frontiers in Virtual Reality

TECHNOLOGY AND CODE
published: 03 November 2020
doi: 10.3389/frvir.2020.561558

The Rocketbox Library and the Utility of Freely Available Rigged Avatars

Mar Gonzalez-Franco[1], Eyal Ofek[1], Ye Pan[2], Angus Antley[1], Anthony Steed[1,3], Bernhard Spanlang[4], Antonella Maselli[5], Domna Banakou[6,7], Nuria Pelechano[8], Sergio Orts-Escolano[9], Veronica Orvalho[10,11,12], Laura Trutoiu[13], Markus Wojcik[14], Maria V. Sanchez-Vives[15,16], Jeremy Bailenson[17], Mel Slater[6,7] and Jaron Lanier[1]*

[1] Microsoft Research, Redmond, WA, United States, [2] Disney Research, Los Angeles, CA, United States, [3] Computer Science Department, University College London, London, United Kingdom, [4] Virtual Bodyworks S.L., Barcelona, Spain, [5] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy, [6] Department of Psychology, Institute of Neurosciences of the University of Barcelona, Barcelona, Spain, [7] EventLab, Universitat de Barcelona, Barcelona, Spain, [8] Computer Science Department, Universitat Politècnica de Catalunya, Barcelona, Spain, [9] Google, Mountain View, CA, United States, [10] Faculdade de Ciências, Universidade do Porto, Porto, Portugal, [11] Instituto de Telecomunicações, Porto, Portugal, [12] Didimo Inc., Porto, Portugal, [13] Independent Researcher, Seattle, WA, United States, [14] Independent Researcher, Hannover, Germany, [15] Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain, [16] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain, [17] Department of Communication, Stanford University, Stanford, CA, United States*

OPEN ACCESS

Edited by:
Stefania Serafin, Aalborg University Copenhagen, Denmark

Reviewed by:
Daniel Roth, Technical University of Munich, Germany
Mark Billinghurst, University of South Australia, Australia

****Correspondence:***
Mar Gonzalez-Franco
margon@microsoft.com

Specialty section:
This article was submitted to Technologies for VR, a section of the journal Frontiers in Virtual Reality

Received: *12 May 2020*
Accepted: *16 September 2020*
Published: *03 November 2020*

Citation:
Gonzalez-Franco M, Ofek E, Pan Y, Antley A, Steed A, Spanlang B, Maselli A, Banakou D, Pelechano N, Orts-Escolano S, Orvalho V, Trutoiu L, Wojcik M, Sanchez-Vives MV, Bailenson J, Slater M and Lanier J (2020) The Rocketbox Library and the Utility of Freely Available Rigged Avatars. Front. Virtual Real. 1:561558. doi: 10.3389/frvir.2020.561558

As part of the open sourcing of the Microsoft Rocketbox avatar library for research and academic purposes, here we discuss the importance of rigged avatars for the Virtual and Augmented Reality (VR, AR) research community. Avatars, virtual representations of humans, are widely used in VR applications. Furthermore many research areas ranging from crowd simulation to neuroscience, psychology, or sociology have used avatars to investigate new theories or to demonstrate how they influence human performance and interactions. We divide this paper in two main parts: the first one gives an overview of the different methods available to create and animate avatars. We cover the current main alternatives for face and body animation as well introduce upcoming capture methods. The second part presents the scientific evidence of the utility of using rigged avatars for embodiment but also for applications such as crowd simulation and entertainment. All in all this paper attempts to convey why rigged avatars will be key to the future of VR and its wide adoption.

Keywords: avatars, virtual reality, augmented reality, rigging, animation, motion capture, blendshapes, Microsoft Rocketbox

1. INTRODUCTION

When representing users or computer-controlled agents within computer graphics systems we have a range of alternatives from abstract and cartoon-like, through human-like to fantastic creations from our imagination. However, in this paper we focus on anthropomorphically correct digital human representations: avatars. These digital avatars are a collection of geometry (meshes, vertex) and textures (images) combined to look like real humans in three dimensions (3D). When these avatars are rigged they have a skeleton system, can walk and be animated to resemble people. A human-like avatar is defined by its morphology and behavior. The morphology of an avatar refers to the definition of the shape and structure of the geometry of the 3D model, and it usually complies with the anatomical structure of the human body. The behavior of an avatar is defined by the movements the 3D model can perform. Avatars created with computer graphics can reach such

a level of realism that they can substitute real humans inside Virtual Reality (VR) or Augmented Reality (AR). An avatar can represent a real live participating person, or an underlying software agent. When digital humans are controlled by algorithms they are referred to as embodied agents (Bailenson and Blascovich, 2004).

When people enter an immersive virtual environment through a VR/AR system they may experience an illusion of being in the place depicted by the virtual environment, typically referred to as "presence" (Sanchez-Vives and Slater, 2005). Presence has been decomposed into two different aspects, the illusion of "being there," referred to as "Place Illusion," and the illusion that the events that are occurring are really happening, referred to as "Plausibility" (Slater, 2009). These illusions and their consequences occur in spite of the person knowing that nothing real is happening. However, typically, the stronger these illusions the more realistically people will respond to the events inside the VR and AR (Gonzalez-Franco and Lanier, 2017).

A key contributor to plausibility is the representation and behavior of the avatars in the environment (Slater, 2009). Are those avatars realistic? Do their appearance, behavior, and actions match with the plot? Do they behave and move according to expectations in the given context? Do they respond appropriately and according to expectations with the participant? Do they initiate interactions with the participant of their own accord? A simple example is that a character should smile back, or at least react, when a person smiles toward it. Another example is that a character moves out of the way, or acknowledges the participant in some way, as she or he walks by.

Avatars are key to every social VR and AR interaction (Schroeder, 2012). They can be used to recreate social psychology scenarios that would be very hard or impossible to recreate in reality to evaluate human responses. Avatars have helped researchers in further studying bystander effects during violent scenarios (Rovira et al., 2009; Slater et al., 2013) or paradigms of obedience to authority (Slater et al., 2006; Gonzalez-Franco et al., 2019b), to explore the effects of self-compassion (Falconer et al., 2014, 2016), crowd simulation (Pelechano et al., 2007), or even experiencing the world from the embodied viewpoint of another (Osimo et al., 2015; Hamilton-Giachritsis et al., 2018; Seinfeld et al., 2018). In many cases of VR social interaction, researchers use embodied agents (i.e., procedural avatars). Note that in this paper we do not use the term "procedural" to refer to how they were created, but rather how they are animated to represent agents in the scene, for example, following a series of predefined animations potentially driven by AI tools.

A particular case of avatars inside VR are self-avatars, or embodied avatars. A self-avatar is a 3D representation of a human model that is co-located with the user's body, as if it were to replace or hide the real body. When wearing a VR Head Mounted Display (HMD) the user cannot see the real environment around her and in particular, cannot see her own body. The same is true in some AR configurations. Self-avatars provide users with a virtual body that can be visually coincident with their real body (**Figure 1**). This substitution of the self-body with a self-avatar is often referred to as embodiment (Longo et al., 2008; Kilteni et al., 2012a). We use the term "virtual embodiment" (or just "embodiment") to describe the physical process that employs the VR hardware and software to substitute a person's body with a virtual one. Embodiment under a variety of conditions may give rise to the subjective illusions of body ownership and agency (Slater et al., 2009; Spanlang et al., 2014). Body ownership is enabled by multisensory processing and plasticity of body representation in the brain (Kilteni et al., 2015). For example, if we see a body from a first-person perspective that moves as we move (i.e., synchronous visuo-motor correlations) (Gonzalez-Franco et al., 2010), or is touched with the same spatio-temporal pattern as our real body (i.e., synchronous visuo-tactile correlations) (Slater et al., 2010b), or is just static but co-located with our own body (i.e., congruent visuo-proprioceptive correlation) (Maselli and Slater, 2013), then an embodiment experience is generated. In fact, even if the participant is not moving, nor being touched, in some setups, the first-person co-location will be sufficient to generate embodiment (González-Franco et al., 2014; Maselli and Slater, 2014; Gonzalez-Franco and Peck, 2018), to result in the perceptual illusion that this is our body (even though we know for sure that it is not). Interestingly, and highly useful as a control, if there is asynchrony or incongruence between sensory inputs (either in space or in time) or in sensorimotor correlations, the illusion breaks (Berger et al., 2018; Gonzalez-Franco and Berger, 2019). This body ownership effect that was first demonstrated with a rubber hand (Botvinick and Cohen, 1998), has now been replicated in a large number of instances with virtual avatars, (for a review see the Slater and Sanchez-Vives, 2016; Gonzalez-Franco and Lanier, 2017).

FIGURE 1 | Two of the Microsoft Rocketbox avatars being used for a first-person substitution of gender matched bodies in an embodiment experiment at University of Barcelona by Maselli and Slater (2013).

This means that self-avatars not only allow us to interact with others as we would do in the real world, but are also critical for non-social VR experiences. The body is a basic aspect required for perceptual, cognitive, and bodily interactions. Inside VR, the avatar becomes our body, our "self". Indeed participants that have virtual body show better perceptual ability when estimating distances than non-embodied participants in VR (Mohler et al., 2010; Phillips et al., 2010; Gonzalez-Franco et al., 2019a). Self-avatars also change how we perceive touch inside VR (Maselli et al., 2016; Gonzalez-Franco and Berger, 2019). Even more interestingly, self-avatars can even help users to better perform cognitive tasks (Steed et al., 2016b; Banakou et al., 2018), modify

implicit racial bias (Groom et al., 2009; Peck et al., 2013; Banakou et al., 2016; Hasler et al., 2017; Salmanowitz, 2018) or even change, for example, their body weight perception (Piryankova et al., 2014).

Such examples of research are just the tip of the iceberg, but show the importance of avatars that are controllable for developing VR/AR experiments, games, and applications. Hence, multiple commercial and non-commercial tools (such as the Microsoft Rocketbox library, Autodesk Character Generator, Mixamo/Adobe Fuse or iClone Character Creator, to name a few) aim to democratize and extend their use of avatars among developers and researchers.

Avatars can be created in different ways and in this paper we will detail how they can be fitted to a skeleton, and animated. We give an overview of previous work as well as future avenues for avatar creation. We also describe the particularities of the use and creation of the Microsoft Rocketbox avatar library and we discuss the consequences of the open source release (Mic, 2020).

2. AVATAR CREATION

The creation of avatars that can move and express emotions is a complex task and relies on the definition of both the morphology and behavior of its 3D representation. The generation of believable avatars requires a technical pipeline (**Figure 2**), that can create the geometry, textures, control structure (rig) and movements of the avatar (Roth et al., 2017). Anthropomorphic avatars might be sculpted by artists, scanned from real people or a combination of both. Microsoft Rocketbox avatars are based on sculpting. With current technological advances it is also possible to create avatars automatically from a set of input images or through manipulation of a small set of parameters. At the core rigging is the set of control structures attached to selected areas of the avatar, allowing its manipulation and animation. A rig is usually represented by a skeleton (bone structure). However, rigs can be attached to any structure that is useful for the task needed (animators may use much more complex rigs than a skeleton). The main challenge when rigging an avatar is to accurately mimic the deformation of an anthropomorphic shape, so artists and developers can manipulate and animate the avatar. Note that non-anthropomorphic avatars, such as those lacking legs or other body parts, are widely used in social VR. However, although previous research has shown that non-anthropomorphic avatars can also be quite successful in creating self-identification, there are more positive effects to using full anthropomorphic avatars (Aymerich-Franch, 2012).

2.1. Mesh Creation

2.1.1. Sculpting

Much like the ancient artists sculpted humans in stone, today digital artists can create avatars by combining and manipulating digital geometric primitives and deforming the resulting collection of vertex or meshes. Manually sculpting, texturing, and rigging using 3D content creation tools, such as Autodesk 3ds Max (3dm, 2020), Maya (May, 2020), or Blender (Ble, 2020), has been the traditional way to achieve high-quality avatars.

This work requires artists specializing in character design and animation, and though this can be a long and tedious process, the results can be optimized for high-level quality output. Most of the avatars currently used for commercial applications, such as AAA games and most VR/AR applications, are based on a combination of sculpted and scanned work with a strong artistic involvement.

In fact, at the time of writing, specialized artistic work is generally still used to fine tune avatar models to avoid artifacts produced with the other methods listed here, even though some of these avatars can still suffer from the uncanny valley effect (Mori, 1970), where extreme but not perfect realism can cause a negative reaction to the avatar.

2.1.2. Data Driven Methods and Scanning

In many applications, geometry scanning is used to generate an avatar that can be rigged and animated later (Kobbelt and Botsch, 2004). Geometry scanning is the acquisition of physical topologies by imaging techniques and accurate translation of this surface information to digital 3D geometry; that is, a mesh (Thorn et al., 2016).

That later animation might itself use a motion capture system to provide data to animate the avatar. The use of one-time scanning enables the artist to stylize the scanned model and optimize it for animation and the application. Scanning is often a pre-processing step where the geometry and potentially animation is recorded and used for reference or in the direct production of the final mesh and animation.

Scanning is particularly useful if an avatar needs to be modeled on existing real people (Waltemate et al., 2018). Indeed, the creation of avatars this way is extremely common in special effects rendering, for sports games or for content around celebrities.

Depth cameras or laser techniques allow developers and researchers to capture their own avatars, and create look-alike avatars (Gonzalez-Franco et al., 2016). These new avatars can later be rigged in the same way as sculpted ones. Scanning is a fast alternative to modeling of avatars. But depending on the quality, it comes with common flaws that need to be tweaked afterwards.

Indeed, special care needs to be taken to scan the model from all directions to minimize the number and size of occluded regions. To enable future relighting of the avatar, textures have to be scanned under known natural and sometimes varying illuminations to recover the original albedo or bi-directional reflectance function (BRDF) of the object (Debevec et al., 2000).

