each experiment, we model the probability of transitioning from any given configuration to another as a Markov chain, which can be represented as the transition matrix \mathbf{P} of probabilities where the element \mathbf{P}_{ij} of the matrix describes the probability of a transition from configuration i to configuration j . The character animation and the self-avatar animation experiments produced transition matrices of dimension 36×36 and 54×54 , respectively. However, due to the transition constraints imposed by our experimental design, these matrices are very sparse. In fact, the former contains only 50 non-zero elements while the latter contains only 84 non-zero elements.

Let \mathbf{v} be the $1 \times n$ vector, with n equal the number of rows in \mathbf{P} , representing configuration $\{0, 0, 0, 0\}$. That is, a vector with a 1 in the element corresponding to configuration $\{0, 0, 0, 0\}$ and 0 elsewhere. Then, using the operation \mathbf{vP}^k we can estimate the probability of attaining any given configuration after k transitions for the initial configuration $\{0, 0, 0, 0\}$. Lastly, the most probable configurations in the character animation experiment after $k \in \{1, 2, 3, 4, 5\}$ transitions and the most probable configurations in the avatar experiment after $k \in \{1, 2, 3, 4, 5, 6\}$ transitions are presented in Fig 8a and b, respectively. We emphasize that participants could perform up to six transitions in the character animation experiment and seven transitions in the self-avatar animation experiment, the figures omit the last transition since the absorbing configuration was the only option available at that stage. Note that only probabilities above 5% are presented in the figure.

It is possible to observe that the most probable sequence of transitions in the character animation experiment was: $\{0, 0, 0, 0\} \rightarrow \{0, 1, 0, 0\} \rightarrow \{0, 1, 1, 0\} \rightarrow \{1, 1, 1, 0\} \rightarrow \{1, 1, 1, 1\} \rightarrow \{2, 1, 1, 1\} \rightarrow \{2, 2, 1, 1\}$. This path results in a single improvement for Lower Body, Face, Upper Body and Hands respectively, followed by maxing the Upper body and then Lower body. Moreover, the most probable sequence of transitions in the self-avatar animation experiment was: $\{0, 0, 0, 0\} \rightarrow \{0, 1, 0, 0\} \rightarrow \{1, 1, 0, 0\} \rightarrow \{1, 1, 0, 1\} \rightarrow \{1, 1, 0, 2\} \rightarrow \{1, 1, 1, 2\} \rightarrow \{2, 1, 1, 2\} \rightarrow \{2, 2, 1, 2\}$. This path describes a single improvement for Lower and Upper Body, followed by maxing the Hand feature to achieve fingers movement control, then maxing the Face animation feature, and finally maxing

Upper and Lower Body.

4.4 Sense of Control

The change in sense of control of the self-avatar (*avatar experiment*) for a given animation factor and level was assessed with the response to the statement “I experienced an increased feeling of control over the virtual body”. Answers were provided in a 5-point scale from -2 to 2, where -2 means “Disagree” and 2 means “Agree”. The summary of results, with the average score per participant, is shown as a box and whiskers plot in Fig. 9. We run the Wilcoxon signed-rank test to identify whether participants generally agreed that the features added to the sense of control of the virtual body. The responses expressed statistically significant agreement with the statement (i.e. $p < .05$) for all animation feature improvements except for $HA = 1$ ($p > .7$). In addition, the highest agreement was obtained for $LB = 1$.

5 DISCUSSION

5.1 Foot tracking is of the highest priority

Lower body tracking received the highest priority in both experiments. When, at the start of a trial, Lower Body animation was procedurally generated ($LB = 0$, no foot tracking), participants improved (LB) from $LB = 0$ to $LB = 1$ first $\approx 77\%$ (40/52) of the time in the *character experiment* and $\approx 95\%$ (91/96) of the time in the *avatar experiment*. In addition, one out of 52 trials starting with $LB = 0$ terminated without any lower body improvement in the *character experiment* (Table 2), while none of the 96 trials starting with $LB = 0$ terminated without at least one lower body improvement in the *avatar experiment* (Table 2). A chi-square test shows a dependency between the distribution of the sum of occurrences of (LB) levels (0, 1 and 2) at the matching configurations and experiment type (character or self-avatar animation, $\chi^2_{(2)} = 13.8$ $p < .01$). Notably, participants in the *character experiment* were satisfied with $LB = 1$ more often than participants in the *avatar experiment* (81.5% compared to 57.5%), while participants in the *avatar experiment* requested the maximum level of lower body animation ($LB = 2$) more often (42.5% compared to 16.9% in the *character experiment*).

