Investigating User Embodiment of Inverse-Kinematic Avatars in Smartphone Augmented Reality

Elhassan Makled* Florian Weidner† Wolfgang Broll‡

Ilmenau University of Technology

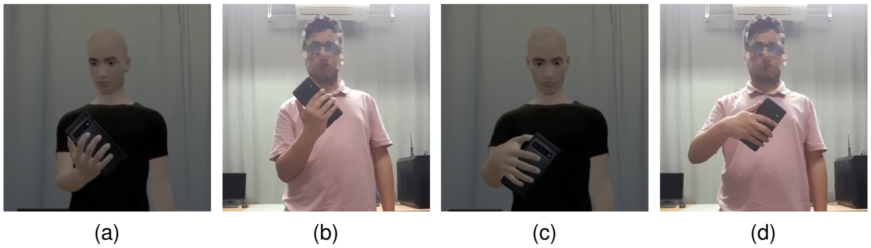


Figure 1: A user tilting the smartphone (right and center-left) and the IK avatar representation (left and center-right) through Augmented Reality IK Avatars (ARIKA).

ABSTRACT

Smartphone Augmented Reality (AR) has already provided us with a plethora of social applications such as Pokemon Go or Harry Potter Wizards Unite. However, to enable smartphone AR for social applications similar to VRChat or AltspaceVR, proper user tracking is necessary to accurately animate the avatars. In Virtual Reality (VR), avatar tracking is rather easy due to the availability of hand-tracking, controllers, and HMD whereas smartphone AR has only the back- (and front) camera and IMUs available for this task.

In this paper we propose ARIKA, a tracking solution for avatars in smartphone AR. ARIKA uses tracking information from ARCore to track the users hand position and to calculate a pose using Inverse Kinematics (IK). We compare the accuracy of our system against a commercial motion tracking system and compare both systems with respect to sense of agency, self-location, and body-ownership. For this, 20 participants observed their avatars in an augmented virtual mirror and executed a navigation and a pointing task.

Our results show that participants felt a higher sense of agency and self location when using the full body tracked avatar as opposed to IK avatars. Interestingly and in favor of ARIKA, there were no significant differences in body-ownership between our solution and the full-body tracked avatars. Thus, ARIKA and it's single-camera approach is valid solution for smartphone AR applications where body-ownership is essential.

Keywords: Augmented reality, inverse kinematics, embodiment

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking

1 INTRODUCTION

AR and VR technologies have been offering several possibilities for immersive experiences. They have impacted various industries and topics including cultural, entertainment, and educational [29, 31, 49]. By enabling users to occupy virtual spaces and to augment their real spaces with virtual objects, possibilities of remote collaborative or social applications are feasible. Further research explores collaborative and social environments and their impact on immersive experiences [5, 40, 43, 61, 62]. With Meta's Horizon Home and Horizon Worlds [34] announcement, public interest of consumers and designers alike peaked regarding the potential of social VR and AR experiences. This is not limited to VR to VR collaboration, but also asymmetric cross-platform collaborative experiences between AR and VR as explored through Grandi et al. [17].

Collaborative Virtual Environments (CVEs) are becoming more integral for remote collaboration between users. Several studies have explored the importance of user representation within these types of environments [41, 60, 61]. In these CVEs, users are commonly represented with avatars that represent their looks and behaviors within the virtual experience [23, 24, 47, 58].

Avatars have been commonly used in interactive experiences, such as video games, to represent users. Properly representing users in VR is an important factor in single and multi-user VR applications as it enhances user embodiment and sense of presence [58]. This is usually achieved by accurately representing the user's appearance and behavior using tracking devices such as controllers or sensors. In today's VR setups, a user's head typically is tracked with the Head-mounted display (HMD), and both of their hands are tracked via controllers [9, 10, 61]. This is done with the help of either an inside-out or outside-in tracking system. An inside-out system uses the headset itself to track the environment around it alongside the controllers [20]. As for outside-in tracking, it uses cameras or other sensors carefully stationed around the user's tracking volume. The tracking fidelity of those systems have the ability to enhances the user's sense of embodiment and presence within the experience [9]. Outside-in tracking is more common for VR applications while inside-out tracking is becoming widely used for both, VR and AR applications. AR is becoming more publicly and widely accessible through smartphones. Integrating the technology in smartphones has opened up possibilities for diverse applications to be accessible to a large target audience. Smartphone AR allows the user to view the new augmented world through the smartphone's screen. With the ubiquity of this technology, social and collaborative AR applications could help in enhancing different remote collaborative experiences and social interactions. However, a challenge in

---

*e-mail:elhassan.makled@tu-ilmenau.de

†e-mail:florian.weidner@tu-ilmenau.de

‡e-mail:wolfgang.broll@tu-ilmenau.de

these types of applications comes from the smartphone AR's lack of tracking data, other than the phone itself, to represent users in these social or collaborative experiences through avatars. To overcome this limitation, we present ARIKA, an IK avatar solution for smartphone AR Applications. ARIKA uses the phone's pose information delivered by ARCore [15] to position a 3D model of the phone in the augmented space. The 3D model of the phone is attached to the 3D model of the user — the avatar – and drives its movements. Thus, ARIKA animates the avatar through IK relative to the phone's pose information. We compare ARIKA with a professional grade motion capture system with respect to accuracy and self-embodiment in a study were 20 participants observed their avatars and performed a navigation and a pointing task. We present evidence that using IK avatars through ARIKA does not negatively impact body-ownership when compared with a motion capture system. We further enable researchers and designers to explore smartphone AR social applications using ARIKA.

The main contributions of this paper are:

- Present an accessible method for users to represent themselves in social or collaborative smartphone AR applications.
- Using only tracking information from the smartphone, infer user's position to drive their avatar's pose through IK providing a solution for avatar behavioral realism in smartphone AR.
- Maintain overall sense of body ownership when compared to professional grade full body tracking.

The remainder of this paper is structured as follows: InSect. 2 we outline related work about avatars, user tracking, and embodiment. Followed by Sect. 3 we discuss and further elaborates on ARIKA's implementation. In Sect. 4 and Sect. 5, we present our study design and procedure to evaluate ARIKA as well as our results respectively. In Sect. 6, Sect. 7, and Sect. 8 we contextualize our findings, discuss our limitations, and conclude our work respectively.

2 RELATED WORK

In this section, we discuss previous research that explored avatars and their impact on user's sense of self-embodiment in VR and AR applications as well as the different tracking techniques for user tracking. We further discuss the role of IK in avatar animations and user representation.

2.1 Avatars

In most VR applications, users occupy virtual spaces by being represented through avatars co-located with the users' bodies [9, 46, 50, 51, 61]. This allows users to have a sense of self-embodiment and presence in the virtual space. By that, the avatar acts as an interface between them and the virtual environment, allowing them to interact and engage with the virtual world [46]. Avatars allow users to be further immersed within experiences, making them more believable and realistic, improving elements such as cognitive abilities [53] and haptic performance [11, 32]. We further discuss embodiment in Sect. 2.4. Prior research explored the impact of avatar appearance on overall sense of embodiment and on user perception. Through comparing different render styles such as realistic or abstract render styles, Lurgin et al. [29] report a significantly higher embodiment for avatars that were realistic.