To capture the full body (**Figure 3**), the scanning requires a large set of images to be taken around the model by a moving camera (Aitpayev and Gaber, 2012). Alternatively a large number of cameras in a capture stage can collect the required images at one instant (Esteban and Schmitt, 2004). Surface reconstruction using multi-view stereo infers depth by finding matching neighborhoods in the multiple images (Steve et al., 2006). When there is a need to capture both the geometry and textures of the avatar, as well as the motion of the model, a large set of cameras can also enable the capture of a large set of images that covers most, yet but not all, of the model surface at each time frame sometimes this is referred to as fusion4d or volumetric capturing (Dou et al., 2016; Orts-Escolano et al., 2016). More

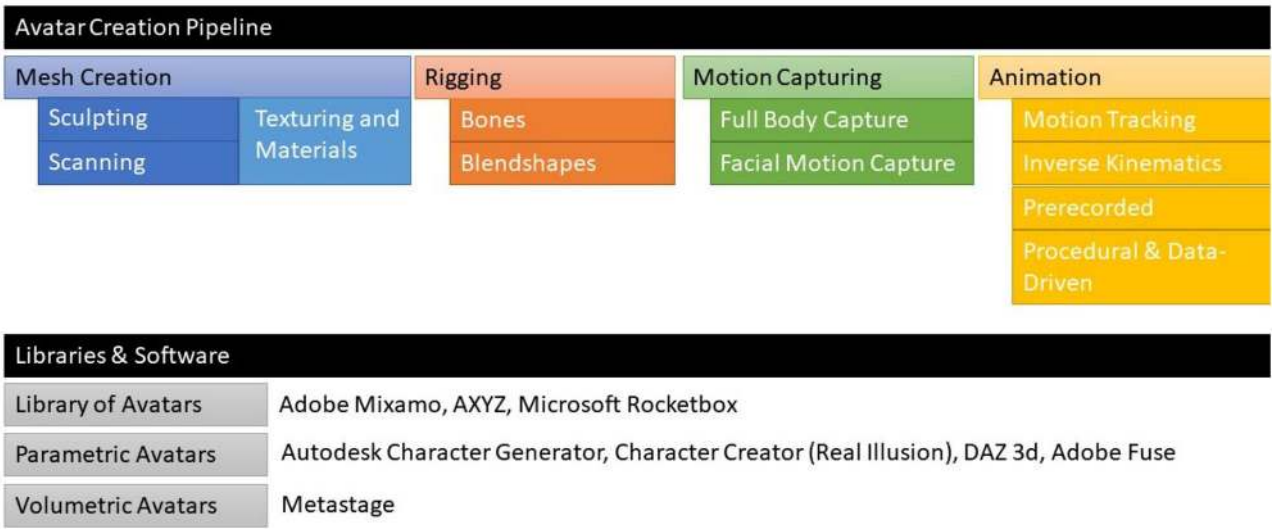

FIGURE 2 | Pipeline for the creation of Microsoft Rocketbox Rigged Avatars and main libraries and software available with pre-created avatars of different types. The mesh creation produces a basic sculpted humanoid form. The rigging attaches the vertices to different bones corresponding to the skeleton of the digital human.

details on volumetric performance capture are presented in section 4.

Special care should be taken when modeling avatar hair. While hair may be an important recognizable feature of a person, it is notoriously hard to scan and model due to its fine structure and reluctance properties. Tools such as those suggested by Wei et al. (2005) use visible hair features as seen by multiple images capturing the head from multiple directions, to fill a volume similar to the image. However, hair and translucent clothing remains a research challenge.

Recent deep learning methods also using retrieved data can be used at many stages of the avatar creation process. Some methods have been successfully used to create avatars from pictures by recreating full 3D meshes from a photo (Hu et al., 2017; Saito et al., 2019), meshes from multiple cameras Collet et al. (2015); Guo et al. (2019), reduce the generated artifacts (Blanz and Vetter, 1999; Ichim et al., 2015), as well as to improve rigging (Weng et al., 2019). Deep learning methods can also generate completely new avatars that are not representations of existing people, by using adversarial networks (Karras et al., 2019).

2.1.3. Parametric Avatars

Another way to reduce the effort needed to model an avatar of an actual person is to use a parametric avatar and fit the parameters to images or scans of the person.

Starting from an existing avatar guarantees that the fitting process will end up with a valid avatar model that can be rendered and animated correctly. Such methods are able to reduce the modeling process to a few images or even a single 2D image (Saito et al., 2016; Shysheya et al., 2019). They can be used to recover physically correct models that can be animated of finer details, even with fine and semi-transparent objects, such as hair (Wei et al., 2019). This is at the cost of some differences compared to the actual person's geometry. There are many things these systems still cannot do well without additional artist ic work: hair, teeth, and fine lines in the face, to name a few.

FIGURE 3 | A Microsoft Rocketbox avatar mesh with less than 10,000 triangle polygons with two texture modes: only diffuse and diffuse, specular and normal (bump mapping). The models include multiple Levels of Detail (LODs) which can be optionally used for performance optimization (on the right).

There are some commercial and non-commercial applications for parametric avatars that provide some of these features, including Autodesk Character Generator, Mixamo/Adobe Fuse or iClone Character Creator.

2.2. Texturing and Materials

At the time of rendering, each vertex and face of the mesh is assigned a value. That value is computed from retrieving

the textures in combination with the shaders that might incorporate enhancements and GPU mixing of multiple layers with information about specular, albedo, normal mapping, and transparency among other material properties. While the specular map states where the reflection should and should not appear, the albedo is very similar to a diffuse map, but with one extra benefit: all the shadows and highlights have been removed. The normal map is used to add details without using more polygons; the computer graphics simulates high-detail bumps and dents using a bitmap image.

Therefore, texturing the avatar is another challenge in the creation process (**Figure 4**). Adding textures requires specific skills as the vertexes and unwrapping are often based on real images that might also need additional processing, but in some occasions this can be automated.

When captured independently, each image covers only part of the object, and there is a need to fuse all sources to a single coherent texture. One possibility is to blend available images into a merged texture (Wang et al., 2001; Baumberg, 2002). However, any misalignment or inaccurate geometry (such as the case of fitting a parametric model) might lead to ghosting and blurring artifacts when the textures are geometrically misaligned.

To reduce such artifacts, texturing has been addressed as an image stitching problem (Lempitsky and Ivanov, 2007; Gal et al., 2010). This approach targets each surface triangle that is then projected onto the images from which it is visible, and the final texture is assigned entirely from one image in this set. The goal is to select the best texture source and to penalize mismatches across triangle boundaries. Lighting variations are corrected at post-processing using a piece-wise continuous function over the triangles.

A shader program runs in the graphics pipeline and determines how the computer will render each pixel of the screen based on the different texture images and the material and physical properties associated with the different objects. Shader programs are often used to control lighting and shading effects, and are programmed using GLSL (OpenGL Shading Language).

Shader programming also plays a very important role in applying all the additional textures (albedo, specular, bump mapping etc.). Most of the rendering engines will provide some basic shaders that will map the textures of the material assigned to the avatar.

2.3. Rigging Avatars

A key step in most animation systems is to attach a hierarchy of bones to the mesh in a process called rigging. The intention is that animations are defined by moving the bones, not the individual vertices. Therefore during rigging, each vertex of the mesh is attached to one or more bones. If a vertex is attached to one bone, it is either rigidly attached to that bone or has a fall-out region of influence. Alternatively if a vertex is attached to multiple bones then the effect of each bone on the vertex's positions is defined by weights. This allows for a mesh to smoothly interpolate as the bones move. These vertexes affected by multiple bones are typical in the joint sections.

Animators may also use much less complex rigs than a skeleton by combining, for example, different expressions (i.e., blendshapes) and movements to a combined trigger (e.g., happy expression).

2.3.1. Bones

Artists typically structure the bones to follow, crudely, the skeletal structure of the human or animal (**Figure 5**).

There are several pipelines available today for rigging a mesh so that it can be imported into a game engine such as Unreal or Unity. One common way is using the CATS plugin for Blender (Cat, 2020) that can export in a compatible manner with the Mecanim retargeting system used in Unity. Another method is to import the mesh into Maya or another 3D design program and rig it by hand, or alternatively, use Adobe Fuse (Fus, 2020) to rig the mesh. Nonetheless, each of these methods has its advantages and disadvantages. As with any professional creation tool, they have steep learning curves in becoming familiar with the user interfaces and with all of the foibles of misplacing bones in the mesh that can lead to unnatural poses. Although with the CATS plugin and Fuse the placement of bones is automated, if they are not working initially it can be difficult to fix issues.

Mixamo (Mix, 2020) makes rigging and animation easier for artists and developers. Users upload the mesh of a character and then place joint locators (wrists, elbows, knees, and groin) by following the onscreen instructions. Mixamo is not only an auto-rigging software but also provides a character library.

Independently of the tool used, the main problems usually encountered during the animation of avatars are caused by various interdependent design steps in the 3D character creation process. For example, poor deformation can come from bad placement of the bones/joints, badly structured mesh polygons, incorrect weighting of mesh vertex to their bones or non-fitting animations. Even for an avatar model that looks extremely realistic in a static pose, bad deformation/rigging would lead to a significant loss of plausibility and immersion in the virtual scenario as soon as the avatar moves its body. In fact, if not done carefully rigged avatars can exhibit pinched joints or unnatural body poses when animated. This depends on the level of sophistication of the overall process, weight definition, deformation specs, and kinematic solver. The solver software can also create such effects.

Overall, creating a high quality rig is not trivial. There is a trend toward easier tools and automatic systems for rigging meshes and avatars (Baran and Popović, 2007; Feng et al., 2015). They will be easier and more accessible as the technology evolves but having access to professional libraries of avatars, such as Microsoft Rocketbox, can simplify production and allow researchers to focus on animating and controlling avatars at a high level rather than at the level of mesh movement.

2.3.2. Blendshapes

While bones are intuitive control systems for bodies, for the face, bones are not created analogously to real bones, but rather to pull and push on different mesh parts of the face. While bones can be used for face animation on their own, a common alternative to bone-based facial animation is to use blendshape animation. (Vinayagamoorthy et al., 2006; Lewis et al., 2014).

FIGURE 4 | A close-up of an avatar face from Microsoft Rocketbox and corresponding set of textures that are mapped to the avatar mesh to produce the best possible appearance. Textures need to be mapped per each vertex and can contain information about the diffuse, specular, normal, and transparency colors of each vertex.

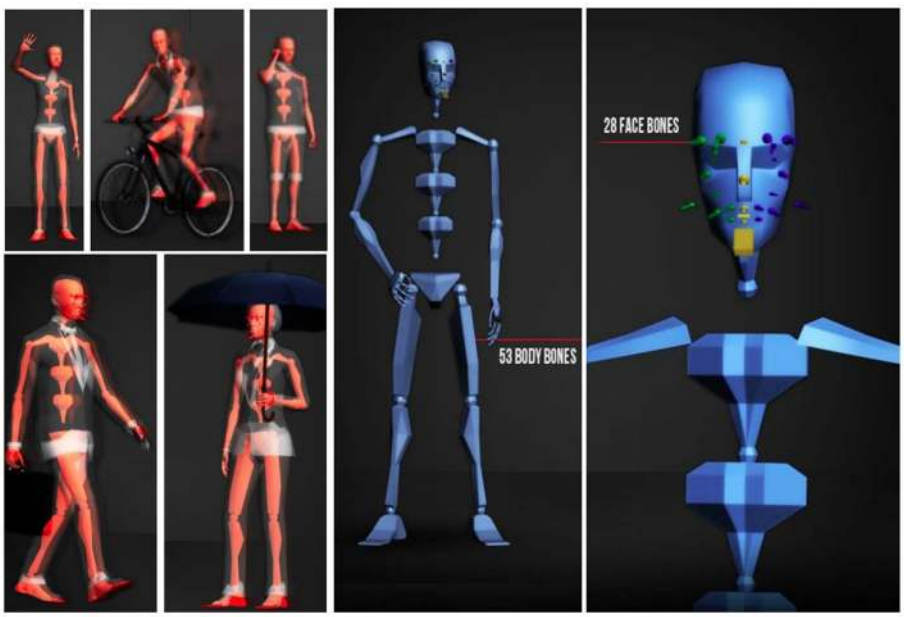

FIGURE 5 | Bones in one of the Microsoft Rocketbox avatars (53 body bones and 28 facial bones).

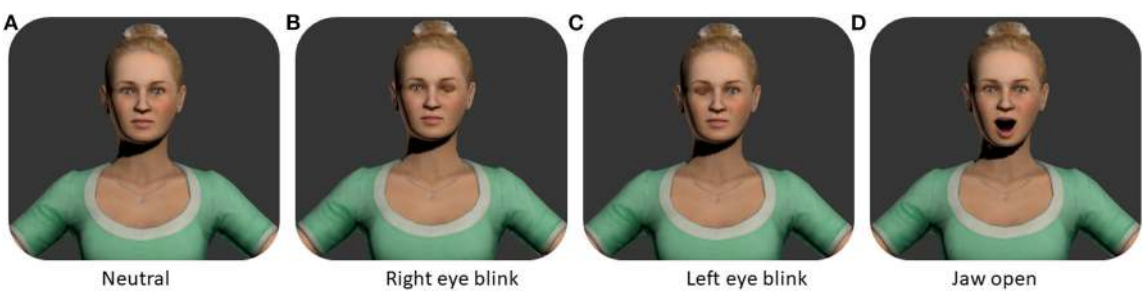

FIGURE 6 | One of the Microsoft Rocketbox avatars—Blendshapes by Pan et al. (2014).

Blendshapes are variants of an original mesh with each variant representing a different non-rigid deformation or, in this context, a different isolated facial expression. The meshes have the same number of vertices and the same topology. Facial animation is created as a linear combination of blendshapes. At key instances in time, blendshapes are combined as a weighted sum into a keypose mesh. Different types of algebraic methods are used to interpolate between keyposes for all frames of the animation (Lewis et al., 2014). For example, one can select 10% of a smile, and 100% of left eye blink and the system would combine both as in **Figure 6**. Blendshapes can then be considered as the units of facial expressions and despite the fact that they are seldomly followed there are standards proposed such as the Facial Action Coding System (FACS) proposed by Ekman and Friesen (1976). The number of blendshapes in facial rigs varies. For example, some facial rigs use 19 blendshapes plus neutral (Cao et al., 2013). However, if facial expressions have more complex semantics the sliders can reach much higher numbers, and 45 or more poses are often used (Orvalho et al., 2012). For reference, high-end cinematic quality facial rigs often require hundreds of blendshapes to fully represent subtle deformations (the Gollum character in the Lord of the Rigns movies had over 600 blendshapes for the face).

Blendshapes are mostly used for faces because (i) they capture well non-rigid deformations common for the face as well as small details that might not be represented by a skeleton model, and (ii) the physical range of motion in the face is limited, and in many instances can be triggered by emotions or particular facial expressions that can be later merged with weights (Joshi et al., 2006).

Blendshapes can be designed manually by artists but also captured using camera tracking systems (RGB-D) from an actor and then blended through parameters at the time of rendering (Casas et al., 2016). In production settings, hybrid blendshape and bones based on facial rigs are sometimes used.

3. MOTION CAPTURING AND ANIMATION

At run-time, rigged avatars can be driven using pre-recorded animations or procedural programs to create movements during the VR/AR experience (Roth et al., 2019). A self-avatar can be animated to follow real-time motion tracking of the participants. The fidelity of this type of animation depends on the number of joints being captured and the techniques used to do extrapolations and heuristics between those joints (Spanlang et al., 2014).

VR poses a unique challenge for animating avatars that are user driven as this needs to happen at low latency so there is little opportunity for sophisticated animation control. It is particularly tricky if the avatar is used as a replacement for the user's unseen body. In such cases, to maximize the feeling of body ownership of the user, VR should use a low latency motion capture of the user's body and animate the avatar in the body's place. Due to limitations of most commercial VR systems, the capture of the body is typically restricted to measuring the 6 degrees of freedom of two controllers and the HMD. Other techniques, such as inverse kinematics, can be used to infer movements of the body that are not actually tracked. From these, the locations of two hands and head can be determined and other parts of the body can be inferred.

FIGURE 7 | Body motion capture via OptiTrack motion capture system (Opt, 2020) and the corresponding mapping to one of the Microsoft Rocketbox avatars in a singing experiment by Steed et al. (2016a).