Participants in the *character experiment* often reported that, with $LB = 0$, the movement looked unnatural, and even robotic.

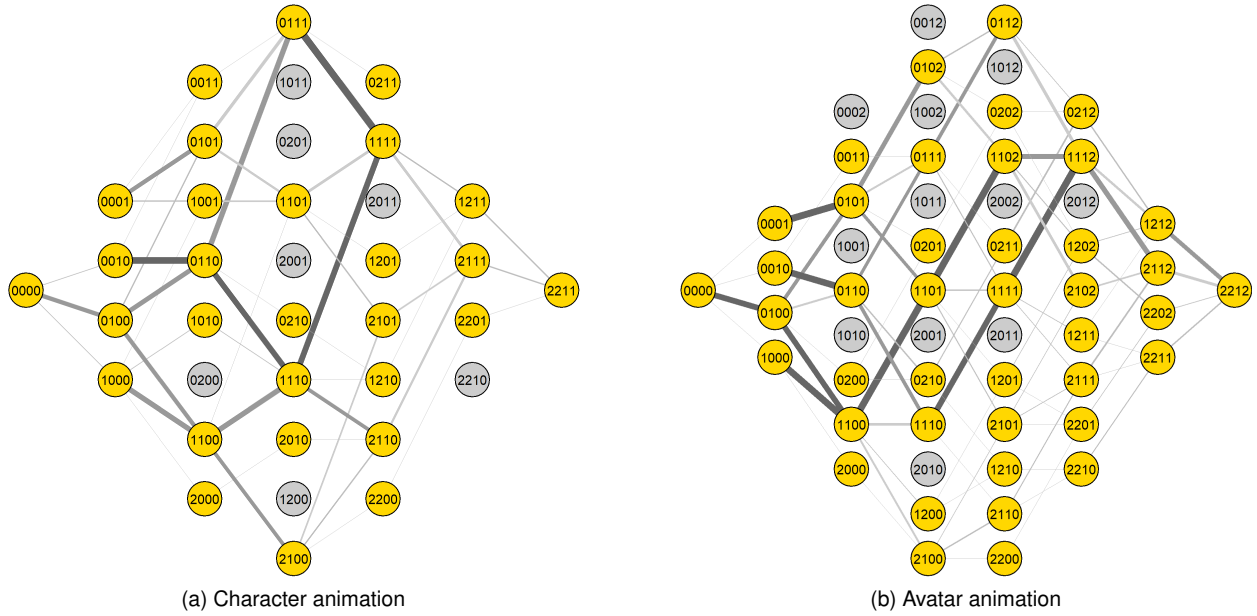


Fig. 7. Diagram representation of the configurations and transitions taken in the experiments. The nodes represent all possible $\{UB, LB, FA, HA\}$ configurations in the character animation (a) and self-avatar animation (b) experiments. Participants could only move from nodes in the left to nodes in the right. The width and shading of the lines represent the proportion of trials taking a particular transition path proportional to the total trials in that experiment, i.e. the wider and darker a connection is, the more often that transition was observed in the experiment. Gray shaded nodes are configurations that have not been visited in any of the trials.

They also reported that, as the legs of the character were partially occluded by a table during part of the animation clip, this feature could be even more relevant in a different situation. Participants in the *avatar experiment* reported problems with the unnatural dynamic of the procedurally generated leg movements ($LB = 0$), akin to the first experiment. In addition, they reported the feeling of strangeness and surprise in seeing the virtual lower body moving asynchronously to their own movements. That is, seeing the virtual legs moving when their legs are still and vice versa. We could observe that the segment of the task in which participants had to walk over a pair of footprints on the floor was destabilizing as participants tended to look at the floor and realize that, in spite of the fact that they knew that their real feet were over the footprints, the feet that they saw were in a different position. This behavior suggests that the use of procedural lower body animation was particularly detrimental to self-avatar animation. In fact, participants generally registered the highest agreement with the sense of control statement right after improving this particular feature (Fig. 9).

