International Journal of Advanced Virtual Reality

Volume: 2, Issue: 1, Page: 56-68, Year:2025
http://journals.leukolion.com

3-point Tracking with Procedural and Baked Animation and Inverse Kinematic to Create a Closer Realistic Full-body Movement in Virtual Reality

Raymond Leonardo CHANDRA[1]

[1] RWTH Aachen University, Germany, chandra@lfi.rwth-aachen.de, ORCID: 0000-0002-9184-5062

Article History:
Received: 11 Jan 2024
Revised: 03 May 2024
Accepted: 31 Dec 2025
Online: 31 Dec 2025

Keywords:
Virtual Reality
3-Point Tracking
Baked Animation
Procedural Animation
Inverse Knimatics

Abstract
This paper presents a full-body avatar animation that combines 3-point tracking, inverse kinematics, and procedural and baked animation to improve avatar realism in virtual reality (VR). The technique allows users to control a full-body avatar using only a VR headset and controllers without the need for further sensors. The scalability of the proposed system in terms of multiple animated avatars being in one scene has been evaluated. A study with 22 participants was conducted in Aachen, Germany to evaluate the level of immersion provided by the combined animation avatar. Voluntary sampling method was used to recruit the participants and the quantitative experiment method was used as methodology approach. All four hypotheses tested were supported. The first hypothesis was evaluated using the Chi-Squared test, while the remaining hypotheses were assessed through descriptive statistics and the Wilcoxon Signed-Rank test. The results showed that the combined animation technique improved immersion and increased users' sense of presence and engagement in the VR environment. This study also found that the combined animation technique adequately reproduced the user's movements. Thus, the full-body avatar system presented in this work has the potential to enhance the immersive experience of VR applications.

Contents

How to Cite

Chandra, R., L. (2025). 3-point Tracking with Procedural and Baked Animation and Inverse Kinematic to Create a Closer Realistic Full-body Movement in Virtual Reality. *International Journal of Advanced Virtual Reality*, 2(1), 56-68.

1. Introduction

Virtual reality (VR) technology has evolved over the past decade from heavy and expensive prototype systems into mobile, affordable consumer products that offer highly immersive user experiences (Berni and Borgianni, 2020). This growth and technological advancement have been driven by several key factors, including the development of high-quality display technologies, accurate motion tracking systems, increased computational power, and improved wireless connectivity (Hamilton, McKechnie, Edgerton, and Wilson, 2021). As a result, VR has become increasingly popular as its potential for training, education, entertainment, and various other application domains has become more widely recognized (Al-Ansi, Jaboob, Garad, and Al-Ansi, 2023).

Most contemporary consumer VR systems rely on three-point tracking, which estimates the position and orientation of the user's head and both hands through the VR headset and handheld controllers, respectively (Thampan, Razak, and Krishna, 2023). Although full-body tracking can be achieved by augmenting three-point tracking with external sensors, this approach requires users to wear and calibrate additional trackers on body parts such as the hips and legs. To avoid these setup complexities, many VR applications instead employ floating avatars or partial avatars that omit the rendering of limbs, particularly the legs (Yang, Chen, Qin, Lam, and Landay, 2022).

The closest approximation to a full-body avatar without additional hardware is achieved by combining three-point tracking with pre-recorded leg animations. These pre-recorded animations, commonly referred to as baked animations, consist of motion sequences generated using traditional animation techniques such as motion capture or keyframe animation to replicate realistic human movement (van Gumster, 2020). The resulting animation data are precomputed and stored as static files containing positional and rotational information for playback. During runtime, these baked animations are stitched together to animate the lower body, while three-point tracking data are used to control the upper body.

However, this approach presents notable limitations. Because baked animations are static and predefined, they frequently fail to align accurately with the user's real-time movements. Furthermore, avatar interactions that were not anticipated during animation authoring cannot be performed. For instance, when a user stands on an inclined surface, the avatar's feet may appear to float above the ground rather than adapting naturally to the terrain. Such inconsistencies can significantly reduce movement realism and disrupt user immersion, as illustrated in Fig. 1.

In addition, although some VR applications use a full-body avatar with baked animation, the user controlling that avatar sees only floating hands, with no existing body part rendered in the user's view like shown in Fig. 2. This is potentially distracting and can break the user's immersion.

Figure 1. Unrealistic foot placement when walking on uneven surface

Figure 2. The users able to see their virtual hands, but not their full-body avatar (spatial.io)

To address these limitations, we developed an open-source virtual reality (VR) application named *MyScore* at the Teaching and Research Field of Engineering Hydrology, RWTH Aachen University, Germany. The primary objective of this study is to present a full-body avatar system implemented in *MyScore* that enhances user immersion in VR environments. The proposed system integrates the advantages of three-point tracking and baked animation with procedural animation and inverse kinematics (IK) through a mathematical approach, resulting in more realistic and adaptive avatar motion.

A key advantage of the proposed system is that it relies exclusively on tracking data obtained from the VR headset and handheld controllers, eliminating the need for additional hip or leg sensors. Moreover, users are able to perceive their avatar as a complete body rather than as disembodied or floating hands, which contributes to a more coherent embodiment experience. The technical details of this approach are discussed comprehensively in Chapter 3.