In recent years some systems have offered additional sensors such as finger tracking around the hand-held controllers, eye tracking in the HMD, and optional stand-alone trackers. Research is examining the use of additional sensors and tracking systems such as external (Spanlang et al., 2014; Cao et al., 2017; Mehta et al., 2017) or wearable cameras (Ahuja et al., 2019) for sampling more information about the user's pose and perhaps in the future we may see full-body sensing.

3.1. Motion Capture

3.1.1. Full Body Motion Capture

Motion capture is the processes of sensing of a person's pose and movement. Pose is usually represented by a defining skeletal structure, and then for each joint giving its 3 or 6 degrees of freedom transformation (rotations and sometimes position) from its parent in the skeletal structure. A common technique for motion capture has the actor wearing easily recognized markers such as retro-reflective markers or active markers that are observed by multiple high-speed cameras (Ma et al., 2006). Optitrack (Opt, 2020) and Vicon (Vic, 2020) are two commercial motion capture systems that are commonly used in animation production and human- computer interaction research (**Figure 7**).

Another set of techniques uses computer vision to track a person's movement without markers, such as using the Microsoft Kinect sensor (Wei et al., 2012), and more recently RGB cameras (Shiratori et al., 2011; Ahuja et al., 2019), RF reflection (Zhao et al., 2018a,b), and capacitive sensing (Zheng et al., 2018). A final set of techniques uses Worn sensors to recover the user's motion with no external sensors. Examples include inertial measurement units (IMUs) (Ha et al., 2011), wearable cameras (Shiratori et al., 2011) or mechanical suits such as METAmotion's Gypsy (Met, 2020).

3.1.2. Facial Motion Capture

For facial motion capture we have similar options to the ones proposed in the full body setup: either marker-less with a camera and computer vision tracking algorithms, such as Faceshift (Bouaziz et al., 2013) (**Figure 8**), or using marker-based systems such as Optitrack.

However, the problem with facial animation is that on instances, when needed in real-time, users might be wearing HMDs that occlude their real face, and therefore capturing their expressions is very hard (Lou et al., 2019). Some researchers have shown that it is possible to add sensors to the HMDs to record facial expressions (Li et al., 2015). The BinaryVR face tracking device attaches to your HMD and tracks facial expressions (Bin, 2020).

3.2. Animation

For animating an avatar, translations and rotations of each of the bones are changed to match the desired pose at an instant in time. If the motions have been prerecorded, then this constitutes a sort of playback. However, data-driven approaches can be much more sophisticated than simple playback.

Independently of whether the animation is performed in real-time or using pre-recorded animations, in both cases at some point there has been a motion capture that either sensed the full body and created prerecorded animations, or a subset of body parts to later do motion retargeting. If is done offline the animations are then saved and accessed during runtime. Aternatively, if the motion capture is done in real-time then this can be used directly to animate the avatars that match the users' motions.

3.2.1. Motion Tracking

When each joint of the participant can be recorded in real-time through a full body motion capturing system (Spanlang et al., 2014), they can be transferred directly to the avatar (Spanlang et al., 2013). For best performance this assumes that the avatar is resized to the size of the participant. In that situation participants achieve the maximum level of agency over the self-avatar and rapidly embody it (Slater et al., 2010a). However, this requires wearing a motion capturing suit or having outside sensors monitoring the motions of the users, which can be expensive and/or hard to setup (Spanlang et al., 2014). Since direct mapping between the tracking and avatar skeletons can sometimes be non trivial, intermediate solvers are typically put in place in most approaches today Spanlang et al. (2014).

Most commercial VR systems capture only the rotation and position of the HMD and two hand controllers. In some cases the hands are well-tracked with finger capacitive sensing around the controller. Some new systems are introducing more sensors such as eye tracking. However, in most systems, most of the body degrees of freedom are not monitored. Even if only using partial tracking (Badler et al., 1993) it is possible to infer a good approximation of the full body, but the end-effectors need to be well-tracked. Nevertheless, many VR applications limit the rendering of the user's body to hands only. Although such representations may suffice for many tasks, this lowers the realism of the virtual scenario and may reduce the level of embodiment of the user (De Vignemont, 2011) and even have further impact even on their cognitive load (Steed et al., 2016b).

3.2.2. Inverse Kinematics

A common technique to generate a complete avatar motion given the limited sensing of the body motion, is Inverse Kinematics (IK). The position and orientations of avatar joints that are not monitored are estimated given the known degrees of freedom. IK originated from the field of robotics and exploits the inherent movement limitations of each of the skeleton's joints to retrieve a plausible pose (Roth et al., 2016; Aristidou et al., 2018; Parger et al., 2018). The pose fits the position and direction of the sensed data and lies within the body's possible poses. IK addresses the challenge that there are many possible combinations and sometimes it is not a perfect one-to-one method. Additionally it can do a reasonable job in reconstructing motions in joints that are within ground truth nodes, but does a worse job in extrapolating to other joints, for example to know the leg motions from the hand trackers is harder than it is to know the elbow motions.

3.2.3. Prerecorded Animations

Most applications of avatars in games and movie productions use a mix of prerecorded motion captures of actors and professionals (Moeslund et al., 2006) and artists' manual animations, together with AI, algorithm ic and procedural animations. The main focus of these approaches is the generation of performances that are believable, expressive, and effective. Added animation aims at minimizing artifacts that can result due to a lack of small motions that may exist during a real person's motions, such as clothes bending, underlying facial movements (Li et al., 2015) and others.

Prerecorded animations alone can be a bit limiting in the interactivity delivered by the VR, especially if they are driving a first-person embodied avatar where the user wants to retain agency over the body. For that reason, prerecorded animations are generally not used for self-avatars; however, recent research has shown that they can be used to trigger enfacement illusions and realistic underlying facial movements (Gonzalez-Franco et al., 2020b).

Indeed, the recombination of existing animations can become quite complex and procedural or deep learning systems can be in charge of generating the correct animations for each scene.

3.2.4. Interactive Procedural, Deep Learning and Data-Driven Animation

A combination of procedural animation with data-driven animation is a common strategy for real-time systems. However, this approach also allows compensation for incomplete sensing data for real-time animation, especially self-avatars.

In some scenarios the procedural animation systems can use the context of the application itself to animate the avatars. For example, in a driving application where the legs can be rendered sitting in the driver seat performing actions consistent with the driving inputs even if the user is not actually in a correct driving position.

In other scenarios the approach is more interactive and depends on the user input. It can incorporate tracking

FIGURE 8 | One of the Microsoft Rocketbox avatars animated with markerless facial motion capture via Faceshift by Steed et al. (2016a).

information from the user motion capture and start to complete parts of the self-avatar or their motions if they are not complete. This could be an improvement over traditional IK, in which a data-driven animation can incorporate the popularity of the sampled data and retrieve as a result the most likely pose within the whole range of possible motions of the IK (Lee et al., 2010).

Using data-driven machine learning algorithms the system can pick the most likely pose out of the known data given the limited capture information. In many cases this might result in a motion retargeting or blending scenario.

More recently, we have seen rigged avatars that can be trained through deep learning to perform certain motion actions, such as dribbling (Liu and Hodgins, 2018), adaptive walking and running (Holden et al., 2017), or physics behaviors (Peng et al., 2018). These are great tools for procedural driven avatars and can also solve some issues with direct manipulation.

However, with any such procedural animation, care needs to be taken that the animation doesn't stray too far from other plausible interpretations of the tracker input as discrepancies in representation of the user's body may reduce their sense of body ownership and agency (Padrao et al., 2016; Gonzalez-Franco et al., 2020a).

4. VOLUMETRIC AVATARS

There are other forms of avatar creation beyond the pipeline defined previously. For example using volumetric capture, we can capture meshes and their animations without the need for rigging (**Figure 9**). Volumetric capturing can be considered an advanced form of scanning in which the external optical sensors (often numerous) allow a reconstruction of the whole point cloud and mesh of a real person and can record the movement of the person over time, essentially leading to a 3D video.

When we talk about volumetric avatars, we refer to different types of 3D representations, such as point clouds, meshes (Orts-Escolano et al., 2016), voxel grids (Loop et al., 2016), or light fields (Wang et al., 2017; Overbeck et al., 2018). However, the 3D triangle mesh is still one of the most common 3D representations that is used in current volumetric capture systems, mostly because GPUs are traditionally designed to render triangulated surface meshes. Moreover, from a temporal point of view, we can also distinguish between capturing a single frame, and then performing rigging on the single frame, or capturing a sequence of these 3D representations, which is usually known as volumetric video. Finally, volumetric avatars can also be categorized as those that are captured offline (Collet et al., 2015; Guo et al., 2019) and played back as volumetric streams of data or those that are being captured and streamed in real-time (Orts-Escolano et al., 2016).

Either real-time or offline, volumetric video capture is a complex process that involves many stages, from low-level hardware setups and camera calibration, to sophisticated machine learning and computer vision processing pipelines. A traditional volumetric capture pipeline includes the following steps: image acquisition, image preprocessing, depth estimation, 3D reconstruction, texture UV atlas parameterization, and data compression. Some additional steps are usually performed to remove flickering effects in the final volumetric video, as mesh tracking (Guo et al., 2019) or non-rigid fusion (Dou et al., 2016, 2017) in the case of offline approaches and real-time ones, respectively. More detailed information about state-of-the-art volumetric performance capture systems can be found in Guo et al. (2019), Collet et al. (2015), and Orts-Escolano et al. (2016).

The latest frontier in volumetric avatars has been for the computation to be performed in real-time, and this was achieved in the Holoportation system (Orts-Escolano et al., 2016). When these real-time systems are combined with mixed reality displays

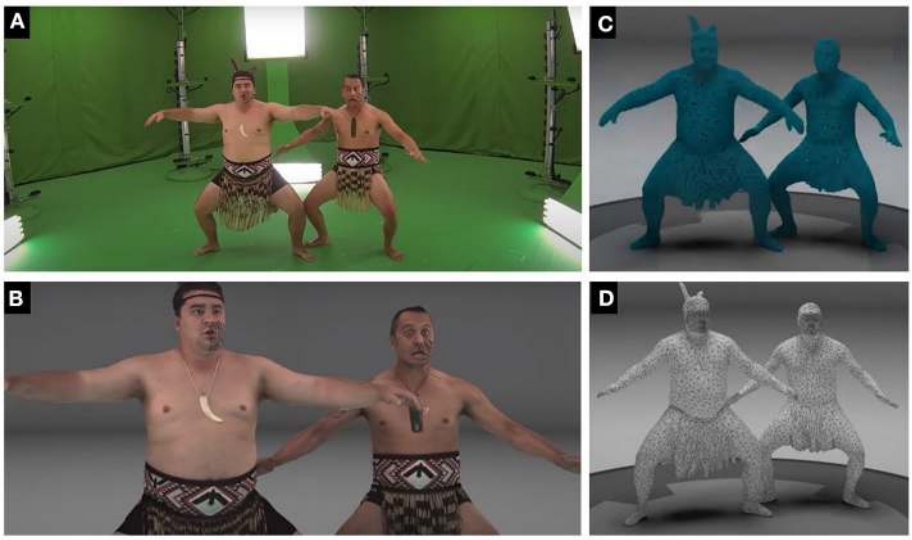

FIGURE 9 | Two volumetric avatars captured in the Microsoft Mixed Reality studio Collet et al. (2015). **(A)** Input setup. **(B)** Avatar output. **(C)** Scanned point cloud. **(D)** Processed mesh.

such as HoloLens, this technology also allows users to see, hear, and interact with remote participants in 3D as if they are actually present in the same physical space.

Both pre-recorded and real-time volumetric avatars offer a different level of realism that cannot be currently achieved by using rigging techniques. For example, volumetric avatars automatically capture object interaction, extremely realistic cloth motion capture and accurate facial expressions.

Real-time volumetric avatars can enable a more interactive communication with remote users, allowing also the user to interact and perform eye contact with the other participants. However, these systems have to deal with a huge amount of computing, sacrificing the quality of the reconstructed avatar for a more interactive experience, which can also affect the perception of eye gaze or other facial clues (MacQuarrie and Steed, 2019). A good rigged avatar with animated facial animation can do an equal or better job for communication (Garau et al., 2003; Gonzalez-Franco et al., 2020b). Moreover, it is important to note that if the user is wearing a HMD while being captured, either a VR or an AR device, it will require additional processing to remove the headset in real-time (Frueh et al., 2017). For this purpose, eye-tracking cameras are often mounted within the headset, which allows in-painting of occluded areas of the face in a realistic way in the final rendered 3D model. Another common problem that real-time capture systems need to solve is latency. Volumetric captures tend to involve heavy computational processing and result in larger data sets. Thus, they require high bandwidth and this increases the latency between the participants in the experience.

Most recent techniques or generating real-time volumetric avatars are focusing on learning-based approaches. By leveraging neural rendering techniques these novel methods have achieved unprecedented results, enabling the modeling of view-dependent effects such as specularity and also correcting for imperfect geometry (Lombardi et al., 2018, 2019; Pandey et al., 2019). Compared to traditional graphics pipelines, these new methods require less computation and also can deal with a coarser geometry, which is usually used as a proxy for rendering the final 3D model. Most of these approaches are able to achieve compelling results, with substantially less infrastructure than previously required.

5. MICROSOFT ROCKETBOX AVATARS

The Microsoft Rocketbox avatar library creation process deployed a lot of research and prototyping work into "base models" for male and female character meshes (**Figure 10**). These base models were already rigged and tested with various (preferably extreme) animations in an early stage of development. The base models were used as a starting point for creation of all character types later on so as to guarantee a high-quality standard and consistent specifications throughout the whole library. Optimization opportunities identified during the production phase of the library also flowed back into these base models. In order to be able to mix different heads with different bodies of the avatars more easily, one specific polygon edge around the neck area of the characters is identical in all avatars of the same gender.

UV Mapping was also predefined in the base meshes already. UV Mapping is the process of "unfolding" or "unwrapping" the mesh to a 2D coordinate system that maps onto the texture. To allow for mixing and matching texture elements of different characters from the library, many parts of the character UV coordinates were standardized; for example, the hands or the face. Therefore, it is possible to exchange face textures across different characters of the library with some knowledge of image editing and retouching. The UV Mapping of polygons that move

FIGURE 10 | Several of the 115 Microsoft Rocketbox rigged avatars released in the library that show a diversity in race, gender, and age, as well as attire and occupation.

or stretch significantly when animation is applied achieves a higher texture resolution to avoid blur effects (**Figure 4**).

Another important set of source data for the creation of the Microsoft Rocketbox library were photographs of real people that were taken in a special setup with photo studio softbox lights. A whitebox and a turntable to guarantee neutral lighting of the photo material that was used as a reference for modeling the meshes and as source material for creating the textures. Photos of people from many different ethnic groups, clothing styles, age classes, and so on, were taken in order to have high diversity across the library. However, different portions from the source material were mixed and strongly modified for the creation of the final avatars so that the avatars represent generic humans that do not exist in reality.