They further explore the impact of gendered avatars on embodiment. Schwind et al. [51] explores the effect of gender on avatar perception in VR environments. Their results suggest the importance of avatar diversity with respect to gender for embodiment and overall acceptance.

Further studies comparing partial body representation such as full body, torso, and head and hands for avatars in VR [13, 27, 50] suggest the importance of full body representation as opposed to partial.

Not only does realism in appearance impact user's self-embodiment, but also behavioral realism. That is the extent of to which avatars or virtual objects behave similar to their counterparts in the real world. To achieve this in avatars, body tracking is required to animate the avatar's movements with respect to the user.

2.2 Body Tracking (with limited sensors)

Tracking technologies are very diverse in their approach of tracking, specifically when it comes to tracking users. They can be divided into outside-in and inside-out tracking. In the following, we present an overview of tracking in general and for smartphone AR which has the constraint of limited sensor-availability.

2.2.1 Outside-In Tracking

Tracking systems such as Optitrack [44] or Vicon [56] use multiple cameras that are set up around a capture area, to track objects. These systems use marker-based tracking to identify different tracked objects as well as accurately manipulate their transforms in space. Their price range and complex setup nature make them not accessible for average consumers.

There are also various markerless tracking solutions in the market or publicly accessible ones such as OpenPose [4] or Kinect [36]. However, with them, occluded body parts are not registered and properly tracked. Furthermore, they restricts users to be facing the sensor at most times. While Optitrack uses marker-based tracking and OpenPose uses markerless tracking, they both successfully use the reconstructed skeleton information to puppeteer the user's avatar. Again, the need for inward facing cameras prevents such solutions for ubiquitous smartphone AR applications.

In the XR space, the most common tracking comes from commercial VR-HMDs such as the HTC VIVE [20] or Oculus Quest [35]. HTC VIVE uses trackers from the controllers and headset to track the user's hands and head respectively [9, 10]. Users can add more tracked points or objects by using HTC's VIVE Trackers [21]. In most use cases for smartphone AR, such methods are usually not applicable due to the absence of external tracking cameras.

2.2.2 Inside-Out Tracking

The opposite of this method is inside-out tracking, where the sensors tracking the device are placed on the device itself. Unlike outside-in tracking, inside-out tracking does not require and is not limited to a specific tracking area and setup, which makes it more accessible and less limiting. Inside-out method is common in various VR-HMDs such as the Oculus Quest, Windows Mixed Reality headsets [38].

Some inside-out tracking devices such as ARCore and Oculus Quest use Simultaneous Localization and Mapping (SLAM), a method for localization and mapping used in various applications from self-driving vehicles to AR applications, to localize and position the headset in the world. This helps in tracking the user's head however, it does not give information with regards to other parts of the user's body such as hands or legs. For Oculus Quest, HoloLens, and Windows Mixed Reality headsets, multiple outward-facing cameras on the headset not only track the position of the headset in space, but also used track either the user's hands or controllers. However, for smartphone AR, such additional cameras are not available — thus, hand-tracking is not possible.

As mentioned earlier in Sect. 2.1, avatar behavioral realism is an important aspect of user's self-embodiment. Furthermore, higher level of behavioral realism, through tracking, leads to better non-verbal behavior as well as higher levels of self presence as described by Herrera et al. [18]. Therefore, tracking only the three points —head and hands — does not fully represent a user's expression through their avatar. Thus, IK is used to fill in the gaps based on positional information about the hands and head of the user [9, 34, 57].

2.2.3 Tracking for Smartphone AR

For smartphone AR, tracking the user is a complex problem due to the unavailability of external cameras for outside-in tracking and because the internal cameras are usually not able to capture the hands.

Thus, various approaches have been presented that use external devices. Here, thermal cameras on the wrist, [22], EMG-sensing wearables [28], phone mounted wide view RGB camera [26], body-mounted fish-eye cameras [45] have been presented. Wu et al. have presented an alternative outside-in approach that uses the reflections and disturbances in WiFi-signals to reconstruct body pose [59]. While all these approaches can be used to reconstruct the users pose and by that, drive an avatar, they are hardly easy-to-use and all require either external sensors or additional devices. Ahuja et al. [1] to overcame this limitation by using only the smartphone. Their approach relies on the back- and front camera as well as the IMU of an iPhone. Using these sensors, they were able to reconstruct the pose and animate an avatar (see Sect. 2.3.2).

2.3 Inverse Kinematics

2.3.1 Inverse Kinematics in Avatars

IK is a method in animations of calculating the posture of joints based on their individual degree of freedom [3]. Therefore, IK can be used to estimate untracked users' joints [9], a method used in various commercial applications such as VRChat [57] or Meta's Horizon Worlds [34] to achieve behavioral realism in VR applications. Commercial SDK's like FinalIK [48] are accessible for developers and designers to use out of the box with game engines such as Unity [55] or Unreal [8] to further implement IK for avatars. Furthermore, commercial add-on trackers for the HTC Vive could be used to improves the accuracy of the IK solver by tracking more points on the user body such as hips, feet, and knees [9]. A study by Eubanks et al. [9] explored the impact of tracking fidelity on users' sense of embodiment and presence in VR. By increasing the tracked points on the user's body — and by that, the tracking fidelity — users felt more present and felt a more sense of self-embodiment. With most smartphone's quickly becoming AR capable, IK could be used to improve social AR applications and allow interactions similar to that of VR social applications.

2.3.2 IK in Smartphone AR

IK avatars for smartphone AR have been proposed for tracking users in a remote scenario by Murugan et al. [39]. In their paper they propose a study design to measure the effect of different avatar types on social presence during a collaborative experience using only positional data received from the device.

Ahuja et al. [1] proposes, Pose-On-The-Go, a solution using a fusion of sensors for iPhone [2] users. Their solution mixes an IK avatar with information from the IMU (inertial measurement unit) sensor in the phone as well as the front facing depth camera to more accurately estimate the user's pose. The depth camera is used to determine the user's torso orientation. Further, the IMU sensor's data is used to check if the user's lower body motion through detecting the user's steps. The IK is then used to reconstruct the avatar's bone position based on the estimated foot location and hand location. Their work compares the accuracy of bone estimation with Vicon, an external sensing motion capturing system. As hypothesized, Vicon was more accurate than Pose-On-The-Go however, applications that allow iPhone AR users to be represented in AR social and collaborative applications are now more accessible for through Pose-On-The-Go.

Even though Pose-On-The-Go successfully represented users through an IK avatar, they did not explore the impact of the avatars on users' self-embodiment in smartphone AR. Moreover, with the focus on iPhone as the platform, further limitations exist as most smartphone's do not include diverse sensors, such as the depth camera. Therefore, with ARIKA, we aim to explore the impact of IK avatars on user's self-embodiment through using only positional data acquired from an Android [14] device via ARCore.

2.4 Embodiment

Kilteni et al. [25] define sense of embodiment as a sense "towards a body B that emerges when B's properties is processed as if they were the properties of one's own biological body". They further explore properties pertaining to the human body to derive three main sub-components – sense of body, sense of agency, and sense of self-location.