5.2 Body animation with eight tracked body parts

For the *character experiment*, configurations containing a combination of ($UB \leq 1$) and ($LB \leq 1$) were accepted as a match in 49% (32/65) of the trials. That is, participants in the first experiment deemed the simplified body tracking setup, relying on feet, pelvis, hands and head position and rotation, plus elbows positions, to appear equivalent to the professional capture pipeline nearly half of the time. Taken individually, Table 2 shows that, at the matching configuration, both upper and lower bodies (UB and LB) presented higher marginal probability of being set to level 1 or lower than to level 2. In fact, the particular configuration $\{1, 1, 1, 1\}$ was accepted as equivalent to the absorbing configuration the more times in the experiment (24 out of 65 trials, Fig. 6, blue bars).

For the *avatar experiment*, configurations containing a combination of ($UB \leq 1$) and ($LB \leq 1$) were accepted as a matching configuration in 32% (38/120) of the trials. That is, nearly one third of the trials were interrupted at a point where neither upper body nor lower body were using the body pose reconstruction computed with the Vicon system. Based on feedback from users, a common issue of $UB = 0$ was that the chest felt stiff, and was often visible at a location that was not congruent with the actual chest of the participant. This is expected given that the pelvis tracking information is not available and that the pose of the whole spine, including the chest, is controlled by the pose of the HMD. In fact, pelvis information only became available with $UB \geq 1$ or $LB = 2$. Once both UB and LB were set to 1, most of the salient body issues, such as lower body movements being controlled by the HMD, the incongruent position of lower body, spine and elbows, and the stiffness of the upper body, were solved (as detailed in Section 5.3). The fact that, after the absorbing configuration, the configuration that was perceived to match the more times was $\{1, 1, 1, 2\}$ seems to support this observation (27 out of 120 trials, Fig. 6, blue bars).

Compared to the character, Table 2 and Fig. 6 show that participants in the *avatar experiment* requested more animation improvements before a match and, thus, were able to perceive a difference between the best configuration ($\{2, 2, 1, 2\}$) and a simplified configuration more clearly. A chi-square test showed a dependency between the total of UB level occurrences (0, 1 or 2) at the matching configurations and the experiment type (character or self-avatar animation, $\chi^2_{(2)} = 7$ $p < .05$). Notably, participants in the character animation experiment were satisfied with $UB = 1$ more often than participants in the *avatar experiment*. This might be partially due to the fact that, in the avatar animation experiment, participants could control the virtual body and monitor for the mismatch between intended and resulting movements. In