The models are available in multiple Levels of Detail (LODs) which can be optionally be used for performance optimization for scenes with larger numbers of characters, or for mobile platforms. The LODs in the models include "hipoly" (10.000 triangles), "midpoly" (5.000 triangles), "lowpoly" (2.500 triangles), and "ultralowpoly" (500 triangles) levels, (**Figure 3**). Textures are usually included with a resolution of 2048x2048 pixels, with one separate texture for the head and another one for the body. This means that the head texture has a higher detail level (pixel resolution per inch) than the body texture. A set of textures are then mapped to the avatar mesh to produce the best possible appearance. Textures need to be mapped per each vertex and can contain information about the diffuse, specular, or normal colors of each vertex (**Figure 4**).

All in all, the Rocketbox avatars provide a reliable rigged skeletal system with 56 body bones, that despite not being procedurally generated can be used for procedural animation thanks to the rigged system.

5.1. Facial Animation

Facial animations for the Microsoft Rocketbox library animation sets (**Figure 11**) were created though generating different facial expressions manually by setting up and saving different face bone constellations (**Figure 5**). Additionally visemes (face bone positions) for all important phonemes (speech-sounds) were created. These facial expressions were saved as poses and used for keyframe animation of the face adapted to the body movement of the respective body animation. The usage of animation sets across all characters of the library without retargeting the animations required some general conditions to be accounted for during creation of the Rocketbox HD library. The bones in the face only transform in x,y,z and do not rotate. Therefore, facial animation also looks correct on character variants with face elements such as eyebrows or lips positioned deviant from the base mesh. The only exception are the bones for the eyeballs which have rotation features only—this requires the eyeballs of all avatars that belong to the same class (e.g., adult female) to be at the same position in 3D space.

5.2. Limitations

When compared to the other available avatar libraries and tools, the Microsoft Rocketbox present some limitations (**Table 1**).

The Rocketbox avatars do not use blendshapes for facial animation but bones, meaning that, it is up to the users to create their own blendshapes on top, as shown in **Figure 11**. However, creating blendshapes is a process that needs skill and cannot be replicated easily to other avatars but has to be performed manually for each avatar (or scripted).

The library also shows limitations for extreme scaling, that can result in errors regarding the predefined skeleton skin weights and additionally result in unrealistic proportions (see **Figure 12** where the texture stretches unrealistically). Other limitations are the exchangeability of clothes, hair styles, face looks, gender, and so on. Nevertheless, given that the avatars use a base model this simplifies the access to and editing of avatar shapes and clothes. For cases where the avatars are required to look like the participants, or in which there is a need to adapt the body sizes of participants, there might be a way to either change the texture

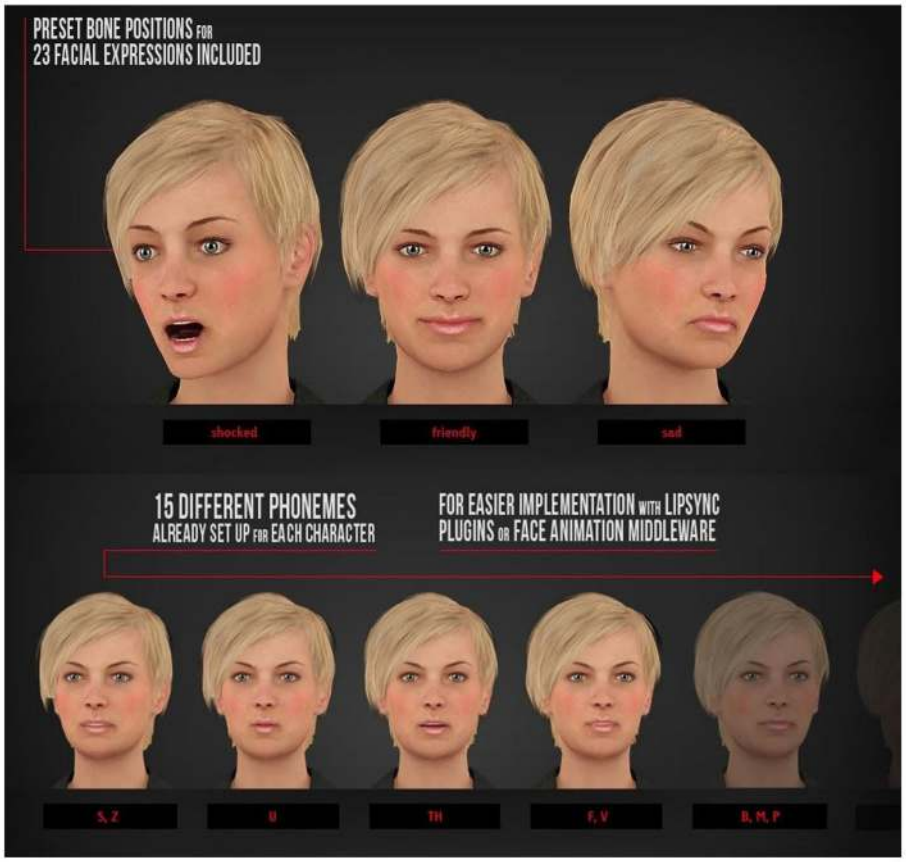

FIGURE 11 | (Top) 3 of the 23 facial expressions included in the Microsoft Rocketbox avatars together with **(Bottom)** 5 of the 15th phonemes already setup in each character for easier implementation with lipsync plugins and mace animation software. Underlying there are 28 facial bones that allow researchers to create their own expressions.

of the facial map, through a transformation, or modifying the mesh, or directly stitching a different mesh to the face of a given library avatar. Initiatives like the virtual caliper (Pujades et al., 2019) could help in adapting the current avatars; however, these possibilities are not straightforward with the release. Future work on the library would also need to bring easier tools for animation for general use, as well as tools to facilitate the exchangeability of clothes based on the generic model, as well as easier blendshape creation.

6. AVATAR SCIENCE RESEARCH AND OTHER APPLICATIONS

The first experiences with avatars and social VR were implemented in the 1980s in the context of the technology startup (VP Research). While many of the results were reported in the popular press (Lanier, 2001) that early work was also the germ of what later became a deep engagement from the academic community into exploring avatars and body representation illusions as well as their impact on human behavior (Lanier et al., 1988; Lanier, 1990). In that regard VPL Research not only did the initial explorations on avatar interactions but also provided crucial VR instrumentation for many laboratories and pioneered a whole industry in the headsets, haptics and computer graphics arenas. Many of the possibilities of avatars and VR for social and somatic interactions were explored (Blanchard et al., 1990) and were later formalized empirically by the scientific community as presented in this section.

Researchers in diverse fields have since explored the impact of avatars on human responses. The underlying aim of the research in many of the avatar scientific studies is to better understand the behavioral, neural, and multisensory mechanisms involved in shaping the facets of humanness, and to further explore potential applications for therapeutic treatments, training, education, and entertainment. A significant number of these studies has used the Microsoft Rocketbox avatars (with over 145 papers and 5307 citations, see the **Supplementary Materials** for the full list). In this paper, at least 49 citations (representing 25% of the citations) used the Rocketbox avatars, and all are referenced in the following section about applications, where they represent almost 50% of all the citations in this section. Considering that the Rocketbox avatars started being used around 2009 (see our list of papers using the avatars in **Supplementary Materials**) and

TABLE 1 | Avatar Libraries and tools comparisons.

	Daz-3D	Mixamo	Autodesk character generator	Microsoft Rocketbox
Avatar	Fully textured and rigged (Improved shoulder, collar, and abdomen bends).	Fully textured and rigged. Automatic rigging of human skeleton to character created in AFCC.	Fully textured and rigged.	Fully textured and rigged.
Range of characters	A very wide ecosystem of models, cloths, accessories. (Every day, Sci-Fi, Animals, etc.).	A wide range of themes, but a small dataset of characters and motions.	Everyday models (Specific focus on this theme), a few motions.	Realistic humans of diverse raze, age, gender, and occupation.
Face	Blend shapes (no bones).	Blend shapes (no bones).	Facial expressions blends including phonemes.	Bones.
Uniqueness and Limitations	Muscle flexion. Blend between different models. Share clothing between genders via morph projection. Body proportions can be modified including child stature. No body mesh under cloths area to prevent artifacts. IRAY shader–sub-surface features. Hi-res UV map.	Avatar is made of different parts that can be mixed and matched. No body mesh under cloths area to prevent artifacts.	Web based service. Limitation: One mesh only–hard to modify or change shaders. Clothing–is part of the mesh.	All avatars created with similar structure so relatively easy to interchange body parts and or outfits. Multiple HOD poly levels. Three submeshes: body, head, and hair. Limitation: Clothing–is part of the mesh.
Paid/Free	Mixed.	Data is free.	Some free data but low poly and texture res, and body rig only, no face animation. Paid–blend shapes no bones, facial expressions.	Free for research and academic use.

that 23 of the cited papers in this section were published before 2009, we can consider that 60% of the research referenced in this section has been carried out with the Rocketbox avatars. The remaining citations are included to provide further context for the importance of the research in this area, as well as of the potential applications of the avatars that range from simulations to entertainment to embodiment.

Indeed, a substantial focus of interest has been on how using and embodying self-avatars can change human perception, cognition and behavior. In fact, the possibility of making a person, even if only temporarily, "live in the skin" of a virtual avatar has tremendously widened the limit of experimental research with people, allowing scientific questions to be addressed that would not have been possible otherwise.

6.1. Bodily Illusions Over Virtual Avatars

In the past two decades there has been an explosion of interest in the field of cognitive neuroscience on how the brain represents the body, with a large part of the related research making use of body illusions. A one-page paper in Nature (Botvinick and Cohen, 1998), showed that a rubber hand can be incorporated into the body representation of a person simply by application of multisensory stimulation. This rubber-hand illusion occurs when a rubber hand in an anatomically plausible position is seen to be tactilely stimulated synchronously with the corresponding real out-of-sight hand. For most people, after about 1 minute of this stimulation and visuo-tactile correlation, proprioception shifts to the rubber hand, and if the rubber hand is threatened there is a strong physiological response (Armel and Ramachandran, 2003) and corresponding brain activation (Ehrsson et al., 2007). This research provided convincing evidence of body ownership illusions. The setup has since been reproduced with a virtual arm in virtual reality (Slater et al., 2009), also with appropriate brain activation to a threat (González-Franco et al., 2014).

Petkova and Ehrsson (2008) showed how the techniques used for the rubber-hand illusion could be applied to the whole body thus producing a "body ownership illusion" of a mannequin's body and this also has been reproduced in VR (Slater et al., 2010b; Yuan and Steed, 2010). First-person perspective over the virtual body that visually substitutes the real body in VR is a prerequisite for this type of body ownership illusion (Petkova et al., 2011; Maselli and Slater, 2014). The illusion induced by visuomotor synchrony, where the body moves synchronously with the movements of the person, is typically more vivid than those triggered by visuotactile synchrony, where the virtual body is passively touched synchronously with real touch (Gonzalez-Franco et al., 2010; Sanchez-Vives et al., 2010; Kokkinara and Slater, 2014). Illusory body ownership over a virtual avatar, however, can be also experienced in static conditions, when the virtual body is seen in spatial alignment with the real body, a condition that could be easily met using immersive stereoscopic VR displays (Maselli and Slater, 2013; Gonzalez-Franco et al., 2020a). The visuomotor route to body ownership is a powerful one, integrating visual and proprioceptive sensory inputs in correlation with motor outputs. This route is uniquely suited to self-avatars in VR, since it requires the technical capabilities of the forms of body tracking and rendering discussed in the animation section. Additional correlated modalities (passive or active touch, sound) can enhance the phenomenon.

Several studies have focused on casting the illusions of body ownership into a coherent theoretical framework. Most accounts have been discussed in the context of the rubber-hand illusion

and point to multisensory integration as the main key underlying process (Graziano and Botvinick, 2002; Makin et al., 2008; Ehrsson, 2012). More recently, computational models developed in the framework of causal inference (Körding et al., 2007; Shams and Beierholm, 2010) described the onset of illusory ownership as the result of an "inference process" in which the brain associates all the available sensory information about the body (visual cues from the virtual, together with somatosensory or proprioceptive cues from the physical body) to a single origin: the own body (Kilteni et al., 2015; Samad et al., 2015). According to these accounts, all visual inputs associated with the virtual body (e.g., its aspect, the control over it, the interactions that it has with the surrounding environment) are processed as if emanating from the own body. As such, these could have a profound implications on perception and behavior.

What is even more interesting than the illusion of ownership over a virtual body is indeed the consequences it can have for changed physiology, behaviors, attitudes, and cognition.

6.1.1. Self-Avatar Impact on Behavior

Self-perception theory argues that people often infer their own attitudes and beliefs from observing themselves—for example their facial expressions or styles of dress—as if from a third party (Bem, 1972). Could avatars have a similar effect? Early work focused on what Yee and Bailenson called "The Proteus Effect," the notion that one's avatar would influence the behavior of the person embodied in it. The first study examined the consequences of embodying older avatars, where college students were embodied in either age-appropriate avatars or elderly ones (Yee and Bailenson, 2007). The results showed that negative stereotyping toward the elderly was reduced when participants were placed into older avatars compared with those placed into young avatars. Subsequent work extended this finding to other domains. For example, people embodied in taller avatars negotiated more aggressively than those embodied in shorter ones, and attractive avatars caused more self-disclosure and closer interpersonal distance to a confederate than unattractive ones during social interaction (Yee and Bailenson, 2007). The embodiment effects of both height and attractiveness in VR extended to subsequent interactions outside of the laboratory, predicting players' performance in online games, but also face-to-face interactions (Yee et al., 2009). Another study from the same laboratory has also shown how embodying avatars of different races can modulate implicit racial prejudice (Groom et al., 2009). In this case, Caucasian participants embodying "Black" avatars demonstrated increased level of racial prejudice favoring Whites, with respect to participants embodying "White" avatars, in the context of a job interview. These effects of embodying an avatar only occurred when the avatar's movements were controlled by the user—simply observing an avatar was not sufficient to cause changes in social behavior (Yee and Bailenson, 2009).

While these pioneering studies give a solid foundation to understanding the impact of self-avatars on social behavior, they relied on fairly simple technology at the time. The avatars were only tracked with head rotations and the body translated as a unit, while the latency and frame-rate were courser than today's standards. Also, these studies entailed a third-person view of self-avatars, either reflected in a mirror from a "disembodied" perspective (i.e., participants could see the avatar body rendered in a mirror as from a first-person perspective, but looking down they would see no virtual body), or from a third-person perspective as in an on-line community. More recent work has demonstrated that embodying life-size virtual avatars from a first-person perspective, and undergoing the illusion of having one's own physical body (concealed from view) substituted by the virtual body seen in its place and moving accordingly, could further leverage the impact of self-avatars on implicit attitudes, social behavior, and cognition (Maister et al., 2015).

Several studies have shown that in neutral or positive social circumstances embodiment of White people in a Black virtual body results in a reduction in implicit racial bias, measured using the Implicit Association Test (Peck et al., 2013). Similar results had been found with the rubber-hand illusion over a black rubber hand (Maister et al., 2013), and these results and explanations for them were discussed in (Maister et al., 2015). Banakou et al. (2016) explored the impact of the number of exposures and the duration of the effect. The results of Peck et al. (2013) were replicated, but in Banakou et al. the racial-bias measure was taken one week after the final exposure in VR, suggesting the durability of the effect. These results stand in contrast to Groom et al. (2009), which had found an increase in the racial-bias IAT, as discussed above. Recent evidence suggests, however, that when the surrounding social situation is a negative one (in the case of Groom et al., 2009 a job interview) then the effect reverses. These results have been simulated through a neural network model (Bedder et al., 2019).