In addition to the system development, an empirical study was conducted to evaluate whether the proposed improvements in avatar animation lead to a measurable enhancement in user immersion. Based on the conceptual framework and the underlying research objectives, four hypotheses were formulated:

- **H1**: The use of combined avatar animation techniques (procedural animation, inverse kinematics, and three-point tracking) improves immersion in general VR scenarios more than baked animation alone.
- **H2**: The inclusion of arms and lower body parts in avatar representation enhances user immersion in VR.
- **H3**: Avatar motion that closely correlates with user motion increases the user's sense of immersion.
- **H4**: Realistic avatar motion is positively perceived by other users within a shared virtual environment (VE).

The results of the empirical study and the evaluation of the proposed hypotheses are presented and discussed in Chapter 4.

2. Literature Review

This chapter begins by defining essential concepts such as three-point tracking, procedural and baked animation, and inverse kinematics (IK), which constitute the fundamental building blocks for the development of a full-body avatar in virtual reality (VR). Subsequently, the definition of VR and the theoretical framework guiding this study are presented. This framework explains how the integration of different animation techniques can enhance user immersion and overall experience in VR environments. Finally, relevant literature is reviewed to position this work within existing research and to highlight its contribution to advancing avatar realism in VR.

2.1 Key Concept Definitions

This section revisits the key terms and central animation concepts relevant to this study.

2.1.1 Three-Point Tracking

The term *three-point tracking* refers to a VR tracking approach that monitors the position and orientation of the user's head and both hands using a VR headset and handheld controllers (Żuk, Wojtków, Popek, Mazur, and Bulińska, 2022). Most contemporary VR headsets and controllers support six degrees of freedom (6-DoF), enabling precise tracking across three translational and three rotational axes (Caserman, Garcia-Agundez, Konrad, Göbel, and Steinmetz, 2019). Kelkkanen, Lindero, Fiedler, and Zepernick (2023) demonstrated a positive relationship between VR tracking accuracy and the user's actual hand position during aiming and interaction tasks, emphasizing that predictability of target motion becomes increasingly important as latency and target speed increase. Similarly, Chang, Kim, and Yoo (2020) identified latency between user head movement and the corresponding visual feedback as a contributing factor to motion sickness in VR. Given that the majority of consumer VR systems rely on three-point tracking via headsets and controllers, generating a full-body avatar using only these devices presents a practical and advantageous solution.

2.1.2 Procedural and Baked Animation

Procedural animation refers to the generation of motion in real time through algorithmic methods rather than relying on predefined animation sequences or keyframes (Curtis, Adalgeirsson, Ciurdar, Knox, et al., 2022). For instance, Alberto, Luo, Navarro Newball, Zúñiga, and Lozano-Garzón (2019) employed procedural animation techniques to achieve lifelike virtual character behavior that adapts dynamically without requiring additional pre-authored animations.

In contrast, baked animation consists of precomputed motion data, typically created using motion capture systems or keyframe animation techniques, and stored for later playback (van Gumster, 2020). Baked animation is widely used in film and offline animation pipelines due to its consistency and predictability. Combining the adaptability of procedural animation with the stability of baked animation offers the potential to produce more realistic and responsive full-body avatars in VR.

2.1.3 Inverse Kinematics

Inverse kinematics (IK) is a widely used technique for controlling articulated character motion in animation systems. It operates by specifying desired end-effector positions, allowing joint configurations to be calculated automatically while preserving motion realism and structural constraints (Aristidou, Lasenby, Chrysanthou, and Shamir, 2017). Liu, Zhou, Wei, Su, Song, Kou, and Jin (2023) describe IK as a physics-driven approach that employs solvers to compute character poses, ensuring natural and physically plausible reactions to external forces. Similarly, Burdka and Rohleder (2017) presented a system that adapts character animations in real time based on environmental conditions, such as uneven terrain or obstacles. Overall, IK enables precise and adaptive control of character poses by dynamically adjusting joint positions and orientations, thereby enhancing interaction with virtual environments.

2.1.4 Virtual Reality

In this study, virtual reality (VR) is defined as a computer-generated, immersive three-dimensional environment that users can interact with using a head-mounted display

(HMD) and handheld controllers (Al-Ansi et al., 2023). Other researchers describe VR systems as primarily tracking the user's head and hand movements while omitting full-body tracking (Yang et al., 2022). Berni and Borgianni (2020) further characterize VR as a technology that supports design and engineering activities by enabling interaction within immersive digital environments. Accordingly, this work adopts a general definition of VR as an immersive 3D environment in which user interaction is facilitated through VR headsets and controllers, with tracking mechanisms focused on head and hand movements.

2.2 Theoretical Framing

The theoretical framework of this study integrates concepts from VR tracking technologies, animation techniques, and human–computer interaction to enhance user immersion and presence. This framework emphasizes the importance of realistic avatar motion achieved through the combined use of three-point tracking, procedural animation, baked animation, and inverse kinematics. The four hypotheses introduced earlier are grounded in this framework and focus on the following aspects:

- **H1**: Combining multiple avatar animation techniques enhances immersion more effectively than baked animation alone.
- **H2**: The inclusion of arms and lower body components improves immersion by providing a more complete avatar representation.
- **H3**: A close correspondence between avatar motion and user motion increases perceived immersion.
- **H4**: Realistic avatar motion is positively perceived by other users, contributing to a shared sense of realism and presence in virtual environments.