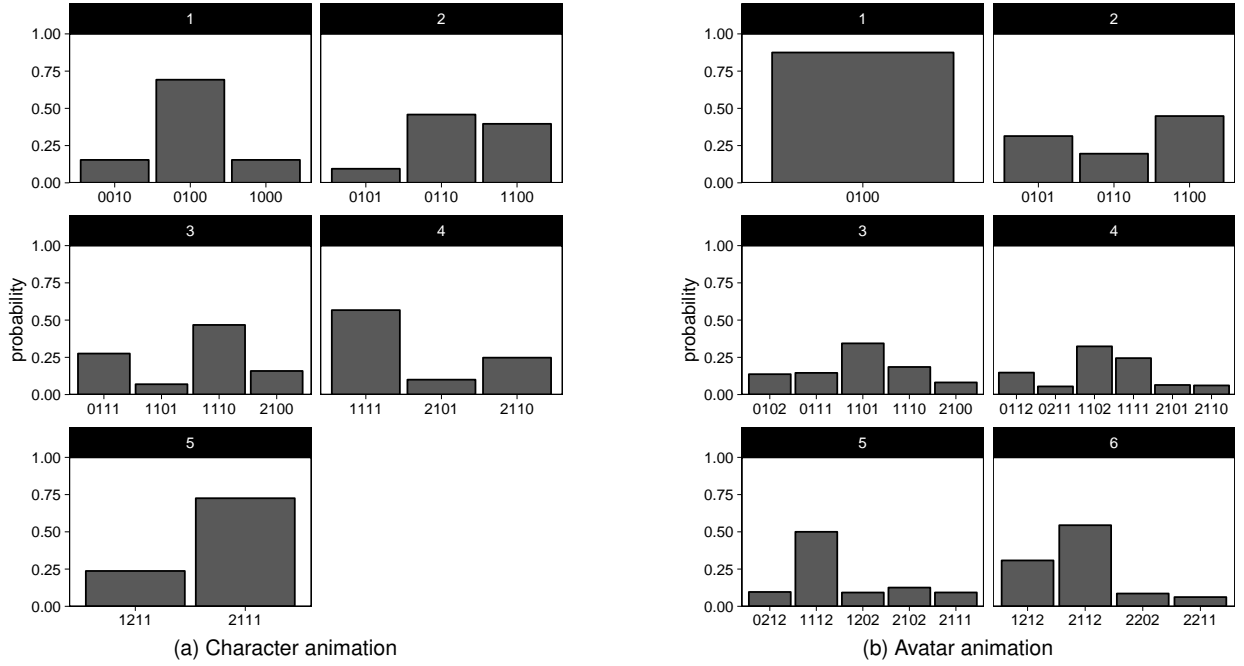


Fig. 8. Probability distribution of any given configuration UB, LB, FA, HA after each transition for the initial condition 0,0,0,0. Only probabilities greater than 0.05 are presented.

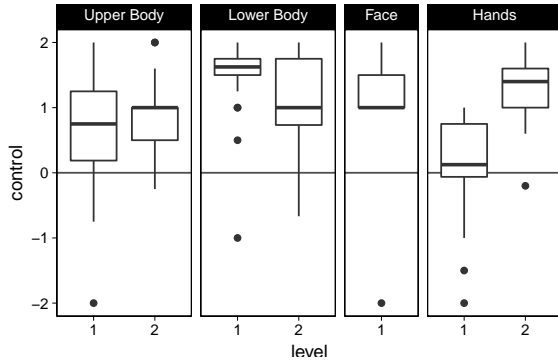


Fig. 9. Box and whiskers plot of the per participant mean agreement to the statement "I experienced an increased feeling of control over the virtual body", where 2 means "Agree" and -2 means "Disagree".

such situations, realistic body animation is not sufficient since the segments of the avatar and user bodies do not move coherently (i.e. the animation fidelity is low). For instance, shoulders could only be directly controlled by the participant with $UB = 2$ while at $UB \leq 1$ the shoulder is controlled by inverse kinematics to satisfy the pose of the (tracked) end-effectors.

5.3 Interaction between upper and lower body

With $LB = 0$ and $UB = 0$, there is no tracking information about the elbows, shoulders, pelvis, spine and lower body. Spine, pelvis, and feet are controlled by the HMD tracking, and feet stepping tries to ensure the stability of the center of mass of the virtual body (details on [35]). The feet also rotate to face approximately the same direction as the HMD, which seemed to disturb participants in the *avatar experiment*. In fact, three of them reported having noticed a connection between the HMD or spine with leg movements or feet direction. Moreover, the lack of elbow tracking could cause

interpenetration between arms and torso. This limitation seemed to disturb participants in the *character experiment* more often. In addition, the stiffness and incongruent pose of the chest was often reported as a problem by participants in both experiments.

With $LB = 1$ and $UB = 0$, real feet tracking information is available, which improves the behavior of the legs (participants in the *avatar experiment* get to control the lower body, participants in the *character experiment* report a more natural movement) and reduces the range of movement of the (still not tracked) pelvis, which can also improve the behavior of the spine. Yet, the movement of the spine is still stiff and, for the *avatar experiment*, incongruent with the movement of the participant since it relies on tracking that is a few joints away in the kinematic chain (i.e. HMD and pelvis exert control over the spine).

Most of these issues are solved with $LB = 1$ and $UB = 1$, when the tracking of pelvis and elbows become available. This enforces a reduced range of movement to the spine and results in more realistic behavior. It also prevents the interpenetration of arm segments with the spine, which was reported by a few users for $UB = 0$. It is worth noting that $LB = 2$ and $UB = 0$ would also produce similar results for the spine control since pelvis tracking also becomes available when $LB = 2$. But this would not improve the control of the arms of the virtual body.