When adults are embodied in the body of a 5-year-old child with visuomotor synchrony they experience strong body ownership over the child body. As a result, they self-identify more with child-like attributes and see the surrounding world as larger (Banakou et al., 2013; Tajadura-Jiménez et al., 2017). However, when there is body ownership over an adult body of the same size as the child, the perceptual effects are significantly lower, suggesting that it is the form of the body that matters, not only the height. This result was also replicated (Tajadura-Jiménez et al., 2017).

Virtual embodiment allows us not only to experience having a different body, but to live through situations from a different perspective. One of these situations is the experience of a violent situation. In (Seinfeld et al., 2018) men who were domestic violence offenders experienced a virtual domestic violent confrontation from the perspective of the female victim. Such perspective has been found to modulate the brain network that encodes the bodily self (de Borst et al., 2020). Violent offenders often have deficits in emotion recognition, in the case of male offenders, with a deficit in recognizing fear in the faces of women. This deficit was found to be reduced after embodiment in the female subject to domestic abuse by a virtual man (Seinfeld et al., 2018). Similarly, mothers of young children tend to improve in empathy toward their children after spending some time embodied as a child in interaction with a virtual mother (Hamilton-Giachritsis et al., 2018).

Embodying virtual avatars can further influence the engagement and performance on a given task or situation,

depending on the perceived appropriateness of the embodied avatar for the task. For example, in a drumming task, participants showed more complex and articulated movement patterns when embodying a casually dressed avatar that when embodying a business man in a suit (Kilteni et al., 2013). Effects at the cognitive level have also been found. For example, participants embodied as Albert Einstein tend to improve their performance on a cognitive test than when embodied in another "ordinary" virtual body (Banakou et al., 2018). It has been shown that people embodied as Sigmund Freud tend to offer themselves better counseling than when embodied in a copy of their own body, or embodied as Freud with visuomotor asynchrony (Osimo et al., 2015; Slater et al., 2019). Moreover, being embodied as Lenin, the leader of the October 1917 Russian Revolution in a crowd scene leads to people being more likely to follow up on information about the Russian Revolution (Slater et al., 2018). All these studies form a body of accumulated evidence of the power of embodying virtual avatars not only in modifying physiological responses and perception and the world and the others, but also modifying behavior and cognitive performance.

6.1.2. Self-Avatar Impact on Agency, Self-Perception and Pain

It has been largely demonstrated that embodying a virtual avatar affects the way bodily related stimuli are processed. Experimental research based on the rubber hand illusion paradigm (Botvinick and Cohen, 1998) provided robust evidence for the impact of illusory body ownership on the perception of bodily related stimuli (Folegatti et al., 2009; Mancini et al., 2011; Zopf et al., 2011). Following this tradition, the embodiment of virtual avatars was shown to affect different facets of perception, from tactile processing to pain perception and own body image.

When in an immersive VR scenario, embodying a life-size virtual avatar enhances the perception of touch delivered on a held object, with respect to having no body in VR (Gonzalez-Franco and Berger, 2019). It was also shown that experiencing ownership toward a virtual avatar modulates the temporal constraints for associating two independent sensory cues, visual and tactile, to the same coherent visuo-tactile event (Maselli et al., 2016; Gonzalez-Franco and Berger, 2019). Virtual embodiment could therefore grant a larger flexibility to spatiotemporal offsets with respect to the constraints that apply in the physical world or when having an embodied self-avatar in VR.

Not only can embodiment modify the perception of timing of sensory inputs (Berger and Gonzalez-Franco, 2018; Gonzalez-Franco and Berger, 2019), but also the perception of other sensorial modalities such as temperature (Llobera et al., 2013b) and pain (Llobera et al., 2013a). Several studies have demonstrated that virtual embodiment can modulate pain (Lenggenhager et al., 2010; Martini et al., 2014). In particular, it has been shown that pain perception is modulated by the visual appearance of the virtual body, including its color (Martini et al., 2013), shape (Matamala-Gomez et al., 2020), or level of transparency (Martini et al., 2015), as well as by the degree of spatial overlap between the real and the virtual bodies (Nierula et al., 2017). This is a relevant area regarding chronic pain-therapeutical applications (Matamala-Gomez et al., 2019b), although pain of different origins may require different manipulations of the embodied-avatars (Matamala-Gomez et al., 2019a).

The embodiment of a virtual avatar further allows the temporal reshaping of peripersonal space, the space surrounding the body where external stimuli (e.g., visual or auditory) interact with the somatosensory system (Serino et al., 2006). This can be done by manipulating the visual perspective over an embodied avatar, as for the case of illusory out-of-body experiences (Lenggenhager et al., 2007; Blanke, 2012; Maselli and Slater, 2014), as well as by modulating the size and/or shape of the virtual body that interacts with this peripersonal space (Abtahi et al., 2019), for example by having an elongated virtual arm (Kilteni et al., 2012b; Feuchtner and Müller, 2017).

Such flexibility of the virtual body shape modification has been leveraged to study psychological phenomena such as body image in anorexia and other eating disorders (Piryankova et al., 2014; Mölbert et al., 2018) (**Figure 12**).

Agency is also an important element of the perceptual experience associated with the embodiment of self-avatars. Suppose you are embodied in a virtual body visuomotor synchrony and after a while the body does something that you did not do (in this case talk). It was found that participants have agency over the speaking, and the way they themselves speak later is influenced by how their self-avatar spoke (Banakou and Slater, 2014). The rigged self-avatar which has its movements driven by the movements of the participant was crucial to this result, since it was later found that if the body ownership is induced through tactile synchrony, then although the subjective illusory agency still occurs, it is not complemented by the behavioral after-effect (Banakou and Slater, 2017). Agency can also be induced when the real body does not move but the virtual body moves, after embodiment has been induced. This is the case when the virtual body is moved by means of a brain-computer interface (Nierula et al., 2019) or when seated participants can have the illusion of walking, based solely on the feedback from their embodiment in a walking virtual body (Kokkinara et al., 2016).

Interesting is also the case of pain perception. Several studies have demonstrated that virtual embodiment can modulate it (Lenggenhager et al., 2010; Martini et al., 2014) with important implication for pain treatment applications (Matamala-Gomez et al., 2019a,b). In particular, it was shown that pain perception is modulated by the visual appearance of the virtual body, including its color (Martini et al., 2013), shape (Matamala-Gomez et al., 2020), and level of transparency (Martini et al., 2015), as well as by the degree of spatial overlap between the real and the virtual bodies (Nierula et al., 2017).

6.2. Applications

The use and applications of avatars beyond basic scientific research is potentially as vast as the use of VR and AR and more generally computer graphics. In sum avatars are the main way to realistically represent humans, or in some situations, computational agents, inside digital content even if they are not displayed in immersive technologies, for example they are often part of animation movies or console games. Therefore, avatars have potential applications across many

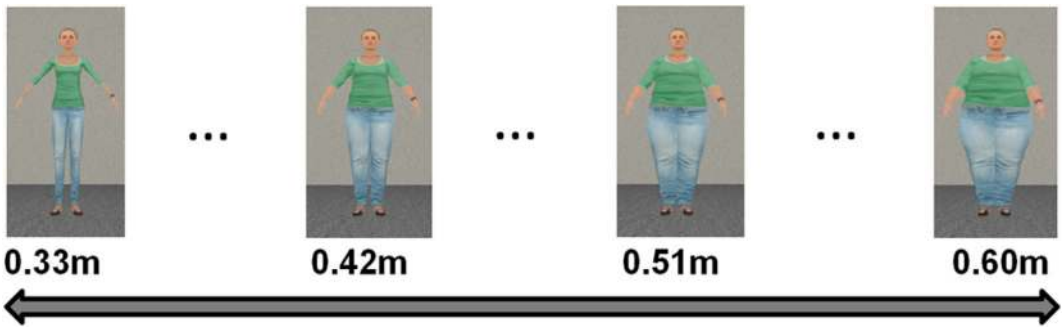

FIGURE 12 | A Microsoft Rocketbox avatar with modified mesh being used in a body weight study at Max Plank by Piryankova et al. (2014).

FIGURE 13 | One of the Microsoft Rocketbox avatars being used for a recreation of the Stanley Milgram experiment on obedience to authority at University College London by Gonzalez-Franco et al. (2019b).

areas including therapeutic treatments, training, education, and entertainment. In this section we will only explore two areas of application that have traditionally had a big emphasis on avatars: the entertainment industry and simulations. We will try to summarize why having access to high-quality rigged avatars is important for these fields. For more possible applications of avatars and VR in general we recommend a more in-depth review by Slater and Sanchez-Vives (2016).

6.2.1. Entertainment and Movies

Avatars or 3D characters can create a long-lasting emotional bond with the audience, and thus play an essential role in film-making, VR and computer games. For example, characters must exhibit clearly recognizable facial expressions that are consistent with their emotional state in the storyline (Aneja et al., 2018). Manually creating character animation requires expertise and hours of work. To speed up the animation process, we could use human actors to control and animate a character using a facial motion capture system (Li et al., 2015). Many of the techniques to create and animate avatars have been described above. However, there are particularities for this industry and despite recent advances in modeling capabilities, motion capture and control parameterization, most current animation studios still rely on artists manually producing high-quality frames. Most motion capture systems require skilled actors and laborious post-processing steps. The avatars need to match or exaggerate the physiology of performer limits to the possible motions (e.g., actor cannot perform exaggerated facial expressions).

Alleviating artist workload, but creating believable and compelling character motions, is arguably the central challenge in animated storytelling. Some professionals are using VR-based interfaces to pose and animate 3D characters: (PoseVR, 2019) or (Pan and Mitchell, 2020) developed by the Walt Disney Animation Studios, is a recent example in this direction. Meanwhile, a couple of VR-based research prototypes and commercial products have also recently been developed; however, these have mainly targeted to non-professional users (Cannavò et al., 2019). Alternatively, some researchers have proposed methods for generating 3D character expressions from humans in a geometrically consistent and perceptually valid manner using machine learning models (Aneja et al., 2018).

By open sourcing high quality rigs, the Microsoft Rocketbox avatar library, we are providing opportunities for researchers, engineers, and artists to work together to discover new tools and techniques that will shape the future of animated storytelling.

6.2.2. Avatars in Simulation

No matter how realistic a virtual environment looks from the point of view of geometry and rendering, it is essential that the virtual environment appears populated by realistic people and crowds in order to bring those virtual environments to life.

The needs for other people with whom a person interacts in VR/AR ranges from simulating one or two avatars to a whole crowd. Multiple studies have found that people behave realistically when interacting with avatars inside VR/AR. In recent years these realistic responses gained the interest of sociologists and psychologists who want to explore scientifically increasingly complex scenarios. Inside VR, researchers have replicated obedience to authority paradigms such as the Milgram experiments, that became almost impossible to run in real setups due to the ethical considerations regarding the deception scheme underlying the learner and punishment mechanisms (Slater et al.,

FIGURE 14 | Microsoft Rocketbox avatars being used for a recreation of a train station for studies of human behavior in crowded VR situations at the Universitat Politcnica de Catalunya, by Ríos and Pelechano (2020).

2006; Gonzalez-Franco et al., 2019b). Indeed, the replication of famous results such as the Milgram studies, further validate the use of avatars for social-psychological studies (**Figure 13**). Indeed, VR Milgram paradigms have recently been used to study empathy levels, predisposition and conformity to sexual harassment scenarios (Neyret et al., 2020).

Researchers have also used VR to create violent scenarios. For example, converting the avatar into a domestic abuser to see what the response of a real offender would be when exposed to this role scenario (Seinfeld et al., 2018), or studying what happens when a violent scenario has bystanders. In an experiment with soccer fans in a VR pub, researchers found a strong in-group, out-group effect for the victim of a soccer bulling interaction (Rovira et al., 2009). However, the response of the participants would vary depending on whether there were other bystander avatars and whether they were or not fans of the same soccer team (Slater et al., 2013). Some of these scenarios recreated in VR would be impossible in reality.

Besides the interactions with avatars in violent scenarios, researchers have also explored how users of different personalities interact with avatars. For example (Pan et al., 2012) studied how socially anxious and confident men interacted with a forward virtual woman, and how medical doctors respond to avatar patients who insist and demand unreasonably being treated with antibiotics (Pan et al., 2016).

Research in the area of social-psychology has also utilized avatars, for example to study moral dilemmas on how people would react when exposed to a shooting in a museum (Pan and Slater, 2011; Friedman et al., 2014). Or how different types of audiences would affect public speaking anxiety (Pertaub et al., 2002), and phobias (Botella et al., 2017).

Simulations have evolved from a different angle in the field of crowd simulation (**Figure 14**). In that area researchers have spent a great deal of effort in improving the algorithms that move agents smoothly between two points while avoiding collisions. However, no matter how close the simulation gets to real data, it is essential that each agent's position is then represented with a natural looking fully rigged avatar. The crowd simulation field has focused its work in the development of a large number of algorithms based on social forces (Helbing et al., 2000), geometrical rules (Pelechano et al., 2007), vision-based approaches (López et al., 2019), velocity vectors (Van den Berg et al., 2008), or data driven (Charalambous and Chrysanthou, 2014). Often psychological and personality traits can be included to add heterogeneity to the crowd (Pelechano et al., 2016). The output of a crowd simulation model is typically limited to just a position, the motion and sometimes some limited pose data such as a torso orientation. This type of output is repeatedly rendered as a crowd of moving points, or simple geometrical proxies such as 3D cylinders. Ideally the crowd simulation output should be seamlessly input ted into a system that could provide fully animated avatars with animations naturally matching the crowd trajectories. Research in real-time animation is not yet at the stage of providing a good real-time solution to this problem, but having high-quality fully rigged avatars is already a big step forward into making crowd simulation more realistic and thus, being ready to enhance the realism of immersive virtual scenarios.

Research efforts into simulations of few detailed humans and large crowds are gradually converging. The simulation research community needs realistic looking rigged avatars. For large crowds it also needs them to have flexibility in the number of polygons so that the cost of skinning and rendering does not become a bottleneck (Beacco et al., 2016). Natural looking avatars are not only critical to small simulations but can also greatly enhance the crowd simulation appearance when being rendered in 3D on a computer screen. This effect is even more important when being rendered in a HMD, where avatars are seen at eye level and from close distances.

Having realistic and well-formed avatars is especially relevant if we are studying human behavior in crowded spaces using immersive VR. Such setups can be used, for example, to evaluate human decision making during emergencies based on the behavior of the crowd (Ríos and Pelechano, 2020). Or to explore cognitive models of locomotion in crowded spaces (Thalmann, 2007; Luo et al., 2008; Olivier et al., 2014). Or for one-to-one interactions with avatars inside VR while performing locomotion (Pelechano et al., 2011; Ríos et al., 2018).