Kilteni et al. [25] further describe self-location as the sense of being self-located inside one's body. This is not to be confused with the sense of presence, which is the sense of being there in a virtual world or environment, place illusion as described by Slater [52]. They further describe sense of body as the sense of body ownership over one's body. Finally, they describe the sense of agency as the sense of having motor control over one's body. Overall, sense of embodiment is a combination of those three main components.

2.4.1 Embodiment in AR

Unlike VR applications, smartphone AR applications users usually do not see their avatar or a representation of themselves within the experience. They only see their avatar if they consciously point the smartphone at their own body and look at it. Nimcharoen et al conducted a study where they used AR-HMDs to evaluate user embodiment [42], participants used a Microsoft HoloLens [37] and a Microsoft Kinect [36] gathered point cloud data of them. With the point cloud data, the system reconstructed a 3D representation and displayed it in front of the participants in a mirror-like scenario. It was concluded that users felt a sense of ownership towards their digital representation.

Through ARIKA we aim to investigate the impact of full body IK avatars on embodiment using only tracking information from ARCore's positional data. In Sect. 3 we discuss ARIKA's implementation.

3 ARIKA IMPLEMENTATION

Core objective of ARIKA is to track the user without any external sensors and with data available on common smartphones: the rear camera and the IMU. Similar to IK avatars with 3-point tracking in VR from Eubanks et al. [9], ARIKA uses the phone's position and orientation to track the user's right hand and therefore implement the IK from this point.

ARIKA system was built using Unity 2021.2 and tested on a Pixel 6 Pro [16] Android smartphone. We used Unity's ARFoundation version, Unity's AR cross platform SDK that supports Android's ARCore and iOS's ARKit.

3.1 ARIKA's Avatars

For avatar's 3D models we use Genesis 8 [7] from Daz Studio [6]. As previously mentioned in Sect. 2.4, diverse representation of avatars with respect to gender has improved user's body-ownership, a component in self-embodiment. Therefore, we provide participants with a choice to pick between a female or a male Genesis 8 avatar.

Alongside the avatar's appearance, the avatar's behavioral realism improves the user's sense of self-location and agency through tracking. For ARIKA we track the user only through the smartphone as we describe in Sect. 3.2.

3.2 Getting the phone's pose in world space

We use ARCore to get the device's orientation and position. As mentioned in Sect. 2.2.2, ARCore uses inside-out tracking through SLAM to position the phone in space. While the physical environment is being mapped, ARCore decides on a point to act as the world origin for the augmented world. This point is usually the

(a) Real user (b) ARIKA IK avatar

Figure 2: A user walking using ARIKA and the IK avatar representation from ARIKA as displayed to users.

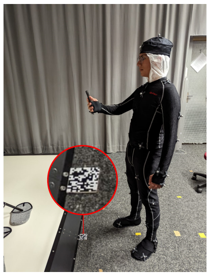


Figure 3: A participant using ARIKA during our studies while wearing the Optitrack suit with the marker placed on the floor.

(a) (b)

Figure 4: Real room a and digital twin b in a side to side comparison. Participants experienced a seamless AR with the smartphone acting as magic lens.

Table 1: Velocity values and their respective walking animations. Our system uses a left hand y-up coordinates system. Therefore, the positive z-axis is the user's forward while the positive x-axis is the user's right-side.

Velocity (x, z)	**Animation**
0.0, 0.0	Idle/standing
1.0, 0.0	Walk left sideways
-1.0, 0.0	Walk right sideways
0.0, 1.0	Walk forward
0.0, -1.0	Walk backward
0.5, 0.5	Blend (walk left sideways, walk forward)

initial position of the phone once mapping and tracking the physical environment, through SLAM, is successful. This means that the world origin for the augmented world is different with every session for ARIKA. Initially, this is not an issue for our solution. However, for the purpose of our study and to compare ARIKA with Optitrack, a marker is placed in a desired point in the real world to act as a reference to map the Android application coordinates to the Optitrack coordinates. This marker is scanned using ARIKA and its position and orientation act as our augmented world's origin point as well as Optitrack's. All our augmented objects are rendered relative to this origin point.

Once the origin is defined, the phone's position and orientation from ARCore are used to manipulate the transform of a 3D model of the phone in ARIKA. The model's position and orientation are then used to infer the wrist, elbow, and shoulder transforms of the avatar through IK. This only estimates the pose of the hand holding the phone as shown in Fig. 1.

Animating only the hand that is holding the phone could be acceptable for stationary scenarios where users will only move their hands. However, to represent a more active scenario, where the users move freely around in space, we need to track the user's location in space.

3.3 ARIKA's Avatar IK

We further develop components in ARIKA to support the purpose of our study. In Sect. 4, we discuss the study design, procedure, room setup for the study, as well as purpose-developed features.

4 STUDY DESIGN

Representing user movement across a scene is important for self-location and agency, which are key structures in achieving user embodiment for avatars. Unity's IK system does not provide out of the box implementation for avatar navigation, therefore an extra component was implemented to support avatar navigation around space: avatar destination controller. To achieve this, we create a virtual 3D point that acts as a point of destination to where the avatar should be standing from the phone. We update the point's position and orientation as the phone moves in space. To find the distance between the avatar and the phone, we explore Merbah et al. [33] study on postures of smartphone users. Through exploring various user postures while using the phone, Merbah et al. measured distances between the user's face and phone. Through a texting and browsing task, they found a minimum of 31.95cm and a maximum of 46.04cm. We use this distance range for the avatar destination point such that if the phone moved beyond the maximum expected distance, the avatar would move towards the phone to maintain the distance range as shown in Fig. 2. Similarly, if the phone moves closer to the user's face than the minimum distance, the avatar moves backwards.

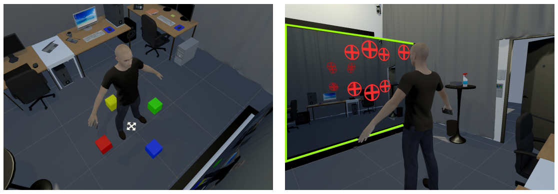
4.0.1 Walking Animations

As the avatar moves in space, it is expected to walk around the environment. This was done by calculating the 2D velocity of the phone during a specific time frame. A window of 10 frames is used to capture the phone's velocity during that time. After removing outliers and averaging the velocity, we pass the average velocity to the avatar's animator controller, a state machine for the walking animations that takes the velcoity as an input. Based on the velocity, it animates the IK avatar. We used a total of 5 predefined animations for the IK avatar. A blend tree blends between those animations based on the 2D values of the phone's velocity in the X and Z axes. Table 1 lists the velocity values in meters per second and the respective animations.

The purpose of our research is to investigate embodiment in IK avatars using only a smartphone for tracking a user and validate results by comparing them with a commercial professional motion tracking system. We had a within-subject design with two factors – type of task and type of tracking. Type of task has two levels: navigation task and pointing task. Type of tracking has also two levels: Optitrack and IK. In the former condition, we use the tracking of Optitrack's Motive:Body to track the avatar (ground-truth). In the IK condition, we use ARIKA's IK (c.f. Sect. 3).

Table 2: List of questions used in the embodiment questionnaire provided to our participants.