With $LB = 1$ and $UB = 2$, the movement of the spine and shoulders are controlled by the full tracking data of these body parts. Thus, movement becomes smoother and behaves according to the participants' movements, as also reported by participants in the *avatar experiment*. However, this level of tracking was often unnecessary, in particular for the *character experiment* ($UB = 2$ in 44.6% and 60% of the matching configurations in the *character* and *avatar* experiments, respectively). Finally, in both experiments, the simultaneous activation of both $LB = 2$ and $UB = 2$ was normally not requested to declare a match (active in 10.8% and 33.3% of the matching configurations in the *character* and *avatar* experiments,

respectively).

5.4 Face and finger movements received higher priority than the best upper and lower body tracking

In the *character experiment*, both face and finger animation were active ($FA = 1$ and $HA = 1$) in 70.1% of the matching configurations. Notably, transition results (Section 4.3 and Fig. 8) show that the face was likely to be animated even before any improvement to the upper body (e.g. while $UB = 0$), while fingers were more likely to be animated before either upper or lower body could reach level 2. Face and finger animation were also important in the *avatar experiment*, and both features were maxed out ($FA = 1$ and $HA = 2$) in 79.2% of the matching configurations. However, setting lower and upper body, in this order, to level 1 ($LB = 1$ then $UB = 1$) first was more important than improvements to FA and HA . Once lower and upper body animation were set to 1, participants were likely to apply a sequence of two improvements to the hand tracking, leading to the control of finger movements ($HA = 2$). Only then the most likely improvement would become the face animation ($FA = 1$). Notably, maxing the available FA and HA options had higher priority for participants than maxing either lower or upper body, as evident from the probability distributions in Fig. 8.

Furthermore, procedural finger animation was explicit to participants in the *character experiment*, but not to participants in the *avatar experiment* (self-avatar animation). For the latter, the $HA = 1$ was the only feature level that did not seem to increase the sense of control of the virtual body (Fig. 9). Moreover, only a few of the participants were able to describe what the effect of $HA = 1$ was during the debriefing session.

5.5 Differences between the experiments

In spite of the fact that the design of the trials of both experiments forced participants to experience events that would stress the faulty features, many factors can play a role on the comparative results. The tasks and virtual environment settings were different across the experiments, affecting the context. Moreover, by controlling a virtual body, participants in the *avatar experiment* had the autonomy to act and control aspects of their experience. They could compare differences between expected and realized outcome of their actions to determine body animation features that felt unrealistic and detrimental to animation plausibility, while participants in the *character experiment* could only see a single animation sequence of the character and could not assess sensorimotor coupling. In addition, the pace of the *character experiment* was enforced by a cinematic segment while the pace of the *avatar experiment* was controlled by participants, which can lead to differences in the exposure time and more opportunities to be affected by the animation defects in the *avatar experiment*. Finally, the possible animation configurations were not identical across experiments, face animation was carried with different capture techniques and hand tracking had an additional level in the *avatar experiment*. This could affect the priority of animation feature improvements.

We argue that the extent to which an incorrect aspect of the simulation, such as a movement behavior, is likely to break scene coherency for an user and affect plausibility is proportional to the chances that this user will get to observe that behavior developing. All of the aforementioned aspects will interfere with those chances and, thus, we must assume that all of them might have played a role in our results.

5.6 Previous experience and presence questionnaire

Half of the participants (twelve) in the avatar animation experiment develop a professional activity in the VR industry. To compare this group with the other participants, we estimate the average sum of active animation feature levels per participant and run a Wilcoxon signed-rank test. The test failed to reject the equivalence of the groups ($p = .726$). Thus, we found no clear evidence that the group of experienced users required higher levels of animation fidelity to be satisfied.

Moreover, we used the pre-experiment characterization form to expand in this analysis. The form includes four questions concerning the previous experience of the participants with VR, HMDs, motion controllers and games, all on a 5 point scale. We found a positive correlation between the sum of the score of the four questions and the sum of active levels in the matching configuration ($r_{22} = .55$, $p = .005$) for the avatar animation experiment but not for the character animation experiment ($r_{11} = -.33$, $p = .28$). Thus, we found evidence that users with more immersive equipment experience require a more complete self-avatar animation setup, but results were not consistent enough to draw strong conclusions.