While in one-to-one interactions the use of facial expressions has been growing in use (Vinayagamoorthy et al., 2006; Gonzalez-Franco et al., 2020b). In current crowd simulations that very important aspect is missing. Current simulated virtual crowds appear as non-expressive avatars that simply look straight ahead while maneuvring around obstacles and other agents. There have been some attempts to introduce gaze so that

avatars appear to acknowledge the user's presence in the virtual environment but without changes in facial expression (Narang et al., 2016).

The Microsoft Rocketbox avatars not only provide a large set of avatars that can be animated for simulations but also provide the possibility of including facial expressions, which can open the doors to achieving virtual simulations of avatars and crowds that appear more lively and that can show real emotions. This will have a massive impact on the overall credibility of the crowd and will enrich the heterogeneity and realism of the populated virtual worlds.

7. CONCLUSIONS

This paper gives an overview the different pipelines that can be used to create 3D avatars that resemble humans, from mesh creation to rigging and animation, and from manual artist ic work to deep learning advancements. For each part of the pipeline we present the different options and outline the process complexities and pitfalls. In many cases there are references to the specific tools that the authors have used in the past for creating avatars themselves. These creation tools are also put into context of the Microsoft Rocketbox avatar library release with more details as to how these particular avatars were created and their limitations.

Furthermore, the paper reviews how these avatars are being used for scientific purposes and emphasizes the unique application of self-avatars that can be used to substitute the human subject's own body inside VR. Further details on applications in other fields such as crowd simulations and entertainment are depicted but for a full review of applications of avatars and VR in general see (Slater and Sanchez-Vives, 2016; Gonzalez-Franco and Lanier, 2017). Altogether, this paper conveys the many needs for rigged avatars that allow manipulation and animation in real time for the future of Virtual Reality. We anticipate that this widely used and freely available library and some of its important applications will enable novel research, and we encourage the research community to complete and share their research and/or enhanced tools based on this paper and avatar library with future publications.

DATA AVAILABILITY STATEMENT

The full Microsoft Rocketbox library is publicly available for research and academic use and can be downloaded in https://github.com/microsoft/Microsoft-Rocketbox (Mic, 2020). Microsoft Rocketbox avatars were a proprietary large and varied set of rigged avatars representing humans of different genders, races, occupations as well as some non-humans examples. Rocketbox Studios GmbH released three different libraries of avatars: starting with "Complete Characters" in 2005, and finally a new generation of highly detailed avatars and animations named "Complete Characters HD" from 2010 to 2015. It includes 115 characters and avatars, created over the course of these 10 years. Rocketbox Studios GmbH was then acquired by Havoc, which is now part of Microsoft. The entity that has now released the library free for academic and research purposes. The diversity of the characters and the quality of the rigging together with a relatively low-poly meshes, made this library the go-to asset among research laboratories worldwide from crowd simulation to real-time avatar embodiment and Virtual Reality (VR). Ever since their launch, laboratories around the globe have been using the library and many of the authors in this paper have extensively used these avatars during their research.

AUTHOR CONTRIBUTIONS

The sections dedicated to the creation of the avatars were mainly contributed by MG-F, EO, AS, YP, AA, BS, MW, LT, SO-E, and VO. The sections dedicated to reviewing avatar science and research were mainly contributed by AM, MS-V, MS, JB, JL, YP, NP, and MG-F. All authors have contributed to the writing of this paper.

FUNDING

MS-V and MS are funded by NEUROVIRTUAL-AGAUR (2017 SGR 1296). MS-V and Virtual Bodyworks are also supported by the European Union's Rights, Equality and Citizenship Programme (2014-2020) under Grant Agreement: 881712 (VRperGenere). DB and MS are supported by the European Research Council Advanced Grant MoTIVE #742989. NP was partly funded by the Spanish Ministry of Economy, Industry and Competitiveness under Grant No. TIN2017-88515-C2-1-R.

ACKNOWLEDGMENTS

The authors would like to thank also to the contribution to the creation of the original Rocketbox library of avatars to: Martin Beyer, Michael Dogan, Denis Dzienziol, Manu Eidner, Adrian Kumorowski, Johannes Ostrowitzki, Aleksander Roman, Artur Sabat, Robert Zimmermann, that together with Markus Wojcik created the original library. And to Dave Garagan from Havok for his help in the release of the public library. We also thank Tony Donegan for the language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frvir.2020.561558/full#supplementary-material

REFERENCES

(2020). *3d Studio Max*. Available online at: https://www.autodesk.com/products/3ds-max

(2020). *Adobe Fuse*. Available online at: https://www.adobe.com/products/fuse.html

(2020). *Binaryvr by Hyprsense*. Available online at: https://www.hyprsense.com/hyprface

(2020). *Blender*. Available online at: https://www.blender.org/
(2020). *Cats*. Available online at: http://vrcat.club/threads/cats-blender-plugin-0-14-0.6
(2020). *Maya*. Available online at: https://www.autodesk.com/products/maya
(2020). *Metamotion*. Available online at: https://metamotion.com/gypsy/gypsy-motion-capture-system.htm
(2020). *Microsoftrocketbox*. Available online at: https://github.com/microsoft/Microsoft-Rocketbox
(2020). *Mixamo*. Available online at: https://www.mixamo.com
(2020). *Optitrack*. Available online at: https://optitrack.com
(2020). *Vicon*. Available online at: https://vicon.com
Abtahi, P., Gonzalez-Franco, M., Ofek, E., and Steed, A. (2019). "I'm a giant: walking in large virtual environments at high speed gains," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–13. doi: 10.1145/3290605.3300752
Ahuja, K., Harrison, C., Goel, M., and Xiao, R. (2019). "MeCap: whole-body digitization for low-cost VR/AR headsets," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New orleans, LU), 453–462. doi: 10.1145/3332165.3347889
Aitpayev, K., and Gaber, J. (2012). Creation of 3D human avatar using kinect. *Asian Trans. Fundam. Electron. Commun. Multimed*. 1, 12–24.
Aneja, D., Chaudhuri, B., Colburn, A., Faigin, G., Shapiro, L., and Mones, B. (2018). "Learning to generate 3D stylized character expressions from humans," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (Lake Tahoe, CA: IEEE), 160–169. doi: 10.1109/WACV.2018.00024
Aristidou, A., Lasenby, J., Chrysanthou, Y., and Shamir, A. (2018). Inverse kinematics techniques in computer graphics: a survey. *Comput. Graph. Forum* 37, 35–58. doi: 10.1111/cgf.13310
Armel, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 1499–1506. doi: 10.1098/rspb.2003.2364
Aymerich-Franch, L. (2012). "Can we identify with a block? Identification with non-anthropomorphic avatars in virtual reality games," in *Proceedings of the International Society for Presence Research Annual Conference* (Philadelphia, CA), 24–26.
Badler, N. I., Hollick, M. J., and Granieri, J. P. (1993). Real-time control of a virtual human using minimal sensors. *Presence Teleoper. Virt. Environ.* 2, 82–86. doi: 10.1162/pres.1993.2.1.82
Bailenson, J., and Blascovich, J. (2004). Avatars. encyclopedia of human-computer interaction. *Berkshire Publ. Group* 64:68.
Banakou, D., Groten, R., and Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12846–12851. doi: 10.1073/pnas.1306779110
Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Front. Hum. Neurosci.* 10:601. doi: 10.3389/fnhum.2016.00601
Banakou, D., Kishore, S., and Slater, M. (2018). Virtually being einstein results in an improvement in cognitive task performance and a decrease in age bias. *Front. Psychol.* 9:917. doi: 10.3389/fpsyg.2018.00917
Banakou, D., and Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17678–17683. doi: 10.1073/pnas.1414936111
Banakou, D., and Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-14620-5
Baran, I., and Popović, J. (2007). Automatic rigging and animation of 3D characters. *ACM Trans. Graph.* 26:72. doi: 10.1145/1276377.1276467
Baumberg, A. (2002). "Blending images for texturing 3D models," in *Conference on British Machine Vision Association* (Cardiff), 404–413. doi: 10.5244/C.16.38
Beacco, A., Pelechano, N., and Andújar, C. (2016). A survey of real-time crowd rendering. *Comput. Graph. Forum* 35, 32–50. doi: 10.1111/cgf.12774
Bedder, R. L., Bush, D., Banakou, D., Peck, T., Slater, M., and Burgess, N. (2019). A mechanistic account of bodily resonance and implicit bias. *Cognition* 184, 1–10. doi: 10.1016/j.cognition.2018.11.010
Bem, D. J. (1972). Self-perception theory. *Adv. Exp. Soc. Psychol.* 6, 1–62. doi: 10.1016/S0065-2601(08)60024-6
Berger, C. C., and Gonzalez-Franco, M. (2018). "Expanding the sense of touch outside the body," in *Proceedings of the 15th ACM Symposium on Applied Perception* (ACM), 10. doi: 10.1145/3225153.3225172
Berger, C. C., Gonzalez-Franco, M., Ofek, E., and Hinckley, K. (2018). The uncanny valley of haptics. *Sci. Robot.* 3:eaar7010. doi: 10.1126/scirobotics.aar7010
Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A., Oberman, M., et al. (1990). "Reality built for two: a virtual reality tool," in *Proceedings of the 1990 Symposium on Interactive 3D Graphics* (Snowbird, UT), 35–36. doi: 10.1145/91385.91409
Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13:556. doi: 10.1038/nrn3292
Blanz, V., and Vetter, T. (1999). "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (Los Angeles, CA), 187–194. doi: 10.1145/311535.311556
Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., and Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Curr. Psychiatry Rep.* 19:42. doi: 10.1007/s11920-017-0788-4
Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391:756. doi: 10.1038/35784
Bouaziz, S., Wang, Y., and Pauly, M. (2013). Online modeling for realtime facial animation. *ACM Trans. Graph.* 32:40. doi: 10.1145/2461912.2461976
Cannavó, A., Demartini, C., Morra, L., and Lamberti, F. (2019). Immersive virtual reality-based interfaces for character animation. *IEEE Access* 7, 125463–125480. doi: 10.1109/ACCESS.2019.2939427
Cao, C., Weng, Y., Zhou, S., Tong, Y., and Zhou, K. (2013). Facewarehouse: a 3D facial expression database for visual computing. *IEEE Trans. Vis. Comput. Graph.* 20, 413–425. doi: 10.1109/TVCG.2013.249
Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). "Realtime multi-person 2D pose estimation using part affinity fields," in *Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 7291–7299. doi: 10.1109/CVPR.2017.143
Casas, D., Feng, A., Alexander, O., Fyffe, G., Debevec, P., Ichikari, R., et al. (2016). "Rapid photorealistic blendshape modeling from RGB-D sensors," in *Proceedings of the 29th International Conference on Computer Animation and Social Agents* (Geneva), 121–129. doi: 10.1145/2915926.2915936
Charalambous, P., and Chrysanthou, Y. (2014). The PAG crowd: a graph based approach for efficient data-driven crowd simulation. *Comput. Graph. Forum* 33, 95–108. doi: 10.1111/cgf.12403
Collet, A., Chuang, M., Sweeney, P., Gillett, D., Evseev, D., Calabrese, D., et al. (2015). High-quality streamable free-viewpoint video. *ACM Trans. Graph.* 34:69. doi: 10.1145/2766945
de Borst, A. W., Sanchez-Vives, M. V., Slater, M., and de Gelder, B. (2020). First person virtual embodiment modulates cortical network that encodes the bodily self and its surrounding space during the experience of domestic violence. *eNeuro* 7:ENEURO.0263-19.2019. doi: 10.1523/ENEURO.0263-19.2019
De Vignemont, F. (2011). Embodiment, ownership and disownership. *Conscious. Cogn.* 20, 82–93. doi: 10.1016/j.concog.2010.09.004
Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagarz, M. (2000). "Acquiring the reflectance field of a human face," in *SIGGRAPH* (New Orleans, LU). doi: 10.1145/344779.344855
Dou, M., Davidson, P. L., Fanello, S. R., Khamis, S., Kowdle, A., Rhemann, C., et al. (2017). Motion2fusion: real-time volumetric performance capture. *ACM Trans. Graph.* 36, 246:1–246:16. doi: 10.1145/3130800.3130801
Dou, M., Khamis, S., Degtyarev, Y., Davidson, P. L., Fanello, S. R., Kowdle, A., et al. (2016). Fusion 4D: real-time performance capture of challenging scenes. *ACM Trans. Graph.* 35, 114:1–114:13. doi: 10.1145/2897824.2925969
Ehrsson, H. H. (2012). "The concept of body ownership and its relation to multisensory integration," in *The New Handbook of Multisensory Process*, ed B. E. Stein (Cambridge, MA: MIT Press), 775–792.
Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., and Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9828–9833. doi: 10.1073/pnas.0610011104
Ekman, P., and Friesen, W. V. (1976). Measuring facial movement. *Environ. Psychol. Nonverbal Behav.* 1, 56–75. doi: 10.1007/BF01115465
Esteban, C. H., and Schmitt, F. (2004). Silhouette and stereo fusion for 3D object modeling. *Comput. Vis. Image Understand.* 96, 367–392. doi: 10.1016/j.cviu.2004.03.016

Falconer, C. J., Rovira, A., King, J. A., Gilbert, P., Antley, A., Fearon, P., et al. (2016). Embodying self-compassion within virtual reality and its effects on patients with depression. *BJPsych. Open* 2, 74–80. doi: 10.1192/bjpo.bp.115.002147

Falconer, C. J., Slater, M., Rovira, A., King, J. A., Gilbert, P., Antley, A., et al. (2014). Embodying compassion: a virtual reality paradigm for overcoming excessive self-criticism. *PLoS ONE* 9:e111933. doi: 10.1371/journal.pone.0111933

Feng, A., Casas, D., and Shapiro, A. (2015). "Avatar reshaping and automatic rigging using a deformable model," in *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games* (Los Angeles, CA), 57–64. doi: 10.1145/2822013.2822017

Feuchtner, T., and Müller, J. (2017). "Extending the body for interaction with reality," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO), 5145–5157. doi: 10.1145/3025453.3025689

Folegatti, A., De Vignemont, F., Pavani, F., Rossetti, Y., and Farné, A. (2009). Losing one's hand: visual-proprioceptive conflict affects touch perception. *PLoS ONE* 4:e6920. doi: 10.1371/journal.pone.0006920

Friedman, D., Pizarro, R., Or-Berkers, K., Neyret, S., Pan, X., and Slater, M. (2014). A method for generating an illusion of backwards time travel using immersive virtual reality-an exploratory study. *Front. Psychol.* 5:943. doi: 10.3389/fpsyg.2014.00943

Frueh, C., Sud, A., and Kwatra, V. (2017). "Headset removal for virtual and mixed reality," in *SIGGRAPH Talks 2017* (Los Angeles, CA). doi: 10.1145/3084363.3085083

Gal, R., Wexler, Y., Ofek, E., Hoppe, H., and Cohen-Or, D. (2010). "Seamless montage for texturing models," in *EuroGraphics* (Norrköping). doi: 10.1111/j.1467-8659.2009.01617.x

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (2003). "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Fort Lauderdale, FL), 529–536. doi: 10.1145/642611.642703