2.3 Combined Three-Point Tracking, Baked and Procedural Animation, and Inverse Kinematics for Full-Body Movement

Avatar animation in VR has been extensively studied, with approaches focusing on facial expressions, finger tracking, and full-body motion to improve user immersion. Yang et al. (2022) proposed *HybridTrak*, a system that uses a standard webcam and neural networks to reconstruct full-body motion by transforming 2D pose estimates into 3D poses combined with inside-out upper-body tracking data. Wu et al. Wu, Wang, Jung, Hoermann, and Lindeman (2021) employed highly expressive full-body avatars captured using multiple Kinect devices to study social interaction and interpersonal attraction in a virtual charades game. Malleson, Gilbert, Trumble, Collomosse, Hilton, and Volino (2017) introduced a real-time full-body motion capture system based on inertial measurement units (IMUs) and video input, while Caserman et al. (2019) combined HTC Vive tracking hardware with inverse kinematics to reconstruct full-body motion. Rogers, Broadbent, Brown, Fraser, and Speelman (2022) investigated social interaction in VR using lifelike avatar motion captured via a desktop-based VR setup.

Although these studies demonstrate significant improvements in avatar realism, they all rely on additional hardware such as webcams, Kinect sensors, IMUs, or external trackers. To the best of our knowledge, none of the existing approaches provides a solution for improving full-body avatar animation in VR environments without the use of supplementary tracking devices beyond the VR headset and controllers.

3. Technical Implementation

The technique presented in this study combines three-point tracking data obtained from VR hardware with inverse kinematics (IK), procedural animation, and baked animation through a mathematical formulation. As a result, users are able to perceive their entire virtual body from a first-person perspective, rather than being limited to floating hands. In addition to the technical implementation, questionnaire data were collected to assess how users perceived their own virtual bodies as well as how they perceived other users within the virtual environment.

3.1 Tools

Two of the most widely used game engines for developing VR applications are Unity (Unity Technologies, 2019) and Unreal Engine (Epic Games, 2023). Unity primarily uses C# as its programming language, whereas Unreal Engine is based on C++. Both engines are freely available and widely adopted by developers. Approximately 45% of game applications worldwide, including VR applications, are developed using Unity (Salama and Elsayed, 2021). By selecting Unity as the development platform for the proposed full-body avatar system, this study aims to maximize accessibility for developers, researchers, and universities.

The Unity XR plugin was employed to track the position and orientation of the VR headset and controllers. Although Unity was chosen for implementation, the proposed technique is not engine-specific and can be extended to other game engines. For evaluation purposes, the Meta Quest 1 was selected as the target device because it is a standalone VR headset that does not require an external computer for rendering. Demonstrating functionality on a low-end device such as the Quest 1 ensures that the proposed approach remains accessible to users with varying hardware capabilities.

To enhance avatar realism, baked animations obtained from Adobe Mixamo were integrated into the avatar ani-

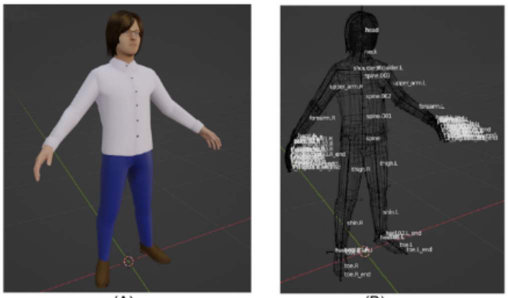

Figure 3. Model of the avatar used in the test study. (A) The avatar's mesh and (B) the armature structure of the avatar

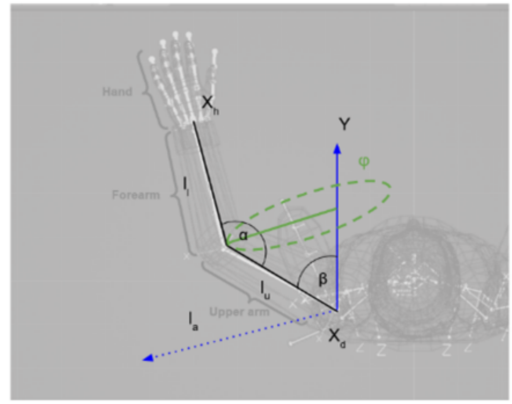

Figure 4. The schematic for arms IK calculations

mation pipeline (Adobe, 2023). Mixamo is an online platform that provides pre-authored animations for digital media projects. However, the proposed technique is not limited to Mixamo-generated animations. Because the system operates on animation keyframes, alternative sources such as motion capture suits can also be used to generate baked animations compatible with the approach.

To represent the user's body within the virtual environment, a full-body three-dimensional avatar with a humanoid skeleton was modeled using Blender 3.3 Blender Foundation (2023). Blender is an open-source 3D creation software commonly used for modeling, rigging, and animation. To further increase immersion, the avatar was equipped with facial expressions created using blendshape modeling. These expressions were triggered, for example, when the user spoke, with approximated blendshape activation derived from the user's voice input. Fig. 3 illustrates the avatar mesh together with the armature (skeleton) used in this project. The armature, which is described in more detail in the following chapter, enables avatar animation using either baked animation alone or the proposed combined animation approach.