No.		Structure
Q1	Overall, I felt as if my body was located where I saw the virtual body to be.	Self-Location
Q2	Overall, I felt that the virtual body was my own body.	Ownership
Q3	The movements of the virtual body were caused by my movements.	Agency
Q4	It seemed as if I might have more than one body.	Ownership
Q5	Overall, I felt that the virtual body belonged to someone else.	Ownership
Q6	I felt like my body was actually there in the environment.	Self-Location
Q7	I felt like my body appeared in the environment.	Self-Location
Q8	I felt like my bodily movements occurred within the environment.	Agency
Q9	I felt like my body affected the environment.	Agency
Q10	I felt like the environment affected my body.	Ownership
MQ1	Overall, I felt as if the virtual body I saw when looking in the mirror was my own body.	Ownership
MQ2	Overall, I felt as if the virtual body I saw when looking in the mirror was another person.	Ownership



(a) Navigation points

(b) Grid points

Figure 5: The digital room with points, navigation and pointing, for the tasks, around the generic male avatar with the mirror with green highlight.

4.1 Procedure

We greeted the participants and briefed them on the experience and set their expectations regarding the tasks. After that, they signed a consent form and filled a demographics questionnaire. Then, the participants wore the Optitrack suit and we attached the Optitrack markers to it as shown in Fig. 3. For each participant, we then created a skeleton through Motive:Body. Before each session, we adjusted ARIKA's avatars' height, according to that of the participant. As every participant experienced both conditions — ARIKA and Optitrack — all participants wore the motion tracking suit all the times. To make the experiment hygienic, participants wore a painters suit underneath the motion tracking suit (white in Fig. 3).

During the study, an AR mirror (4m x 3m) displayed participant's avatars. This was decided to have the mirror aesthetically placed at the edge of the L-shape projector's setup as shown in Fig. 5. We created a digital replica of the room so that not only the user's avatar but also the laboratory were visible in the AR mirror. Thus, the smartphone acted as a magic lens, seamlessly blending AR content with the real world. Fig. 4 illustrates the real laboratory with its digital twin. With the suit, the AR mirror, the avatar, and the tracking systems set up, participants were ready to perform the tasks.

Each participant did a total of two tasks: a navigation task (Sect. 4.2.1) and a pointing task (Sect. 4.2.2). Participants did each task once with the avatar being tracked by Optitrack and once with our ARIKA system. We used a Latin Square to decide on the order of tasks per participant. By that, we control for order and learning effects.

After each task, we asked participants to fill in a questionnaire to evaluate their sense of embodiment; Thus, participants filled the embodiment questionnaire four times.

4.2 Tasks

We designed two tasks to evaluate ARIKA IK system (pointing and navigation). During all tasks, participants were only allowed to use the phone with their right hands.

4.2.1 Navigation Task

We designed a navigation task similar to the one used in Pose-on-the-go [1]. the navigation task consists of two sequences where participants had to do a series of steps for 20 seconds. First, we asked Participants to follow the following steps repeatedly for 20 seconds – Stand in the center of the room, then take a step to the right then, take a step back to the center of the room, then take a step to the left. After that, participants were asked to follow the same steps for moving forward and backward instead of right and left.

Fig. 5 illustrates the points participants had to reach during this task. Here, participants were forced to walk. This was done to evaluate the velocity-based animation system of AKIRA (c.f. Sect. 3).

4.2.2 Pointing Task

Fig. 5 illustrates the pointing task. For the pointing task, a grid of 8 augmented points was placed around a 3D space of 0.5m x 0.5m x 0.5m with the highest point placed at 1.75m from the ground. Only one random point from the grid appeared at a time. We asked the participants to search for that point and approach it. Once they were close enough to the point, they had to tap a button on the phone's screen. The point then disappeared and the next point appeared. The task was done once participants approached all points. With this task, we wanted to nudge participants into moving their hands more often to avoid a static posture of holding the phone in front of them.

4.3 Apparatus

Streaming the motion captured avatar from Optitrack requires Motive to be running on a PC with an Optitrack licence dongle attached. This meant that to stream the Optitrack avatar to the smartphone for the users, requires a network.

We use a windows PC running Optitrack's Motive as well as an ARIKA desktop instance in Unity that acts a server. The user's skeleton data from Optitrack is streamed to ARIKA's desktop instance through Motive. Then, an instance of ARIKA running on the smartphone device connects as a client to the desktop instance over local network. ARIKA's networking part was built using Netcode for GameObjects [54], Unity's network library.

To collect data from ARIKA's IK and Optitrack's full body avatar, both systems — Optitrack Motive:Body and ARIKA — recorded data, regardless of condition. However, depending on the condition, the avatar was only driven by one of the systems. It is important to mention that joint transform data for the IK and Optitrack avatars were recorded on the server instance as csv files with a Unix timestamp.

In order to map the augmented world and the digital replica of the room to the real world, a marker was placed in the center of the room as shown in Fig. 3. This marker was initially tracked to identify a point of origin to be used by the Optitrack system and ARIKA as discussed earlier in Sect. 3.

4.4 Independent Variables

We evaluate ARIKA's avatar pose against that of the Optitrack's. Our independent variables are tracking and task. We collect the avatar's joints transform data of the same joints collected by Ahuja et al. in [1] with Pose-on-the-go to evaluate their system against that of an optical tracking system. The transform data for each study was then compared between both avatars–IK and Optitrack–to calculate the euclidean distance and the rotational error.

4.5 Dependent Variables

We measured embodiment the questionnaire of Peck et al. [46]. As not all questions were relevant to our use case (e.g., question about tactile cues), we have selected the questions that would help us measure embodiment in our study and compiled them into a questionnaire (similar to Eubanks et al. [9] and Gonzalez et al [12]). This selection of questions allows us to make assumptions about body ownership, agency, and self-location — core components of embodiment. Our final set of questions can be found in Table 2.

4.6 Sample

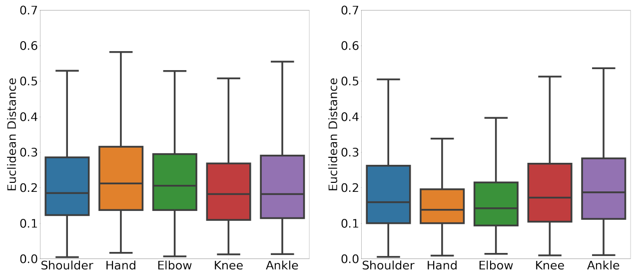
Our study was conducted with 20 participants. 16 participants self-identified as male and 4 self identified as female. The average age of all participants was 27 years (min = 19, max = 62, SD = 10.42). The average participant height was 176.5cm (min = 157cm, max = 193cm, SD = 9.84). Participants were also asked about their level of interest in Augmented Reality applications. On a Likert scale (1-5), the mean was 4.2 (SD = 0.82). Participants spent an average of 27 minutes (min = 12mins, max = 50mins, SD = 11mins) within the application. We needed approximately 20 minutes at the beginning for setting up Optitrack and the motion capture suit.

5 RESULTS

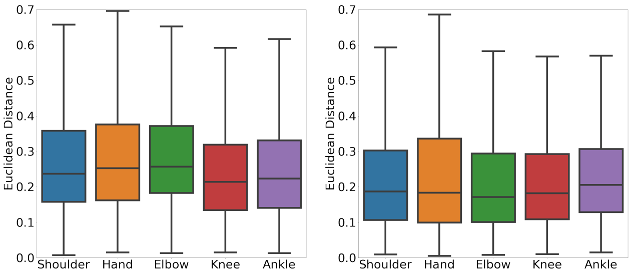
5.1 Euclidean distance and Rotational Error

We collected joint data of the IK and Optitrack avatars to evaluate the euclidean distance and the rotational error between them. Data was measured at a rate of mean(M) = 81(standard deviation(SD) = 22) frames per second.