Lastly, we did not find a statistically significant correlation between the sum of the presence questions (SUS questionnaire) and the sum of the animation levels at the match condition in any of the experiments (character animation: $r_{11} = .19$ $p = .543$, avatar animation: $r_{22} = -.02$ $p = .943$).

5.7 Virtual embodiment vs plausibility illusion

The topic of virtual embodiment with controlled self-avatars, like the avatar implemented for the second experiment, is one that is closely related to the quality of experience of VR users and, we argue, might happen simultaneously to and interact with an enhanced level of presence. As noted earlier, previous empirical research have shown that having and controlling a body affect PI, Psi [10], as well as the sense of embodiment and ownership (i.e. that the virtual body I see and control is my body) of a virtual body [1], [9]. Thus, it is clear that the *avatar experiment* was also capable of driving an enhanced sense of embodiment of the virtual avatar, although participants were explicitly told to reflect on aspects related to animation realism.

Whether plausibility and sense of embodiment can be dissociated is a complex research question in itself and, although we designed an experiment to assess the Psi of controlling a body (*avatar experiment*), this is not to say that embodiment had no role, or was not correlated to, or was not a requirement for Psi of that body. For instance, it is possible that what makes an animation plausible is the same of what makes one feel embodied in an animated virtual body, as correlated events, or that to say that a body that I can control moves in a way that is plausible may imply that one has to feel embodied in that body, so that it can feel plausible (or vice versa), resulting in a causal relationship. Answering such questions would entail the design of an experiment capable of isolating these variables, but at this point, it is not clear if one can represent a virtual body that one can control and experience in a natural way and not feel embodied at the same time.

5.8 Limitations

The analysis of transitions and match probabilities make the simplifying assumption that the trials are statistically independent. However, since participants carried out five trials each, they were

subject to a learning effect, and it is unlikely that the trials are truly independent. In fact, we found that participants in the *avatar experiment* often performed a similar number of transitions in the trials that they completed and, thus, that the number of transitions is dependent on participant (Section 4.1).

In comparison with the original paper (Slater et al. [10]), we believe that two main factors may have had an influence on the correlation tests results. First, the *avatar experiment* had a higher number of participants, all of which performing the same task, which can increase the statistical power of the correlation test. Second, we did not implement a budget control mechanism to penalize or prevent participants from maxing out all animation features. The budget correlation mechanism encourages participants to stop as soon as possible in order to maximise an extrinsic reward, which may make the number of transitions more uniform between the group of participants.

Concerning the implications of the lack of independence in the trials performed by the same participant on the experimental results, we believe that the configuration transitions and its probability distribution were not strongly affected since: (1) even if participants were consistent in their behavior, the number of participants was relatively high ($n=24$ in comparison to $n=10$ per experiment in the original study [10]); (2) certain choices emerged as a preference across participants (i.e. different participants converged to a small group of choices). Concerning the matching configuration results, the dependency on participants might reduce the confidence of results as a confound, but we argue that it does not render the results invalid since our study had a relatively high number of participants.

Another factor that should be taken into consideration is that the feature levels were not equivalent between the four animation features. For instance, ($UB = 0$) would include wrist and head tracking, which could generate a reasonable estimation of the pelvis, spine, shoulders and elbows, and direct control over most of the end-effectors of the body. In contrast, ($LB = 0$) did not allow for the direct control of the legs or feet, and since no tracking information about the lower body was available, these had to be inferred from the upper body. As a consequence, it is not difficult to understand why the initial improvement of the lower body, from ($LB = 0$) to ($LB = 1$), had higher priority than the improvement of the upper body. We should emphasize, however, that this is a consequence of the experimental design, which attempted to contemplate the challenges in the implementation of motion tracking technology in the distribution of factors. For instance, ($FA = 0$) and ($HA = 0$) had yet another behavior that was different than both ($UB = 0$) and ($LB = 0$) (i.e. no movement whatsoever), but in this case, the trend towards a transition choice becomes less obvious. In this context, our experiments proposed to answer how different animation features, that are so different in terms of implementation and hardware, compete for an improved experience of realistic character and self-avatar animation.