3.2 Full-Body Avatar Implementation

The implementation of the proposed full-body avatar system was divided into two main components: upper-body tracking and lower-body tracking. Three-point tracking was employed to capture the position and orientation of the user's head and hands in three-dimensional space using the tracking sensors embedded in the VR headset and handheld controllers. To accurately animate the spine, arms, neck, and head in conjunction with three-point tracking data, a two-bone inverse kinematics (IK) solver was implemented.

Because no trackers or external sensors were attached to the lower body, a combination of inverse kinematics and procedural animation techniques was applied to animate the hips and legs. Additionally, baked animation was integrated to complement both approaches and to introduce fine-grained motion details, resulting in a coherent and realistic full-body avatar animation.

3.2.1 Upper-Body Tracking

To track the position and orientation of the user's head and hands, three-point VR tracking provided by Unity's XR library was utilized. The XR library supports a wide range of VR devices, including Meta Quest and HTC Vive headsets. Within the proposed system, the avatar's head and hands were directly mapped to the corresponding VR headset and controllers, respectively.

Fig. 4 illustrates the bone structure used by the implemented IK solver to animate the connection between the shoulders and arms based on the tracked hand positions. This structure enables the estimation of anatomically plausible upper-body motion while maintaining consistency with the user's real-time movements.

The arm IK solver estimates the position and rotation of the user's upper arm and forearm using tracking data obtained from the VR headset and controllers. The mathematical formulation of this solver is described in the following subsection.

The arm inverse kinematics (IK) solver establishes a kinematic chain between the shoulder joint and the hand target position and determines a feasible elbow position that satisfies anatomical constraints. Assuming fixed upper-arm and forearm lengths, the elbow configuration is computed using the law of trigonometry. Specifically, the elbow angle α and the corresponding shoulder angle are calculated as follows:

Using the law of trigonometry, the elbow angle α and the shoulder angle β are computed as follows:

$$\alpha = \arccos\left(\frac{l_u^2 + l_l^2 - \|\mathbf{x}_t - \mathbf{x}_u\|^2}{2\,l_u\,l_l}\right) \quad (1)$$

$$\beta = \arccos\left(\frac{\|\mathbf{x}_t - \mathbf{x}_u\|^2 + \|\mathbf{x}_h - \mathbf{x}_u\|^2 - \|\mathbf{x}_t - \mathbf{x}_h\|^2}{2\,\|\mathbf{x}_h - \mathbf{x}_u\|\,\|\mathbf{x}_t - \mathbf{x}_u\|}\right) \quad (2)$$

The elbow rotation φ is constrained by predefined minimum and maximum limits. The incremental elbow rotation for each axis i is computed as:

$$\varphi^i = \max\left(\varphi^i_{\min}, \chi^i_{h,s} - t^i \times w^i\right) \quad (3)$$

The final elbow rotation is then obtained by:

$$\varphi = \min\left(\varphi_{\max}, \max\left(\varphi_{\min}, \varphi_0 + \sum_i \varphi^i\right)\right) \quad (4)$$

The pivot point of the hand is denoted by $\mathbf{x}_h$, while the shoulder position is represented by $\mathbf{x}_d$. The forearm pivot point is denoted by $\mathbf{x}_u$, and the desired hand target position is indicated by $\mathbf{x}_t$. The total arm length is represented by l_d, where l_u and l_l denote the lengths of the upper arm and forearm, respectively. Finally, t^i represents the threshold value applied along each rotational axis.

3.2.2 Lower-Body Tracking

Because no additional sensors are attached to the user, lower-body motion is reconstructed primarily using inverse kinematics (IK) and procedural animation techniques. In the proposed system, IK is employed to estimate foot placement and leg articulation based on kinematic constraints and geometric relationships.

The foot IK algorithm is derived from the law of trigonometry and follows a formulation analogous to that used for upper-body articulation. This approach enables plausible leg motion and foot positioning without requiring explicit tracking data from the lower body. An overview of the geometric configuration and the corresponding IK formulation for the foot is illustrated in Fig. 5.

The thigh joint rotations, denoted by φ, have two degrees of freedom (DoF), namely forward/backward flexion and inward/outward rotation, whereas the shin joint has a single degree of freedom corresponding to forward/backward motion. The knee angle β is determined by the geometric configuration defined by segments a, b, and c. Using the inverse kinematics (IK) formulation, the position of the thigh and shin segments, as well as their rotations relative to the foot position, can be computed.

The hip rotation angle φ in the sagittal plane is computed as:

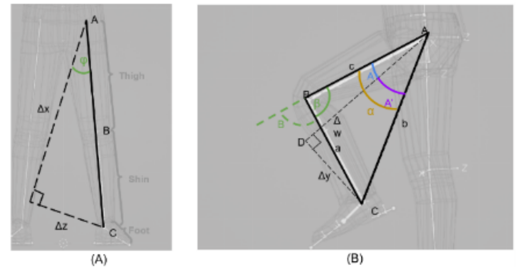

Figure 5. The schematic for the foot IK calculations (A) the front view and (B) is the side view

$$\varphi = \tan^{-1}\left(\frac{\Delta z}{\Delta x}\right) \quad (5)$$

The knee angle is derived using the law of cosines:

$$\beta = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right) \quad (6)$$

Here, Δx and Δz denote the horizontal and vertical displacement components between the hip and foot joints, respectively. Segments a, b, and c correspond to the geometric lengths of the virtual leg, where b and c represent the thigh and shin lengths, and a denotes the distance between the hip and foot target positions.