For pre-processing, we first averaged the position and rotation of each joint per second and then calculated the euclidean distance and rotational difference for each joint between the IK and Optitrack avatars for each window of 1 second length. We used a z-score approach to identify and remove outliers that lie more than 3 standard deviations away from the mean from our calculated euclidean distance and rotational differences. Similar to Pose-on-the-go [1], we explore only a subset of joints from the avatar's skeleton: hand, elbow, shoulder, hip, knee, and ankle. We do this for the left and right side. Thus, we analyze 11 joints.



(a) Navigation task - Left side (b) Navigation task - Right side



(c) Pointing Task - Left side (d) Pointing Task - Right side

Figure 6: Euclidean distance in meters from the navigation task (a, b) and pointing task (c, d) for the left and right side respectively.

Fig. 6 illustrates the euclidean distance error in meters between Optitrack and ARIKA for the joints under observation, split by tasks.

The overall euclidean distance error is M = 22cm (SD = 13cm) across the 11 joints. The euclidean distance for right and left joints across all tasks is M = 21cm (SD = 13cm) and M = 24cm (SD = 14cm) respectively.

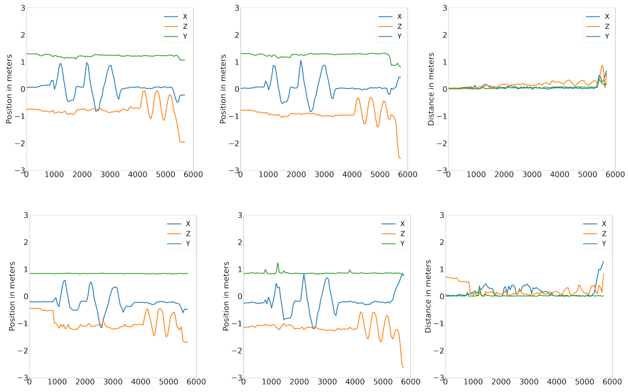
For the pointing task, the overall euclidean distance error was M = 24cm (SD = 14cm) for all observed joints, M = 22cm (SD = 14cm) for the right side, and M = 26cm(SD = 14cm) for the left side. For the navigation task, the overall euclidean distance error was M = 20cm (SD = 12cm) whereas the right side had an error of M = 19cm (SD = 11cm). For the left side, we measured an error of M = 22cm (SD = 12cm) between Optitrack and ARIKA.

We further investigate the observed joints individually across all tasks. As expected, the joint with the least overall euclidean distance error was the elbow, followed by the hip, then the right wrist, then the shoulders. Interestingly after that, the lower right side of the body followed, right knee, then right ankle. After that the left side of the body starting with knee, ankle, shoulders, elbow, then finally the hand. A detailed breakdown of the means and standard deviation is found in Table 3.

Table 3: Mean euclidean distance error per observed joint with standard deviation in brackets.

Joint	Euclidean distance error per joint Right	Left
Elbow	19.3cm (12.6cm)	25.9cm (13.4cm)
Hand	20.2cm (14.3cm)	26.7cm (15.2cm)
Shoulder	20.5cm (13.2cm)	24.2cm (13.4cm)
Knee	20.6cm (12.4cm)	22.2cm (12.6cm)
Ankle	22.2cm (12.9cm)	23.2cm (13.5cm)
Hip	20.1cm (11.9cm, no left/right joint)	

We further investigate data from a single user during a navigation and pointing task. The average difference between Optitrack and IK is different between the right side of the body and the left side of the body as shown in Fig. 7 and Fig. 8. The euclidean distance on the right side has no changes across time. This is due to the phone being held by the user's right hand. The transform data we acquired from the phone via ARCore for the IK user. The spike in differences in the first and last seconds of the data are due to the phone not initially being with the participant and finally being handed over by the participant at the end of the session.

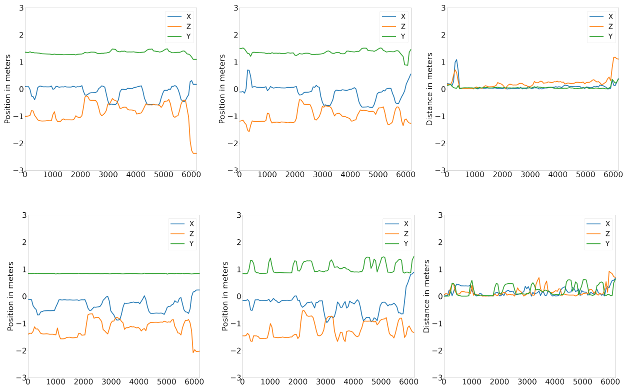


(a) IK Data (b) Optitrack Data (c) Absolute difference

Figure 7: Transform data of the right (top row) and left (bottom row) wrists transformations from a single user during the navigation task, while observing their Optitrack avatar, and the absolute difference between the IK and Optitrack avatar transforms.

Table 4: Mean rotational error per cardinal axis with standard deviation in brackets.

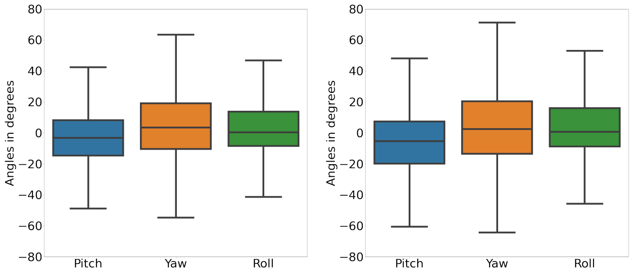
Group	Rotational error per axis Yaw	Pitch	Roll
Nav-Right	5.2° (27.6°)	0.3° (62.4°)	21.6° (56.1°)
Nav-Left	4.8° (29.3°)	-5.7° (42.4°)	-3.8° (34.1°)
Point-Right	7.3° (28°)	0.24° (60°)	24.7° (52.1°)
Point-Left	0.04° (31.1°)	-12.3° (44.5°)	-8.9° (38.6°)



(a) IK Data (b) Optitrack Data (c) Absolute difference

Figure 8: Transform data of the right (top row) and left (bottom row) wrists transformations from a single user during the pointing task, while observing their Optitrack avatar, and the absolute difference between the IK and Optitrack avatar transforms.

Next, we report the rotational error between Optitrack and IK avatars. In Fig. 9 we show the combined error of all axis for the left and right side joints whereas Table 4 shows a breakdown of right and left side according to task.



(a) Navigation Task (b) Pointing Task side

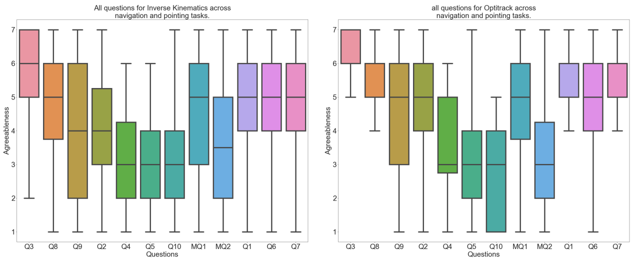
Figure 9: Rotational error in degrees between IK and Optitrack avatars' joints with across the navigation and pointing tasks.