Gonzalez-Franco, M., Abtahi, P., and Steed, A. (2019a). "Individual differences in embodied distance estimation in virtual reality," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka: IEEE), 941–943. doi: 10.1109/VR.2019.8798348

Gonzalez-Franco, M., Bellido, A. I., Blom, K. J., Slater, M., and Rodriguez-Fornells, A. (2016). The neurological traces of look-alike avatars. *Front. Hum. Neurosci.* 10:392. doi: 10.3389/fnhum.2016.00392

Gonzalez-Franco, M., and Berger, C. C. (2019). Avatar embodiment enhances haptic confidence on the out-of-body touch illusion. *IEEE Trans. Haptics* 12, 319–326. doi: 10.1109/TOH.2019.2925038

Gonzalez-Franco, M., Cohn, B., Ofek, E., Burin, D., and Maselli, A. (2020a). "The self-avatar follower effect in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Atlanta, GA). doi: 10.1109/VR46266.2020.00019

Gonzalez-Franco, M., and Lanier, J. (2017). Model of illusions and virtual reality. *Front. Psychol.* 8:1125. doi: 10.3389/fpsyg.2017.01125

Gonzalez-Franco, M., and Peck, T. C. (2018). Avatar embodiment. Towards a standardized questionnaire. *Front. Robot. AI* 5:74. doi: 10.3389/frobt.2018.00074

González-Franco, M., Peck, T. C., Rodríguez-Fornells, A., and Slater, M. (2014). A threat to a virtual hand elicits motor cortex activation. *Exp. Brain Res.* 232, 875–887. doi: 10.1007/s00221-013-3800-1

Gonzalez-Franco, M., Perez-Marcos, D., Spanlang, B., and Slater, M. (2010). "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)* (Waltham, MA: IEEE), 111–114. doi: 10.1109/VR.2010.5444805

Gonzalez-Franco, M., Slater, M., Birney, M. E., Swapp, D., Haslam, S. A., and Reicher, S. D. (2019b). Participant concerns for the learner in a virtual reality replication of the milgram obedience study. *PLoS ONE* 13:e209704. doi: 10.1371/journal.pone.0209704

Gonzalez-Franco, M., Steed, A., Hoogendyk, S., and Ofek, E. (2020b). Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE Trans. Vis. Comput. Graph.* 26, 2023–2029. doi: 10.1109/TVCG.2020.2973075

Graziano, M. S., and Botvinick, M. (2002). *How the Brain Represents the Body: Insights From Neurophysiology and Psychology*. Oxford: Oxford University Press.

Groom, V., Bailenson, J. N., and Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Soc. Influence* 4, 231–248. doi: 10.1080/15534510802643750

Guo, K., Lincoln, P., Davidson, P. L., Busch, J., Yu, X., Whalen, M., et al. (2019). The relightables: volumetric performance capture of humans with realistic relighting. *ACM Trans. Graph.* 38, 217:1–217:19. doi: 10.1145/3355089.3356571

Ha, S., Bai, Y., and Liu, C. K. (2011). "Human motion reconstruction from force sensors," in *Symposium on Computer Animation (SCA '11)* (Vancouver, BC), 129–138. doi: 10.1145/2019406.2019424

Hamilton-Giachritsis, C., Banakou, D., Quiroga, M. G., Giachritsis, C., and Slater, M. (2018). Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-21036-2

Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS ONE* 12:e174965. doi: 10.1371/journal.pone.0174965

Helbing, D., Farkas, I., and Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature* 407, 487–490. doi: 10.1038/35035023

Holden, D., Komura, T., and Saito, J. (2017). Phase-functioned neural networks for character control. *ACM Trans. Graph.* 36, 1–13. doi: 10.1145/3072959.3073663

Hu, L., Saito, S., Wei, L., Nagano, K., Seo, J., Fursund, J., et al. (2017). Avatar digitization from a single image for real-time rendering. *ACM Trans. Graph.* 36, 1–14. doi: 10.1145/3130800.31310887

Ichim, A. E., Bouaziz, S., and Pauly, M. (2015). Dynamic 3D avatar creation from hand-held video input. *ACM Trans. Graph.* 34, 1–14. doi: 10.1145/2766974

Joshi, P., Tien, W. C., Desbrun, M., and Pighin, F. (2006). "Learning controls for blend shape based realistic facial animation," in *ACM Siggraph 2006 Courses* (Boston, MA), 17. doi: 10.1145/1185657.1185857

Karras, T., Laine, S., and Aila, T. (2019). "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA), 4401–4410. doi: 10.1109/CVPR.2019.00453

Kilteni, K., Bergstrom, I., and Slater, M. (2013). Drumming in immersive virtual reality: the body shapes the way we play. *IEEE Trans. Vis. Comput. Graph.* 19, 597–605. doi: 10.1109/TVCG.2013.29

Kilteni, K., Groten, R., and Slater, M. (2012a). The sense of embodiment in virtual reality. *Presence Teleoper. Virtual Environ.* 21, 373–387. doi: 10.1162/PRES_a_00124

Kilteni, K., Maselli, A., Kording, K. P., and Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* 9:141. doi: 10.3389/fnhum.2015.00141

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., and Slater, M. (2012b). Extending body space in immersive virtual reality: a very long arm illusion. *PLoS ONE* 7:e40867. doi: 10.1371/journal.pone.0040867

Kobbelt, L., and Botsch, M. (2004). A survey of point-based techniques in computer graphics. *Comput. Graph.* 28, 801–814. doi: 10.1016/j.cag.2004.08.009

Kokkinara, E., Kilteni, K., Blom, K. J., and Slater, M. (2016). First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep28879

Kokkinara, E., and Slater, M. (2014). Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception* 43, 43–58. doi: 10.1068/p7545

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE* 2:e943. doi: 10.1371/journal.pone.0000943

Lanier, J. (1990). Riding the giant worm to saturn: post-symbolic communication in virtual reality. *ARS Electron.* 2, 186–188.

Lanier, J. (2001). Virtually there. *Sci. Am.* 284, 66–75. doi: 10.1038/scientificamerican0401-66

Lanier, J., Lasko-Harvill, A., Blanchard, C., Smithers, W., Harvill, Y., and Coffman, A. (1988). "From dataglove to datasuit," in *Digest of Papers. COMPCON Spring 88 Thirty-Third IEEE Computer Society International Conference* (IEEE), 536–538. doi: 10.1109/CMPCON.1988.4925

Lee, Y., Kim, S., and Lee, J. (2010). "Data-driven biped control," in *ACM SIGGRAPH 2010 Papers* (Los Angeles, CA), 1–8. doi: 10.1145/1833349.1781155

Lempitsky, V., and Ivanov, D. (2007). "Seamless mosaicing of image-based texture maps," in *Computer Vision and Pattern Recognition* (Minneapolis, MN).

Lenggenhager, B., Hänsel, A., von Känel, R., Curatolo, M., and Blanke, O. (2010). *Analgesic Effects of Illusory Self-Perception*.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099. doi: 10.1126/science.1143439

Lewis, J. P., Anjyo, K., Rhee, T., Zhang, M., Pighin, F. H., and Deng, Z. (2014). Practice and theory of blendshape facial models. *Eurographics* 1:2. doi: 10.2312/egst.20141042

Li, H., Trutoiu, L., Olszewski, K., Wei, L., Trutna, T., Hsieh, P.-L., et al. (2015). Facial performance sensing head-mounted display. *ACM Trans. Graph.* 34:47. doi: 10.1145/2766939

Liu, L., and Hodgins, J. (2018). Learning basketball dribbling skills using trajectory optimization and deep reinforcement learning. *ACM Trans. Graph.* 37, 1–14. doi: 10.1145/3197517.3201315

Llobera, J., González-Franco, M., Perez-Marcos, D., Valls-Solé, J., Slater, M., and Sanchez-Vives, M. V. (2013a). Virtual reality for assessment of patients suffering chronic pain: a case study. *Exp. Brain Res.* 225, 105–117. doi: 10.1007/s00221-012-3352-9

Llobera, J., Sanchez-Vives, M. V., and Slater, M. (2013b). The relationship between virtual body ownership and temperature sensitivity. *J. R. Soc. Interface* 10:20130300. doi: 10.1098/rsif.2013.0300

Lombardi, S., Saragih, J., Simon, T., and Sheikh, Y. (2018). Deep appearance models for face rendering. *ACM Trans. Graph.* 37, 68:1–68:13. doi: 10.1145/3197517.3201401

Lombardi, S., Simon, T., Saragih, J. M., Schwartz, G., Lehrmann, A. M., and Sheikh, Y. (2019). Neural volumes: Learning dynamic renderable volumes from images. *CoRR* abs/1906.07751. doi: 10.1145/3306346.3323020

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004

Loop, C. T., Cai, Q., Orts-Escolano, S., and Chou, P. A. (2016). "A closed-form bayesian fusion equation using occupancy probabilities," in *Fourth International Conference on 3D Vision, 3DV 2016* (Stanford, CA: IEEE Computer Society), 380–388. doi: 10.1109/3DV.2016.47

López, A., Chaumette, F., Marchand, E., and Pettré, J. (2019). Character navigation in dynamic environments based on optical flow. *Comput. Graph. Forum* 38, 181–192. doi: 10.1111/cgf.13629

Lou, J., Wang, Y., Nduka, C., Hamedi, M., Mavridou, I., Wang, F.-Y., et al. (2019). Realistic facial expression reconstruction for VR HMD users. *IEEE Trans. Multimed.* 22, 730–743. doi: 10.1109/TMM.2019.2933338

Luo, L., Zhou, S., Cai, W., Low, M. Y. H., Tian, F., Wang, Y., et al. (2008). Agent-based human behavior modeling for crowd simulation. *Comput. Anim. Virtual Worlds* 19, 271–281. doi: 10.1002/cav.238

Ma, Y., Paterson, M. H., and Pollick, E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behav. Res. Methods* 38, 134–141. doi: 10.3758/BF03192758

MacQuarrie, A., and Steed, A. (2019). "Perception of volumetric characters' eye-gaze direction in head-mounted displays," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Osaka: IEEE), 645–654. doi: 10.1109/VR.2019.8797852

Maister, L., Sebanz, N., Knoblich, G., and Tsakiris, M. (2013). Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition* 128, 170–178. doi: 10.1016/j.cognition.2013.04.002

Maister, L., Slater, M., Sanchez-Vives, M. V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn. Sci.* 19, 6–12. doi: 10.1016/j.tics.2014.11.001

Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behav. Brain Res.* 191, 1–10. doi: 10.1016/j.bbr.2008.02.041

Mancini, F., Longo, M. R., Kammers, M. P., and Haggard, P. (2011). Visual distortion of body size modulates pain perception. *Psychol. Sci.* 22, 325–330. doi: 10.1177/0956797611398496

Martini, M., Kilteni, K., Maselli, A., and Sanchez-Vives, M. V. (2015). The body fades away: investigating the effects of transparency of an embodied virtual body on pain threshold and body ownership. *Sci. Rep.* 5:13948. doi: 10.1038/srep13948

Martini, M., Pérez Marcos, D., and Sanchez-Vives, M. V. (2013). What color is my arm? Changes in skin color of an embodied virtual arm modulates pain threshold. *Front. Hum. Neurosci.* 7:438. doi: 10.3389/fnhum.2013.00438

Martini, M., Perez-Marcos, D., and Sanchez-Vives, M. V. (2014). Modulation of pain threshold by virtual body ownership. *Eur. J. Pain* 18, 1040–1048. doi: 10.1002/j.1532-2149.2014.00451.x

Maselli, A., Kilteni, K., López-Moliner, J., and Slater, M. (2016). The sense of body ownership relaxes temporal constraints for multisensory integration. *Sci. Rep.* 6:30628. doi: 10.1038/srep30628

Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7:83. doi: 10.3389/fnhum.2013.00083

Maselli, A., and Slater, M. (2014). Sliding perspectives: dissociating ownership from self-location during full body illusions in virtual reality. *Front. Hum. Neurosci.* 8:693. doi: 10.3389/fnhum.2014.00693

Matamala-Gomez, M., Donegan, T., Bottiroli, S., Sandrini, G., Sanchez-Vives, M. V., and Tassorelli, C. (2019a). Immersive virtual reality and virtual embodiment for pain relief. *Front. Hum. Neurosci.* 13:279. doi: 10.3389/fnhum.2019.00279

Matamala-Gomez, M., Gonzalez, A. M. D., Slater, M., and Sanchez-Vives, M. V. (2019b). Decreasing pain ratings in chronic arm pain through changing a virtual body: different strategies for different pain types. *J. Pain* 20, 685–697. doi: 10.1016/j.jpain.2018.12.001

Matamala-Gomez, M., Nierula, B., Donegan, T., Slater, M., and Sanchez-Vives, M. V. (2020). Manipulating the perceived shape and color of a virtual limb can modulate pain responses. *J. Clin. Med.* 9:291. doi: 10.3390/jcm9020291

Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., et al. (2017). Vnect: real-time 3D human pose estimation with a single rgb camera. *ACM Trans. Graph.* 36, 7291–7299. doi: 10.1145/3072959.3073596

Moeslund, B. T., Hilton, A., and Krügerc, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Comput. Visi. Image Understand.* 104, 90–126. doi: 10.1016/j.cviu.2006.08.002

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., and Bülthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence Teleoper. Virtual Environ.* 19, 230–242. doi: 10.1162/pres.19.3.230

Mölbert, S. C., Thaler, A., Mohler, B. J., Streuber, S., Romero, J., Black, M. J., et al. (2018). Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: attitudinal components rather than visual body size estimation are distorted. *Psychol. Med.* 48, 642–653. doi: 10.1017/S0033291717002008

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Narang, S., Best, A., Randhavane, T., Shapiro, A., and Manocha, D. (2016). "Pedvr: simulating gaze-based interactions between a real user and virtual crowds," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich), 91–100. doi: 10.1145/2993369.2993378

Neyret, S., Navarro, X., Beacco, A., Oliva, R., Bourdin, P., Valenzuela, J., et al. (2020). An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Sci. Rep.* 10, 1–18. doi: 10.1038/s41598-020-62932-w

Nierula, B., Martini, M., Matamala-Gomez, M., Slater, M., and Sanchez-Vives, M. V. (2017). Seeing an embodied virtual hand is analgesic contingent on colocation. *J. Pain* 18, 645–655. doi: 10.1016/j.jpain.2017.01.003

Nierula, B., Spanlang, B., Martini, M., Borrell, M., Nikulin, V. V., and Sanchez-Vives, M. V. (2019). Agency and responsibility over virtual movements controlled through different paradigms of brain-computer interface. *J. Physiol.* 1, 1–42. doi: 10.1113/JP278167

Olivier, A.-H., Bruneau, J., Cirio, G., and Pettré, J. (2014). A virtual reality platform to study crowd behaviors. *Transport. Res. Proc.* 2, 114–122. doi: 10.1016/j.trpro.2014.09.015

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., et al. (2016). "Holoportation: virtual 3D teleportation in real-time," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo: ACM), 741–754. doi: 10.1145/2984511.2984517

Orvalho, V., Bastos, P., Parke, F. I., Oliveira, B., and Alvarez, X. (2012). "A facial rigging survey," in *Eurographics (STARs)* (Cagliari), 183–204.