An additional rotational component is added to the hip rotation to account for vertical displacement between the hip and foot target positions. This component is computed as:

$$\varphi_{\text{add}} = \tan^{-1}\left(\frac{C_y - D_y}{\sqrt{(C_x - A_x)^2 + (C_z - A_z)^2}}\right) \quad (7)$$

Accordingly, the knee angle is adjusted using the following relation:

$$\beta = \pi - \cos^{-1}\left(\frac{a^2 + c^2 + b^2}{2ac}\right) \quad (8)$$

To ensure realistic foot placement, Unity's raycasting system was employed to detect intersections between the avatar's foot and the ground within the virtual environment. Raycasting is a common technique used to identify object collisions by projecting a ray and testing for intersections with scene colliders. In the proposed system, raycasting was used to determine whether the foot intersected

Volume: 2, Issue: 1, Page: 56-68, Year:2025

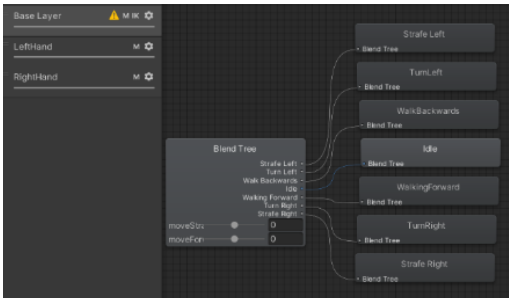

Figure 6. The layer's blending between procedural and baked animation

with the ground collider. Upon detecting ground contact, the foot position was constrained, allowing it to remain fixed to the ground surface and preventing visual artifacts such as foot sliding or floating.

3.2.3 Full-Body Tracking

Following the integration of three-point tracking, procedural animation, and inverse kinematics (IK), users were able to control the full-body avatar using only a VR headset and handheld controllers. Although this configuration enabled full-body control without additional sensors, the resulting avatar motion initially appeared rigid and lacked sufficient realism.

To enhance animation depth and visual fidelity, baked animations were incorporated into the animation pipeline. A set of baked animations, including walking, strafing, idle postures, and finger movements, was obtained from Mixamo and used to validate the proposed approach within a proof-of-concept implementation. Depending on the target application and interaction requirements, additional baked animations can be seamlessly integrated into the system.

Fig. 6 illustrates the Unity Animator window, demonstrating how baked animations are layered on top of procedural and IK-driven motion to produce more natural and expressive full-body avatar animations.

At this stage, the full-body avatar was fully animated and ready for deployment within the VR application. Users were required to wear only the VR headset and handheld controllers, with no need for additional external sensors. Fig. 7 presents examples of the avatar's movements within the virtual environment.

Due to the absence of direct reference data for lower-body tracking, the quantitative evaluation of lower-body motion accuracy is inherently limited. Nevertheless, as long as the avatar's movement appeared natural and visually coherent while navigating the virtual environment, user immersion was preserved. This effect can be attributed to the fact that the user's perceptual frame of reference is constrained to the virtual environment itself, reducing sensitivity to potential lower-body motion inaccuracies.

Figure 7. Poses comparison between real live person and full-body avatar

Figure 8. Virtual environment used for analysis of avatar impact on MyScore

3.3 Full-Body Performance Impact

Because VR meetings frequently involve multiple users interacting within the same virtual environment, evaluating the scalability of the proposed full-body avatar system is essential. Frames per second (FPS) was selected as the primary performance metric. FPS represents the number of frames rendered by the application per second, and maintaining a high FPS is crucial to ensure that avatar movements and interactions are displayed in real time without noticeable lag or stuttering. A consistently high FPS is therefore fundamental to preserving user immersion and to preventing adverse effects such as cybersickness.

To ensure the reliability and consistency of the performance measurements, a simplified virtual environment with a minimal rendering load was employed, as shown in Fig. 8. Complex scene rendering can significantly reduce FPS and confound performance evaluation. Because it was impractical to recruit and equip more than 70 participants, each using different VR hardware, FPS measurements were conducted using computer-controlled non-playable characters (NPCs) instead of human-controlled avatars. To maintain experimental validity, all NPC-driven avatars utilized the same full-body animation system and exhibited behavior equivalent to that of avatars controlled by human.

The *MyScore* application was subsequently deployed on a Meta Quest 1 VR headset, and FPS measurements were conducted under the defined experimental conditions.

Fig. 9 shows that the frame rate began to decrease when

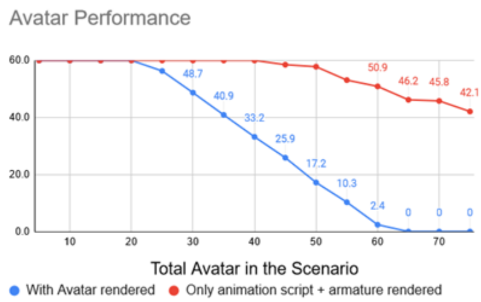

Figure 9. FPS measurement graph with a certain number of avatars in the scenario

more than 20 avatars were rendered simultaneously. When only full-body scripts and armatures were rendered, this threshold increased to approximately 45 avatars. These results indicate that, depending on the intended use case and the underlying VR hardware, the proposed full-body avatar technique scales to a reasonable number of avatars for typical multi-user VR meetings and can be effectively applied in general VR experiences.