5.2 Evaluating Embodiment Questionnaire

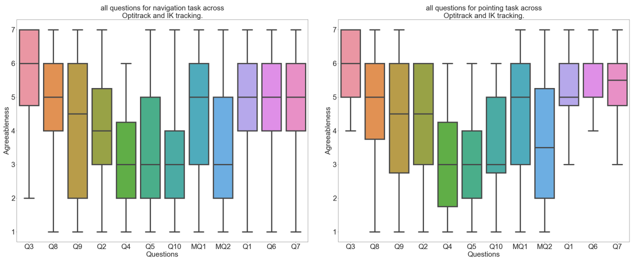
As mentioned in Sect. 2.4, embodiment is a combination of three different structures — self-location, agency, and body-ownership. Overall, questions that explored body ownership had similar mean scores and standard deviation for both Optitrack and IK. With M = 4.31(SD = 0.95) for Optitrack and M = 4.28(SD = 0.92) for IK. For agency, Optitrack had a higher mean of M = 5.12(SD = 1.3) for Optitrack thank IK M = 4.73(SD = 1.46). This is expected, as Optitrack has a higher tracking fidelity than IK. Participants rated self-location with M = 5.17(SD = 1.14) for Optitrack and M = 4.86(SD = 1.40) for IK. A more detailed statistical breakdown per question can be found in Table 5

Table 5: Mean score for individual questions from the embodiment questionnaire with standard deviation in brackets.

No.	Structure	IK	OT
Q1	Self-Location	4.78 (1.56)	5.30(1.29)
Q2	Ownership	4.20 (1.47)	4.50 (1.57)
Q3	Agency	5.50 (1.59)	5.88 (1.22)
Q4	Ownership	3.30 (1.52)	3.40 (1.58)
Q5	Ownership	3.08 (1.44)	3.08 (1.65)
Q6	Self-Location	4.88 (1.52)	5.13 (1.26)
Q7	Self-Location	4.93 (1.64)	5.08 (1.25)
Q8	Agency	4.68 (1.69)	5.25 (1.46)
Q9	Agency	4.03 (2.00)	4.23 (1.89)
Q10	Ownership	3.13 (1.64)	2.68 (1.46)
MQ1	Ownership	4.38 (1.72)	4.50 (1.77)
MQ2	Ownership	3.63 (1.90)	3.33 (1.75)



(a) IK data (b) Optitrack



(c) Navigation Task (d) Pointing Task

Figure 10: Questionnaire results for IK and Optitrack tracking (a, b) and navigation and pointing tasks (c, d) respectively.

We performed a repeated-measures ANOVA to compare the effect of tracking on each question from the questionnaire. There was no statistically significant difference between IK and Optitrack in questions that explored body ownership ($p > .2$).

There was a statistically significant difference between IK and Optitrack in the following questions: Q1, Q3, Q8, and Q9 in Table 2. We found a significant effect of tracking on Q1 (self-location, $F(1, 19) = 4.412$, p=0.049, η_p^2=0.188) We further found significant effect of tracking on Q3 (agency, $F(1, 19) = 4.412$, p=0.036, η_p^2=0.212), Q8 (agency, $F(1, 19) = 4.412$, p=0.039, η_p^2=0.205), and Q9 (agency, $F(1, 19) = 4.412$, p=0.006, η_p^2=0.337).

6 DISCUSSION

6.1 No significant difference in body ownership between Optitrack and IK avatars

Our findings suggest no significant difference in body ownership between IK and Optitrack avatars for smartphone AR. Furthermore, body ownership scores for Optitrack and IK avatars are relatively neutral as displayed in Table 5 and Fig. 10. This could be due to the difference in appearance between the avatar and the participants. Some participants have suggested that they couldn't identify with the avatar as it didn't represent their age, looks, or skin color. Moreover, unlike VR users, AR users are not in complete isolation of the real world. Unlike the study from Nimchareon et al. [42], our participants did not observe a 3D reconstruction of their physical bodies. Therefore, seeing their real body and the virtual body at the same time could have negatively impacted immersion and therefore body ownership.

6.2 Full body tracking is better for self-location and agency in avatars

Significant differences between tracking methods were found in the following questions, Q3, Q8, and Q9. Question Q1 explore self-location within virtual environments. For the IK tracking, scores were relatively similar across tasks. However, there was a significant difference between the Optitrack and the IK avatars in terms of self-location. This is mostly due to ARIKA not having tracking information of parts of the body such as the torso, hips, or legs. Because of this, ARIKA predicts that a step is taken based on the velocity and displacement of the phone. Thus, a small step wouldn't be visualized through ARIKA's IK avatar making it appear in a different location than the user's physical body. Therefore, impacting user's self-location through the avatar. This tracking limitation in IK avatars further caused the significant differences in Q3, Q8, and Q9 — the questions that explore agency. The IK avatar of the participants that moved their left hand, the hand not holding the phone, did not mimic their movement. This impacts the user's sense of agency since, as mentioned in Sect. 2.4, sense of agency is improved when user's seem to have motor control over their avatars. This disparity in user-avatar movement could have also been uncanny for the users. Future studies should include a questionnaire investigating this aspect [19].

Our findings suggest significant difference in overall sense of agency between IK and Optitrack avatars for smartphone AR. Agency scored relatively highest compared to self-location and body-ownership. This means that users felt more in control over their avatars when they were being fully tracked as opposed to using IK. It further implies that users felt as if they are located within their avatars with full body tracking more than with IK. As expected, IK scored less than Optitrack in embodiment. Although, there is room for improvement for ARIKA, the median values for agency and self-location were above the neutral values. Moreover, overall body-ownership was not sacrificed when users were tracked through ARIKA. Furthermore, this was achieved only through accessible AR smartphone technology without the need of external sensors or sensor fusions. Therefore, we believe ARIKA could be used to represent users in smartphone AR applications.

6.3 Comparing ARIKA IK to Optitrack

Ahuja et al. [1] compared Pose-on-the-go with Vicon, an optical tracking system similar to Optitrack. They report an overall euclidean distance across all bones M=18cm (SD=3.0cm). They further report individual joint differences with wrists (M=27.4cm, SD=4.7cm), elbows (M=17.0cm, SD=3.9cm), and shoulders (M=9.7cm, SD=2.1cm). In comparison to Pose-on-the-go [1], ARIKA has an overall larger mean euclidean distance error and standard deviation with a M = 23cm, (SD = 14cm). The difference in the overall mean (delta = 5cm) could be due to the fact that unlike Pose-on-the-go [1], ARIKA only uses ARCore's phone transform data as opposed to a fusion of sensors to further track other parts of the body such as torso, head, and left hand screen interactions. ARIKA is successful in estimating an avatar pose using only data from ARCore without sacrificing euclidean distance error.

7 LIMITATIONS AND FUTURE WORK

It is important to emphasise that ARIKA is a pose estimation solution for smartphone AR applications and is not intended to replace tracking devices such as Optitrack. ARIKA is intended to be a step towards collaborative and social smartphone AR applications through using IK avatars for user representation. There is room for improving ARIKA's features to reach the intended purpose and explore more multi-user scenarios such as presented by Makita et, al [30].