6 CONCLUSION

In this paper, we presented two experiments designed to evaluate the relative importance of different virtual body animation features on the plausibility illusion. Our results suggest that, from the perspective of a user, realistic full-body animation is possible by tracking a relatively small set of body parts. Notably, a total of 8 trackers (hands, elbows, feet, pelvis and head) was often accepted as equivalent to the most complete capture pipeline relying on 53

retro-reflective markers. When taking commercial VR sets as a reference, which include an HMD and a pair of tracked controllers, the addition of foot tracking added the most to the experience of controlling and observing an animated virtual body.

These results help assessing the impact of different animation features on the experience of the user and provide meaningful insights on the requirements of new consumer tracking equipment. Notably, by reporting on information that can be used to estimate the trade-offs between user experience and cost of implementation, our results may help to direct the development of full-body motion capture efforts in different VR application contexts such as consumer VR equipment and location-based VR experiences.

We must emphasize, though, that animation is a complex topic, and the experiments and results presented in this paper should be taken into consideration within the context in which they were performed. There are many variables that can affect and interact in human perception, and the task of evaluating a system may often not generalize well to a broader context.

ACKNOWLEDGMENTS

This work was funded by the European Commission as part of the H2020 program, under the grant agreement 762111, *VRTogether*.

REFERENCES

- [1] M. Gonzalez-Franco, D. Perez-Marcos, B. Spanlang, and M. Slater, "The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment," in *2010 IEEE virtual reality conference (VR)*. IEEE, 2010, pp. 111–114.
- [2] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic, "Characterizing embodied interaction in first and third person perspective viewpoints," in *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2015, pp. 67–72.
- [3] E. Kokkinara and R. McDonnell, "Animation realism affects perceived character appeal of a self-virtual face," in *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. ACM, 2015, pp. 221–226.
- [4] T. J. Dodds, B. J. Mohler, and H. H. Bühlhoff, "Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments," *PloS one*, vol. 6, no. 10, p. e25759, 2011.
- [5] R. Fribourg, F. Argelaguet, L. Hoyet, and A. Lécuyer, "Studying the sense of embodiment in vr shared experiences," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2018, pp. 273–280.
- [6] P. Rubin. (2018) Enter the dreamscape: Location-based vr gets a new player. [Online]. Available: <https://www.wired.com/story/dreamscape-immersive>
- [7] DVgroup. (2017) Alice the virtual reality play. [Online]. Available: <https://alice.dv.fr>
- [8] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009.
- [9] K. Kiltner, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012.
- [10] M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 92:1–92:9, Jul. 2010.
- [11] M. Slater, "A note on presence terminology," *Presence connect*, vol. 3, no. 3, pp. 1–5, 2003.
- [12] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of computer-mediated communication*, vol. 3, no. 2, p. JCMC321, 1997.
- [13] R. Skarbez, F. P. Brooks, Jr, and M. C. Whitton, "A survey of presence and related concepts," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–39, 2017.
- [14] M. Usuh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence: Teleoperators & Virtual Environments*, vol. 9, no. 5, pp. 497–503, 2000.
- [15] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr., "Physiological measures of presence in stressful virtual environments," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 645–652, Jul. 2002.