4. The study

The performance evaluation presented in Section 3.3 demonstrates that the full-body avatar techniques proposed in this study are scalable and capable of approximating user motion. To investigate whether these improvements in avatar animation translate into an increased perception of immersion, an additional user study was conducted. Based on the research objectives, the following hypotheses were tested:

- **H1**: The use of combined avatar animation techniques (procedural animation, inverse kinematics, and three-point tracking) improves immersion in general VR scenarios compared to baked animation alone.
- **H2**: The inclusion of arms and lower body parts in the avatar representation improves user immersion in VR.
- **H3**: Avatar motion that closely correlates with user motion enhances user immersion.
- **H4**: Realistic avatar motion is positively perceived by other users within a virtual environment.

The study employed a quantitative experimental research design to evaluate the four hypotheses. Data were collected from 22 participants, enabling statistical analysis to examine how variations in avatar animation influence the VR experience. Participation in the study was entirely voluntary, ensuring that the collected responses reflected genuine impressions and reactions to the proposed full-body avatar system.

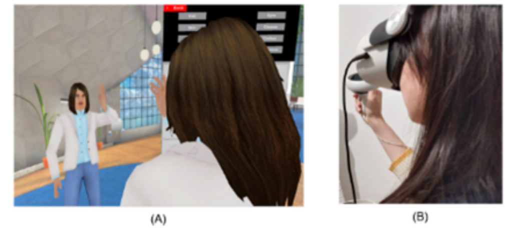

Figure 10. The participant assessing the avatar's movement in front of a virtual mirror (A) VR view (B) real life view

Figure 11. Screenshots taken during the test study (A) real life view (B) VR view

To test the first hypothesis, a chi-square test was applied. Descriptive statistics and the Wilcoxon signed-rank test were used to evaluate the remaining hypotheses. All statistical analyses were conducted using PSPP, an open-source statistical software package comparable to SPSS GNU Project (2023).

4.1 Experiment Phase

A three-part experimental study was conducted in Aachen, Germany, to evaluate users' perceptions of avatar control and the realism of avatar movement. In the first phase of the experiment, participants stood in front of a virtual mirror and performed a self-assessment of their own avatar while controlling it using a VR headset and handheld controllers, as illustrated in Fig. 10.

This procedure was carried out for two animation conditions: avatars animated using the proposed combined animation approach and avatars animated using baked animation alone. By directly comparing these two conditions, the experiment aimed to capture subjective differences in embodiment, control, and perceived realism from the user's perspective.

In the second phase of the experiment, participants evaluated the avatars of other users within the virtual environment. The virtual environment was based on a digital replica of the Katschhof in Aachen, Germany. During this phase, participants rated the perceived realism of the movements exhibited by the other avatars.

Fig. 11 illustrates the experimental setup used for both animation conditions.

Figure 12. Screenshots taken during the test study (A) real life view (B) VR view

In the final phase of the study, participants interacted with a simplified avatar representation that excluded both arms and lower body components, as illustrated in Fig. 12.

5. Findings and Discussion

At the conclusion of the experimental study, each participant completed a set of questionnaires designed to evaluate avatar movement. The questionnaire was divided into two parts. The first part assessed participants' preferences between baked animation and the proposed combined animation approach. The second part focused on evaluating users' perceptions of the combined animation employed throughout the experimental phases.

5.1 Demographic Data

Table 1 summarizes the demographic characteristics of the participants. A total of 22 participants took part in the study, including 17 males and 5 females, with ages ranging from 24 to 37 years. The mean age (μ) of the participants was 26. Eighteen participants were enrolled at a university, while the remaining four were full-time employees or scientific associates.

Regarding prior VR experience, 11 participants were familiar with VR but had never used it, 7 had occasional experience with VR, 2 were frequent VR users, and 2 participants reported no prior knowledge of VR. The majority of participants were students, with 16 individuals enrolled in bachelor's or master's degree programs. In addition, the sample included two scientific associates and four full-time employees.

The participants represented a diverse set of nationalities. The largest group originated from Germany (7 participants), followed by Indonesia (6 participants). The Netherlands, Bulgaria, and South Korea were each represented by two participants, while England, Turkey, and India each contributed one participant. This distribution highlights the international nature of the study and the broad interest in immersive VR technologies.

5.2 Animation Comparison Findings

Table 2 focuses on evaluating Hypothesis H1 by examining participants' sense of immersion across four distinct

Table 1. Demographic data of the participants

VR Familiarity	2	2	2	2	2	4	3	1	2	2	3	3	1	3	4	2	2	2	2	3	3	3
Sex	M	M	M	M	M	M	M	F	F	M	F	M	F	M	M	M	M	M	M	M	F	M
Age	25	25	24	27	25	29	28	25	26	26	26	26	35	37	28	27	24	24	30	27	28	25
Country	Id	Id	Ger	Ger	Ind	Ger	Ger	N	SK	Id	Id	Ger	SK	En	Ger	Ger	Id	Bul	Bul	N	Tur	Ind
Occupation	St	St	St	St	St	E	SA	St	St	St	St	St	E	E	SA	St	St	St	E	St	St	St

VR Familiarity: 1 = No prior knowledge of VR, 2 = Heard of VR but never used it, 3 = First experience with VR, 4 = Frequent VR user.
Sex: M = Male, F = Female.
Country: Id = Indonesia, Ger = Germany, Ind = India, N = Netherlands, SK = South Korea, En = England, Bul = Bulgaria, Tur = Turkey.
Occupation: St = Student, SA = Scientific Associate, E = Employee.