The first focus should be a solution towards user-avatar likeness. Various participants pointed out the desire for their avatar appearance to match their appearance. Including an avatar editor for users could help them better identify with their avatars. This will further impact user embodiment in smartphone AR, specifically body-ownership.

Although ARIKA offered a blend between various animations to accurately estimate user movement, it still requires some fine tuning and more diverse animations. ARIKA's IK avatar will animate with a step regardless if the users are taking smaller or bigger steps. Furthermore, ARIKA only supports walking animations, which is a fraction of the different actions users could take while using the application. Thus, a more diverse and detailed set of animations, such leaning or walking with smaller steps, will enhance overall avatar representation of user movement. Since the current state of ARIKA only allows for minor movement, specifically for navigation and space exploration. Future work should focus towards a more complex system that allows for full body expression by the user and more complex physical coordination.

Similar to Pose-on-the-go [1], the use of sensor fusion, specifically the front facing camera, will create better pose estimation for users. However, this will come with a trade-off for accessibility.

During our study, we had a limitation in diversity of our participants, with 16 participants identifying as males and 4 identifying as females. It is important to further conduct similar studies with a more diverse group to identify differences in results [51].

Finally, the currently available embodiment questionnaires focus on VR and AR applications, specifically looking at a virtual environment through an HMD with a first person perspective. Therefore, further investigation towards a standardized questionnaire that takes into consideration smartphone AR to measure overall embodiment would help future studies in this field have a more accurate and confident results with respect to overall user-embodiment.

8 CONCLUSION

Various Smartphone AR applications have explored social and collaborative AR scenarios. However, a gap exists in user tracking to fully enable social and collaborative experiences in smartphone AR. We presented ARIKA, a tracking solution for smartphone AR using IK avatars to represent user poses and relying on only a single camera. Through ARIKA we explore the impact of IK avatars on user embodiment. We further compare it to a professional motion capturing system, Optitrack. Our results suggest that users feel a higher sense of agency and self location with fully tracked avatars than with IK avatars. Still, there are no significant differences in body-ownership between the IK and the full-body tracked avatars, despite an average euclidean distance error of 23 cm. This indicates that our approach that just relies on a rear-camera and IMU data is a viable approach for avatar tracking in AR when it comes to body-ownership. Thus, ARIKA is a step towards democratizing smartphone AR as it uses only the rear-facing camera and IK for tracking. As many vendors and development kits (like ARCore) prevent parallel access to both cameras, our solution provides an efficient way to drive avatar poses without sacrificing body-ownership compared to a high-fidelity, marker-based tracking system. By that, ARIKA fosters the dissemination of social AR applications and can act as a building block of the future metaverse.

ACKNOWLEDGMENTS

Our research received funding from the Thuringian Ministry for Economic Affairs, Science, and Digital Society under grant 5575/10-5 (MetaReal).

REFERENCES

[1] K. Ahuja, S. Mayer, M. Goel, and C. Harrison. Pose-on-the-go: Approximating user pose with smartphone sensor fusion and inverse kinematics. Association for Computing Machinery, 5 2021. doi: 10.1145/3411764.3445582

[2] Apple Inc. iPhone - Apple, 2022. `https://www.apple.com/de/iphone/`.

[3] A. Aristidou and J. Lasenby. Fabrik: A fast, iterative solver for the inverse kinematics problem. *Graphical Models*, 73:243–260, 9 2011. doi: 10.1016/j.gmod.2011.05.003

[4] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[5] D. Datcu, S. G. Lukosch, and H. K. Lukosch. Handheld augmented reality for distributed collaborative crime scene investigation. vol. 13-16-November-2016, pp. 267–276. Association for Computing Machinery, 11 2016. doi: 10.1145/2957276.2957302

[6] Daz Productions, Inc. Daz 3D - 3D Models and 3D, 2022. `https://www.daz3d.com`.

[7] Daz Productions, Inc. Genesis 8 Starter Essentials, 2022. `https://www.daz3d.com/genesis-8-starter-essentials`.

[8] Epic Games. The most powerful real-time 3d creation tool - unreal engine, 2022. `https://www.unrealengine.com/en-US/`.

[9] J. C. Eubanks, A. G. Moore, P. A. Fishwick, and R. P. McMahan. The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants. pp. 54–63. Institute of Electrical and Electronics Engineers Inc., 11 2020. doi: 10.1109/ISMAR50242.2020.00025

[10] M. Fly. Simulating avatar self-embodiment using 3-points of tracking.; simulating avatar self-embodiment using 3-points of tracking.

[11] M. Gonzalez-Franco and C. C. Berger. Avatar embodiment enhances haptic confidence on the out-of-body touch illusion. *IEEE Transactions on Haptics*, 12:319–326, 7 2019. doi: 10.1109/TOH.2019.2925038

[12] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. vol. 5. Frontiers Media S.A., 2018. doi: 10.3389/frobt.2018.00074

[13] G. Gonçalves, M. Melo, L. Barbosa, J. Vasconcelos-Raposo, and M. Bessa. Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive vr. *Virtual Reality*, 26, 3 2022. doi: 10.1007/s10055-021-00530-5

[14] Google LLC. Android — The platform pushing what's possible, 2022. `https://www.android.com/`.

[15] Google LLC. Google AR & VR, 2022. `https://arvr.google.com/arcore/`.

[16] Google LLC. Google Pixel 6, 2022. `https://store.google.com/product/pixel_6`.

[17] J. G. Grandi, H. G. Debarba, and A. Maciel. *Characterizing Asymmetric Collaborative Interactions in Virtual andAugmented Realities*, vol. 978-1-7281-1377-7/19. 2019.

[18] F. Herrera, S. Y. Oh, and J. N. Bailenson. Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 27:163–182, 2020. doi: 10.1162/PRES_a_00324

[19] C.-C. Ho and K. F. MacDorman. Measuring the Uncanny Valley Effect. *International Journal of Social Robotics*, 9(1):129–139, Jan. 2017. doi: 10.1007/s12369-016-0380-9

[20] HTC. HTC VIVE, 2022. `https://www.vive.com/us/`.

[21] HTC. HTC Vive Tracker, 2022. `https://www.vive.com/us/accessory/tracker3/`.