- [16] A. S. Azevedo, J. Jorge, and P. Campos, "Combining eeg data with place and plausibility responses as an approach to measuring presence in outdoor virtual environments," *Presence*, vol. 23, no. 4, pp. 354–368, Nov 2014.
- [17] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1352–1359, April 2017.
- [18] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1369–1378, April 2017.
- [19] B. Gao, J.-I. Kim, and H. Kim, "Sensory and perceptual consistency for believable response in action feedback loop," in *Proceedings of Computer Graphics International 2018*, ser. CGI 2018. New York, NY, USA: ACM, 2018, pp. 201–210.
- [20] R. Skarbez, D. A. Bowman, J. T. Ogle, T. Tucker, and J. L. Gabbard, "Virtual replicas of real places: Experimental investigations," *arXiv preprint arXiv:1812.03441*, 2018.
- [21] M. Jeannerod, *The cognitive neuroscience of action*. Blackwell Oxford, 1997, vol. 1997.
- [22] R. Fribourg, F. Argelaguet, A. Lécuyer, 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*, vol. 26, no. 5, pp. 2062–2072, 2020.
- [23] J. W. Moore, D. Lagnado, D. C. Deal, and P. Haggard, "Feelings of control: contingency determines experience of action," *Cognition*, vol. 110, no. 2, pp. 279–283, 2009.
- [24] T. I. Nielsen, "Volition: A new experimental approach," *Scandinavian journal of psychology*, vol. 4, no. 1, pp. 225–230, 1963.
- [25] N. Franck, C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, T. d'Amato, and M. Jeannerod, "Defective recognition of one's own actions in patients with schizophrenia," *American Journal of Psychiatry*, vol. 158, no. 3, pp. 454–459, 2001.
- [26] H. G. Debarba, R. Boulic, R. Salomon, O. Blanke, and B. Herbelin, "Self-attribution of distorted reaching movements in immersive virtual reality," *Computers & Graphics*, vol. 76, pp. 142–152, 2018.
- [27] R. McDonnell, M. Breidt, and H. H. Bülthoff, "Render me real?: investigating the effect of render style on the perception of animated virtual humans," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 4, p. 91, 2012.
- [28] Y. Wu, S. V. Babu, R. Armstrong, J. W. Bertrand, J. Luo, T. Roy, S. B. Daily, L. C. Dukes, L. F. Hodges, and T. Fasolino, "Effects of virtual human animation on emotion contagion in simulated inter-personal experiences," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 4, pp. 626–635, 2014.
- [29] P. Wisessing, J. Dingliana, and R. McDonnell, "Perception of lighting and shading for animated virtual characters," in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2016, pp. 25–29.
- [30] M. Volante, S. V. Babu, H. Chaturvedi, N. Newsome, E. Ebrahimi, T. Roy, S. B. Daily, and T. Fasolino, "Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 4, pp. 1326–1335, 2016.
- [31] K. Zibrek, E. Kokkinara, and R. McDonnell, "Don't stand so close to me: investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task," in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2017, p. 3.
- [32] M. Pražák, R. McDonnell, and C. O'Sullivan, "Perceptual evaluation of human animation timewarping," in *ACM SIGGRAPH ASIA 2010 Sketches*. ACM, 2010, p. 30.
- [33] L. Hoyet, K. Ryall, K. Zibrek, H. Park, J. Lee, J. Hodgins, and C. O'Sullivan, "Evaluating the distinctiveness and attractiveness of human motions on realistic virtual bodies," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 204, 2013.
- [34] J. Hodgins, S. Jörg, C. O'Sullivan, S. I. Park, and M. Mahler, "The saliency of anomalies in animated human characters," *ACM Trans. Appl. Percept.*, vol. 7, no. 4, pp. 22:1–22:14, Jul. 2010.
- [35] P. Lang. (2016) Inverse kinematics in dead and buried. [Online]. Available: <http://root-motion.com/2016/06/inverse-kinematics-in-dead-and-buried/>



Henrique Galvan Debarba received a PhD degree in Robotics, Control, and Intelligent Systems from the Ecole Polytechnique Fédérale de Lausanne, Switzerland, in 2017. He was a senior researcher at Artanim Foundation, Switzerland, a research foundation working on virtual reality and motion capture technologies. Henrique currently works as an assistant professor at the IT University of Copenhagen, Denmark. His research topics include virtual and augmented reality, human computer interaction and computer graphics.



Sylvain Chagué is an engineer graduated from the Ecole des Mines de Saint-Etienne, France. He is Co-Founder and Technology Director of Artanim, a multidisciplinary motion capture lab in Switzerland and Co-Founder and CTO of Dreamscape Immersive, a VR entertainment company working on the future of location-based VR. His areas of expertise include motion capture for 3D animation and video games, interactive human-computer interfaces, virtual and augmented reality as well as 3D body scanning technologies.



Caecilia Charbonnier obtained a PhD degree in Computer Science in 2010 at MIRALab - University of Geneva, Switzerland. She is the Co-Founder and Research Director of Artanim, a center specialized in motion capture technologies, and Co-Founder and CIO of Dreamscape Immersive, a VR entertainment company. She is also Privat-docent at the Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva. Her work focus on the interdisciplinary use of motion capture from 3D

animation, live performances to movement science, orthopedics and sports medicine. Her research interests include joint biomechanics, VR/AR, real time motion interaction and 3D body scanning.