Table 2. User preference of avatar animation

No	Question	Baked	Neutral	Combined
1	Which avatar animation do you prefer?	0	3	19
2	Which animation enhances your feeling of immersion in VR?	1	5	16
3	Which type of avatar animation makes you feel more connected or engaged during VR interactions?	1	4	17
4	Which animations make the avatars more responsive in VR interactions?	2	5	15
			Average	17

Note: Expected frequencies $= 22/3 = 7.33$. Degrees of freedom $df = 3 - 1 = 2$. Significance level $\alpha = 0.05$. Critical value $= 5.99$.

questions. These questions were designed to reduce potential response bias and to capture multiple aspects of immersion, including animation preference, perceived immersion, engagement during VR interactions, and avatar responsiveness.

A chi-square goodness-of-fit test was applied to assess Hypothesis H1, which is appropriate for determining whether a statistically significant preference exists among the three animation conditions under the assumption of equal preference (Franke, Ho, and Christie, 2011). The test was conducted with two degrees of freedom ($df = 2$), expected frequencies of 7.33, a significance level of $\alpha = 0.05$, and a critical value of 5.99.

The results of Question 1 indicate that the majority of participants preferred the combined animation approach, while none selected baked animation alone. This preference is supported by a chi-square value of 28.5 ($p = 0.0001$). Question 2 further demonstrated that combined animation was perceived as significantly more immersive, yielding a chi-square value of 16.5 ($p = 0.0003$). Similarly, Question 3 showed that participants felt more connected and engaged during VR interactions when using combined animation, as reflected by a chi-square value of 19.7 ($p = 0.0001$). In addition, participants perceived combined animation as more responsive during VR interactions, supported by a chi-square value of 12.7 ($p = 0.0018$).

Across all four questions, the average chi-square value was 19.4, exceeding the critical value of 5.99, with an average significance level of $p = 0.0005$. These findings provide strong statistical support for Hypothesis H1, indicating that the combined animation approach significantly enhances user immersion compared to baked animation alone.

5.3 Combined Animation Findings

The questions presented in Table Table 3 were designed to evaluate Hypotheses H2, H3, and H4. Questions 1–4 addressed Hypothesis H2, Questions 5–7 targeted Hypothesis H3, and Questions 8–10 focused on Hypothesis H4. Participants rated each statement on a Likert scale ranging from 1 (lowest agreement) to 5 (highest agreement), with a neutral midpoint of 3.

Descriptive statistics, including the mean, median, and mode, were computed to summarize participants' evaluations of the combined animation approach. In addition, the Wilcoxon signed-rank test was applied to assess whether the observed ratings significantly deviated from the neutral score, which is appropriate for ordinal data and does not assume normality (Riffenburgh, 2012).

For Hypothesis H2, the average ratings yielded a mean of 3.7, a median of 4, and a mode of 5, indicating a generally positive perception of the inclusion of avatar arms and lower body components. Although the visibility of the user's own avatar arms did not yield a statistically significant result ($p = 0.38$), the alignment of arm movements with real-world actions ($p = 0.026$) and the visibility of avatar legs ($p = 0.004$) were found to significantly enhance immersion. This suggests that while arm visibility may be sensitive to approximation errors, lower-body representation plays a critical role in perceived immersion.

Hypothesis H3, which concerns the correspondence between avatar motion and user motion, produced consistently positive results. Participants reported a strong sense of embodiment when observing their avatars in a virtual mirror ($p = 0.002$) and expressed confidence in their ability to control the avatar as an extension of their own body ($p = 0.003$). Furthermore, realistic foot placement was positively evaluated ($p = 0.001$), supporting the importance of motion congruence in enhancing immersion.

Hypothesis H4 addresses the perception of avatar realism by other users in the virtual environment. Participants rated the realism of avatar movements positively, supported by a statistically significant result ($p = 0.008$). Accurate foot grounding was also perceived favorably, with a significance level at the acceptance threshold ($p = 0.05$). Finally, participants expressed an overall preference for the combined animation approach ($p = 0.006$), indicating that realistic and consistent avatar animation contributes meaningfully to user satisfaction and engagement.

In summary, the results of the descriptive statistics and Wilcoxon signed-rank tests provide converging evidence in support of Hypotheses H2, H3, and H4. Taken together, these findings demonstrate that the combined animation technique proposed in this study significantly improves user immersion and perceived realism in virtual reality environments.