[22] F. Hu, P. He, S. Xu, Y. Li, and C. Zhang. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes Captured by Miniature Thermal Cameras on Wrist. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):71:1–71:24, June 2020. doi: 10.1145/3397306

[23] D. Jo, K. Kim, G. F. Welch, W. Jeon, Y. Kim, K. H. Kim, and G. J. Kim. The impact of avatar-owner visual similarity on body ownership in immersive virtual reality. vol. Part F131944. Association for Computing Machinery, 11 2017. doi: 10.1145/3139131.3141214

[24] K. Kilteni, I. Bergstrom, and M. Slater. Drumming in immersive virtual reality: The body shapes the way we play. *IEEE Transactions on Visualization and Computer Graphics*, 19:597–605, 4 2013. doi: 10.1109/TVCG.2013.29

[25] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence*, 21(4):373–387, 2012. doi: 10.1162/PRES_a_00124

[26] D. Kim, K. Park, and G. Lee. Oddeyecam: A sensing technique for body-centric peephole interaction using wfov rgb and nfov depth cameras. pp. 85–97. Association for Computing Machinery, Inc, 10 2020. doi: 10.1145/3379337.3415889

[27] M. Kocur, S. Graf, and V. Schwind. The impact of missing fingers in virtual reality. Association for Computing Machinery, 11 2020. doi: 10.1145/3385956.3418973

[28] Y. Liu, S. Zhang, and M. Gowda. NeuroPose: 3D Hand Pose Tracking using EMG Wearables. In *Proceedings of the Web Conference 2021*, WWW '21, pp. 1471–1482. Association for Computing Machinery, New York, NY, USA, Apr. 2021. doi: 10.1145/3442381.3449890

[29] J. L. Lugrin, M. Landeck, and M. E. Latoschik. Avatar embodiment realism and virtual fitness training. pp. 225–226. Institute of Electrical and Electronics Engineers Inc., 8 2015. doi: 10.1109/VR.2015.7223377

[30] K. Makita, M. Kanbara, and N. Yokoya. View management of annotations for wearable augmented reality. pp. 982–985, 2009. doi: 10.1109/ICME.2009.5202661

[31] E. Makled, A. Yassien, P. Elagroudy, M. Magdy, S. Abdennadher, and N. Hamdi. Pathogenius vr: Vr medical training. 2019. doi: 10.1145/3321335.3329694

[32] A. Maselli, K. Kilteni, J. López-Moliner, and M. Slater. The sense of body ownership relaxes temporal constraints for multisensory integration. *Scientific Reports*, 6, 8 2016. doi: 10.1038/srep30628

[33] J. Merbah, P. Gorce, and J. Jacquier-Bret. Interaction with a smartphone under different task and environmental conditions: Emergence of users' postural strategies. *International Journal of Industrial Ergonomics*, 77, 5 2020. doi: 10.1016/j.ergon.2020.102956

[34] Meta. Connect 2021 Recap: Horizon Home, the Future of Work, Presence Platform, and More, 2022. `https://www.oculus.com/blog/connect-2021-recap-horizon-home-the-future-of-work-presence-platform-and-more/`.

[35] Meta. Oculus Quest, 2022. `https://www.oculus.com/experiences/quest/`.

[36] Microsoft. Microsoft HoloLens Mixed Reality Technology for Business, 2022. `https://docs.microsoft.com/en-us/archive/msdn-magazine/2012/september/kinect-working-with-kinect-studio`.

[37] Microsoft. Microsoft HoloLens Mixed Reality Technology for Business, 2022. `https://www.microsoft.com/en-us/hololens`.

[38] Microsoft. Microsoft Virtual Reality, 2022. `https://www.microsoft.com/en-us/mixed-reality/windows-mixed-reality`.

[39] A. Murugan, R. Vanukuru, and J. Pillai. Towards avatars for remote communication using mobile augmented reality. pp. 135–139. Institute of Electrical and Electronics Engineers Inc., 3 2021. doi: 10.1109/VRW52623.2021.00032

[40] J. Müller, J. Zagermann, J. Wieland, U. Pfeil, and H. Reiterer. A qualitative comparison between augmented and virtual reality collaboration with handheld devices. pp. 399–410. Association for Computing Machinery, 9 2019. doi: 10.1145/3340764.3340773

[41] N. Nasser, E. Makled, N. Sharaf, and S. Abdennadher. Social interaction in virtual shopping. 2021. doi: 10.1109/ISM52913.2021.00040

[42] C. Nimcharoen, S. Zollmann, J. Collins, and H. Regenbrecht. Is that me?-embodiment and body perception with an augmented reality mirror, 2018.

[43] P. A. Olin, A. M. Issa, T. Feuchtner, and K. Grønbæk. Designing for heterogeneous cross-device collaboration and social interaction in virtual reality. pp. 112–127. Association for Computing Machinery, 12 2020. doi: 10.1145/3441000.3441070

[44] Optitrack. Optitrack Motion Capture System, 2022. `https://www.optitrack.com/`.

[45] K. Park, S. Kim, Y. Yoon, T.-K. Kim, and G. Lee. DeepFish-

eye: Near-Surface Multi-Finger Tracking Technology Using Fisheye Camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 1132–1146. Association for Computing Machinery, New York, NY, USA, Oct. 2020. https://doi.org/10.1145/3379337.3415818.

[46] T. C. Peck and M. Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1, 2 2021. doi: 10.3389/frvir.2020.575943

[47] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. vol. 2018-April. Association for Computing Machinery, 4 2018. doi: 10.1145/3173574.3173620

[48] RootMotion. Final ik — animation tool — unity asset store, 2022.

[49] R. Rzayev, G. Karaman, K. Wolf, N. Henze, and V. Schwind. The effect of presence and appearance of guides in virtual reality exhibitions. pp. 11–20. Association for Computing Machinery, 9 2019. doi: 10.1145/3340764.3340802

[50] V. Schwind, P. Knierim, L. Chuang, and N. Henze. "where's pinky?": The effects of a reduced number of fingers in virtual reality. pp. 507–515. Association for Computing Machinery, Inc, 10 2017. doi: 10.1145/3116595.3116596

[51] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. "these are not my hands!": Effect of gender on the perception of avatar hands in virtual reality. vol. 2017-May, pp. 1577–1582. Association for Computing Machinery, 5 2017. doi: 10.1145/3025453.3025602

[52] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3549–3557, 12 2009. doi: 10.1098/rstb.2009.0138

[53] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. vol. 2016-July, pp. 67–76. IEEE Computer Society, 7 2016. doi: 10.1109/VR.2016.7504689

[54] Unity. About Netcode for GameObjects, 2022. `https://docs-multiplayer.unity3d.com/netcode/current/about/index.html`.

[55] Unity3D. Unity documentation, 2022. `https://docs.unity3d.com/Manual/index.html`.

[56] Vicon. Vicon Award Winning Motion Capture Systems, 2022. `https://www.vicon.com/`.

[57] VRChat Inc. Vrchat, 2022. `https://hello.vrchat.com/`.

[58] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24:1643–1652, 4 2018. doi: 10.1109/TVCG.2018.2794629

[59] E. Wu, Y. Yuan, H.-S. Yeo, A. Quigley, H. Koike, and K. M. Kitani. Back-Hand-Pose: 3D Hand Pose Estimation for a Wrist-worn Camera via Dorsum Deformation Network. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pp. 1147–1160. Association for Computing Machinery, New York, NY, USA, Oct. 2020. doi: 10.1145/3379337.3415897

[60] A. Yassien, P. Elagroudy, E. Makled, and S. Abdennadher. A design space for social presence in vr. 2020. doi: 10.1145/3419249.3420112

[61] A. Yassien, E. Makled, P. Elagroudy, N. Sadek, and S. Abdennadher. Give-me-a-hand: The efect of partner's gender on collaboration qality in virtual reality. 2021. doi: 10.1145/3411763.3451601

[62] B. Yoon, H. il Kim, G. A. Lee, M. Billinghurst, and W. Woo. *The Effect of Avatar Appearance on Social Presencein an Augmented Reality Remote Collaboration*. 2019.