Osimo, S. A., Pizarro, R., Spanlang, B., and Slater, M. (2015). Conversations between self and self as sigmund freud-a virtual body ownership paradigm for self counselling. *Sci. Rep.* 5:13899. doi: 10.1038/srep13899

Overbeck, R. S., Erickson, D., Evangelakos, D., Pharr, M., and Debevec, P. (2018). A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. *ACM Trans. Graph*. 37:197. doi: 10.1145/3272127.3275031

Padrao, G., Gonzalez-Franco, M., Sanchez-Vives, M. V., Slater, M., and Rodriguez-Fornells, A. (2016). Violating body movement semantics: neural signatures of self-generated and external-generated errors. *Neuroimage* 124, 147–156. doi: 10.1016/j.neuroimage.2015.08.022

Pan, X., Gillies, M., Barker, C., Clark, D. M., and Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: an experimental study. *PLoS ONE* 7:e32931. doi: 10.1371/journal.pone.0032931

Pan, X., and Slater, M. (2011). "Confronting a moral dilemma in virtual reality: a pilot study," in *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction* (Newcastle), 46–51. doi: 10.14236/ewic/HCI2011.26

Pan, X., Slater, M., Beacco, A., Navarro, X., Bellido Rivas, A. I., Swapp, D., et al. (2016). The responses of medical general practitioners to unreasonable patient demand for antibiotics–a study of medical ethics using immersive virtual reality. *PLoS ONE* 11:e146837. doi: 10.1371/journal.pone.0146837

Pan, Y., and Mitchell, K. (2020). "PoseMMR: a collaborative mixed reality authoring tool for character animation," in *IEEE Virtual Reality* (Atlanta, GA). doi: 10.1109/VRW50115.2020.00230

Pan, Y., Steptoe, W., and Steed, A. (2014). "Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (ACM), 1397–1406. doi: 10.1145/2556288.2557276

Pandey, R., Tkach, A., Yang, S., Pidlypenskyi, P., Taylor, J., Martin-Brualla, R., et al. (2019). Volumetric capture of humans with a single RGBD camera via semi-parametric learning. *CoRR* abs/1905.12162. doi: 10.1109/CVPR.2019.00994

Parger, M., Mueller, J. H., Schmalstieg, D., and Steinberger, M. (2018). "Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo), 1–10. doi: 10.1145/3281505.3281529

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn*. 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Pelechano, N., Allbeck, J. M., and Badler, N. I. (2007). "Controlling individual agents in high-density crowd simulation," in *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Prague: Eurographics Association), 99–108.

Pelechano, N., Allbeck, J. M., Kapadia, M., and Badler, N. I. (2016). *Simulating Heterogeneous Crowds With Interactive Behaviors*. CRC Press.

Pelechano, N., Spanlang, B., and Beacco, A. (2011). Avatar locomotion in crowd simulation. *Int. J. Virtual Reality* 10, 13–19. doi: 10.20870/IJVR.2011.10.1.2796

Peng, X. B., Abbeel, P., Levine, S., and van de Panne, M. (2018). Deepmimic: example-guided deep reinforcement learning of physics-based character skills. *ACM Trans. Graph*. 37, 1–14. doi: 10.1145/3197517.3201311

Pertaub, D.-P., Slater, M., and Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence Teleoper. Virtual Environ*. 11, 68–78. doi: 10.1162/105474602317343668

Petkova, V. I., and Ehrsson, H. H. (2008). If i were you: perceptual illusion of body swapping. *PLoS ONE* 3:e3832. doi: 10.1371/journal.pone.0003832

Petkova, V. I., Khoshnevis, M., and Ehrsson, H. H. (2011). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol*. 2:35. doi: 10.3389/fpsyg.2011.00035

Phillips, L., Ries, B., Kaeding, M., and Interrante, V. (2010). "Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments," in *2010 IEEE Virtual Reality Conference (VR)* (Waltham, MA: IEEE), 115–1148. doi: 10.1109/VR.2010.5444802

Piryankova, I. V., Wong, H. Y., Linkenauger, S. A., Stinson, C., Longo, M. R., Bülthoff, H. H., et al. (2014). Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PLoS ONE* 9:e103428. doi: 10.1371/journal.pone.0103428

PoseVR (2019). *Walt Disney Animation Studios*. PoseVR.

Pujades, S., Mohler, B., Thaler, A., Tesch, J., Mahmood, N., Hesse, N., et al. (2019). The virtual caliper: rapid creation of metrically accurate avatars from 3d measurements. *IEEE Trans. Vis. Comput. Graph*. 25, 1887–1897. doi: 10.1109/TVCG.2019.2898748

Ríos, A., Palomar, M., and Pelechano, N. (2018). "Users' locomotor behavior in collaborative virtual reality," in *Proceedings of the 11th Annual International Conference on Motion, Interaction, and Games* (Limassol), 1–9. doi: 10.1145/3274247.3274513

Ríos, A., and Pelechano, N. (2020). Follower behavior under stress in immersive VR. *Virtual Reality* 1, 1–12. doi: 10.1007/s10055-020-00428-8

Roth, D., Lugrin, J., von Mammen, S., and Latoschik, M. (2017). "Controllers & inputs: masters of puppets," in *Avatar, Assembled: The Social and Technical Anatomy of Digital Bodies*, eds J. Bank (New York, NY: Peter Lang), 281–290.

Roth, D., Lugrin, J.-L., Büser, J., Bente, G., Fuhrmann, A., and Latoschik, M. E. (2016). "A simplified inverse kinematic approach for embodied VR applications," in *2016 IEEE Virtual Reality (VR)* (Greenville, SC:IEEE), 275–76. doi: 10.1109/VR.2016.7504760

Roth, D., Stauffert, J.-P., and Latoschik, M. E. (2019). "Avatar embodiment, behavior replication, and kinematics in virtual reality," in *VR Developer Gems* (AK Peters; CRC Press), 321–346. doi: 10.1201/b21598-17

Rovira, A., Swapp, D., Spanlang, B., and Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Front. Behav. Neurosci*. 3:59. doi: 10.3389/neuro.08.059.2009

Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., and Li, H. (2019). "Pifu: pixel-aligned implicit function for high-resolution clothed human digitization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2304–2314. doi: 10.1109/ICCV.2019.00239

Saito, S., Wei, L., Fursund, J., Hu, L., Yang, C., Yu, R., et al. (2016). "Pinscreen: 3D avatar from a single image," in *SIGGRAPH Asia Emerging Technologies* (Seoul). doi: 10.1145/2988240.3014572

Salmanowitz, N. (2018). The impact of virtual reality on implicit racial bias and mock legal decisions. *J. Law Biosci*. 5, 174–203. doi: 10.1093/jlb/lsy005

Samad, M., Chung, A. J., and Shams, L. (2015). Perception of body ownership is driven by bayesian sensory inference. *PLoS ONE* 10:e117178. doi: 10.1371/journal.pone.0117178

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci*. 6:332. doi: 10.1038/nrn1651

Sanchez-Vives, M. V., Spanlang, B., Frisoli, A., Bergamasco, M., and Slater, M. (2010). Virtual hand illusion induced by visuomotor correlations. *PLoS ONE* 5: e10381. doi: 10.1371/journal.pone.0010381

Schroeder, R. (2012). *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Springer Science & Business Media.

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L., Borland, D., et al. (2018). Offenders become the victim in virtual reality: impact of changing perspective in domestic violence. *Sci. Rep*. 8:2692. doi: 10.1038/s41598-018-19987-7

Serino, A., Farné, A., and Ládavas, E. (2006). Visual peripersonal space. *Adv. Conscious. Res*. 66:323. doi: 10.1075/aicr.66.24ser

Shams, L., and Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn. Sci*. 14, 425–432. doi: 10.1016/j.tics.2010.07.001

Shiratori, T., Park, H. S., Sigal, L., Sheikh, Y., and Hodgins, J. K. (2011). "Motion capture from body-mounted cameras," in *ACM SIGGRAPH 2011 Papers* (Vancouver, BC), 1–10. doi: 10.1145/2010324.1964926

Shysheya, A., Zakharov, E., Aliev, K.-A., Bashirov, R., Burkov, E., Iskakov, K., et al. (2019). "Textured neural avatars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA), 2387–2397. doi: 10.1109/CVPR.2019.00249

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci*. 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., et al. (2006). A virtual reprise of the stanley milgram obedience experiments. *PLoS ONE* 1:e39. doi: 10.1371/journal.pone.0000039

Slater, M., Navarro, X., Valenzuela, J., Oliva, R., Beacco, A., Thorn, J., et al. (2018). Virtually being lenin enhances presence and engagement in a scene from the russian revolution. *Front. Robot. AI* 5:91. doi: 10.3389/frobt.2018.00091

Slater, M., Neyret, S., Johnston, T., Iruretagoyena, G., de la Campa Crespo, M. Á., Alabérnia-Segura, M., et al. (2019). An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. *Sci. Rep*. 9, 1–13. doi: 10.1038/s41598-019-46877-3

Slater, M., Pérez Marcos, D., Ehrsson, H., and Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Front. Neurosci*. 3:29. doi: 10.3389/neuro.01.029.2009

Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., et al. (2013). Bystander responses to a violent incident in an immersive virtual environment. *PLoS ONE* 8:e52766. doi: 10.1371/journal.pone.0052766

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074

Slater, M., Spanlang, B., and Corominas, D. (2010a). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph*. 29, 1–9. doi: 10.1145/1778765.1778829

Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010b). First person experience of body transfer in virtual reality. *PLoS ONE* 5:e10564. doi: 10.1371/journal.pone.0010564

Spanlang, B., Navarro, X., Normand, J.-M., Kishore, S., Pizarro, R., and Slater, M. (2013). "Real time whole body motion mapping for avatars and robots," in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology* (Singapore), 175–178. doi: 10.1145/2503713.2503747

Spanlang, B., Normand, J.-M., Borland, D., Kilteni, K., Giannopoulos, E., Pomés, A., et al. (2014). How to build an embodiment lab: achieving body representation illusions in virtual reality. *Front. Robot. AI* 1:9. doi: 10.3389/frobt.2014.00009

Steed, A., Frlston, S., Lopez, M. M., Drummond, J., Pan, Y., and Swapp, D. (2016a). An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Trans. Vis. Comput. Graph*. 22, 1406–1414. doi: 10.1109/TVCG.2016.2518135

Steed, A., Pan, Y., Zisch, F., and Steptoe, W. (2016b). "The impact of a self-avatar on cognitive load in immersive virtual reality," in *2016 IEEE Virtual Reality (VR)* (Greenville, SC:IEEE), 67–76. doi: 10.1109/VR.2016.7504689

Steve, S., Curless, B., J., D., and Rick, S. D. S. (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Conference on Computer Vision and Pattern Recognition* (New York, NY).

Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., and Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Sci. Rep*. 7:9637. doi: 10.1038/s41598-017-09497-3

Thalmann, D. (2007). *Crowd Simulation*. Wiley Encyclopedia of Computer Science and Engineering. doi: 10.1002/9780470050118.ecse676

Thorn, J., Pizarro, R., Spanlang, B., Bermell-Garcia, P., and Gonzalez-Franco, M. (2016). "Assessing 3D scan quality through paired-comparisons psychophysics," in *Proceedings of the 24th ACM International Conference on Multimedia, MM '16* (Amsterdam: Association for Computing Machinery), 147–151. doi: 10.1145/2964284.2967200

Van den Berg, J., Lin, M., and Manocha, D. (2008). "Reciprocal velocity obstacles for real-time multi-agent navigation," in *2008 IEEE International Conference on Robotics and Automation* (Pasadena, CA: IEEE), 1928–1935. doi: 10.1109/ROBOT.2008.4543489

Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., et al. (2006). "Building expression into virtual characters," in *Eurographics* (Vienna).

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph*. 24, 1643–1652. doi: 10.1109/TVCG.2018.2794629

Wang, L., Kang S. B., Szeliski, R., and Shum, H. Y. (2001). "Optimal texture map reconstruction from multiple views," in *Computer Vision and Pattern Recognition* (Kauai, HI).

Wang, T.-C., Zhu, J.-Y., Kalantari, N. K., Efros, A. A., and Ramamoorthi, R. (2017). Light field video capture using a learning-based hybrid imaging system. *ACM Trans. Graph*. 36:133. doi: 10.1145/3072959.3073614

Wei, X., Zhang, P., and Chai, J. (2012). Accurate realtime full-body motion capture using a single depth camera. *ACM Trans. Graph*. 31, 1–12. doi: 10.1145/2366145.2366207

Wei, Y., Ofek, E., Quan, L., and Shum, H. (2005). "Modeling hair from multiple views," in *SIGGRAPH* (Los Angeles, CA). doi: 10.1145/1187112.1187292

Wei, Y., Ofek, E., Quan, L., and Shum, H.-Y. (2019). Realistic facial expression reconstruction for VR hmd users. *ACM Trans. Graph*.

Weng, C.-Y., Curless, B., and Kemelmacher-Shlizerman, I. (2019). "Photo wake-up: 3D character animation from a single photo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Long Beach, CA), 5908–5917. doi: 10.1109/CVPR.2019.00606

Yee, N., and Bailenson, J. (2007). The proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res*. 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x

Yee, N., and Bailenson, J. N. (2009). The difference between being and seeing: the relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychol*. 12, 195–209. doi: 10.1080/15213260902849943

Yee, N., Bailenson, J. N., and Ducheneaut, N. (2009). The proteus effect: implications of transformed digital self-representation on online and offline behavior. *Commun. Res*. 36, 285–312. doi: 10.1177/0093650208330254

Yuan, Y., and Steed, A. (2010). "Is the rubber hand illusion induced by immersive virtual reality?" in *2010 IEEE Virtual Reality Conference (VR)* (Waltham, MA: IEEE), 95–102. doi: 10.1109/VR.2010.5444807

Zhao, M., Li, T., Abu Alsheikh, M., Tian, Y., Zhao, H., Torralba, A., et al. (2018a). "Through-wall human pose estimation using radio signals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 7356–7365. doi: 10.1109/CVPR.2018.00768

Zhao, M., Tian, Y., Zhao, H., Alsheikh, M. A., Li, T., Hristov, R., et al. (2018b). "RF-based 3D skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication* (Budapest), 267–281. doi: 10.1145/3230543.3230579

Zheng, E., Mai, J., Liu, Y., and Wang, Q. (2018). Forearm motion recognition with noncontact capacitive sensing. *Front. Neurorobot*. 12:47. doi: 10.3389/fnbot.2018.00047

Zopf, R., Harris, J. A., and Williams, M. A. (2011). The influence of body-ownership cues on tactile sensitivity. *Cogn. Neurosci*. 2, 147–154. doi: 10.1080/17588928.2011.578208

Conflict of Interest: Microsoft, Disney, Google, Didimo, and Virtual Bodyworks are either private or traded companies with interests in the topic discussed. MG-F, EO, AA, AS, and JL were employed by company Microsoft. SO-E was employed by company Google. YP was employed by company Disney. VO was employed by company Didimo. BS was employed by company Virtual BodyWorks.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.