Table 3. Participants' rating for combined avatar animation

No	H	Statement	Rating (1–5)					Mean	Median	Mode	W	*p*-value
			1	2	3	4	5					
1	H2	The avatar's arms in VR match your real-world arm movements	2	2	4	6	8	3.7	4	5	36	0.026
2	H2	I feel more immersed in VR when the avatar's arms are visible	1	3	4	5	9	3.8	4	5	27	0.009
3	H2	It is important for me to see my own avatar's arms in VR	2	3	8	6	3	3.2	3	3	39	0.380
4	H2	The visibility of the avatar's legs improves immersion	1	2	4	5	10	4.0	4	5	21	0.004
		H2 Average						3.7	3.8			
5	H3	I felt as if the avatar in the mirror was my own body	0	3	3	9	7	3.9	4	4	19.5	0.002
6	H3	I could control the avatar as if it were my own body	1	1	4	8	8	4.0	4	4/5	19	0.003
7	H3	The combined animation accurately mimics real-life foot placement	2	3	3	7	7	3.6	4	4/5	5	0.001
		H3 Average						3.8	4.0			
8	H4	The avatar's movements appear realistic	1	2	6	4	9	3.8	4	5	18.5	0.008
9	H4	The combined animation ensures correct foot grounding	0	1	5	8	8	4.0	4	4/5	46.5	0.050
10	H4	Overall, I like the combined avatar animation	1	1	6	6	8	3.9	4	5	16	0.006
		H4 Average						3.9	4.0			

Note: Neutral score = 3. Acceptance threshold set at $p < 0.05$. Wilcoxon Signed-Rank Test was applied for all hypotheses in this table.

6. Discussion and Conclusion

This chapter concludes the methodology for creating a full-body VR avatar as presented in this study and summarizes participant feedback on the proposed full-body avatar experience. In addition, the implications, recommendations, limitations, and potential future research directions are discussed.

6.1 Implications of the Study

This study presented a full-body VR avatar system based on a combined animation approach that integrates three-point tracking with inverse kinematics (IK), procedural animation, and baked animation. The proposed method enables users to control the upper-body motion of the avatar using only a VR headset and handheld controllers, without the need for additional sensors. Lower-body motion is approximated using the combined animation algorithm, allowing for plausible full-body movement despite limited tracking data.

A total of 22 participants (17 males and 5 females), ranging in age from 24 to 37 years with a mean age of 27, evaluated the quality of the full-body avatar. Four hypotheses related to avatar quality and user immersion were examined. Hypothesis H1 was validated using a chi-square goodness-of-fit test, demonstrating a clear participant preference for the combined animation approach over baked animation alone in enhancing immersion. Descriptive statistics and the Wilcoxon signed-rank test were employed to evaluate Hypotheses H2, H3, and H4.

The results indicate that the inclusion of avatar arms and lower body components significantly enhances user immersion, supporting Hypothesis H2. Hypothesis H3 was also confirmed, as improved correspondence between avatar motion and user movement led to higher perceived immersion. Furthermore, Hypothesis H4 was supported by findings showing that participants within the virtual Katschhof environment appreciated more lifelike movements of both their own avatars and those of other users. Overall, the use of avatars animated with different techniques contributed to a more naturalistic experience and increased users' sense of presence and engagement.

The proposed full-body avatar system demonstrably increases user immersion within the virtual environment. However, breaks in immersion may still occur when discrepancies arise between a user's physical lower-body movements and the avatar's animated response. Because VR applications are predominantly experienced from a first-person perspective, users rarely observe their own lower bodies directly. Consequently, immersion is most likely disrupted only in situations where users encounter reflections of their avatars, such as when standing in front of virtual mirrors.

6.2 Recommendations of the Study

The combined animation technique for full-body VR avatars could be further refined through the integration of advanced tracking systems or machine learning algorithms capable of predicting a wider range of motions from limited sensor input. In addition, expanding the participant pool to include a broader age range and more diverse backgrounds would provide deeper insights into user experience and immersion across different user groups. Ultimately, these efforts should aim to minimize breaks in immersion and ensure that avatar motion more accurately reflects users' actions in VR environments.

6.3 Limitations of the Study

Human movement encompasses a wide variety of motions, each with its own complexity, making it challenging to capture and reproduce every nuance accurately within a virtual environment. This challenge is particularly pronounced when attempting to represent full-body motion without additional body-mounted sensors beyond the VR headset and controllers. Although the majority of participants agreed that the proposed full-body avatar approach enhances immersion, the system is not capable of reproducing all possible movements. For example, complex actions such as cartwheels or rolling motions performed in the physical world cannot currently be translated accurately into VR using the proposed method.

6.4 Future Work

Future research could enhance the proposed technique by incorporating artificial intelligence (AI) models trained to predict lower-body motion more accurately based on the position and orientation of the VR headset and controllers. Additionally, the effectiveness of the combined animation approach should be evaluated using higher-end VR displays, such as the Meta Quest Pro and Meta Quest 3, to determine whether improved visual fidelity further enhances user immersion. Finally, high-performance VR systems could be employed to investigate the scalability of the proposed technique when animating a larger number of avatars simultaneously in shared virtual environments.

Acknowledgments

The authors gratefully acknowledge the support of the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst, DAAD) through their funding of the *MyScore* project.

References

Adobe. Mixamo automatic rigging and animation, 2023. URL https://www.mixamo.com/.

A. M. Al-Ansi, M. Jaboob, A. Garad, and A. Al-Ansi. Analyzing augmented reality (ar) and virtual reality (vr) recent development in education. *Social Sciences & Humanities Open*, 8(1):100532, 2023. doi: 10.1016/j.ssaho.2023.100532.

E. C. Alberto, X. Luo, A. A. Navarro Newball, C. Zúñiga, and C. Lozano-Garzón. Realistic behavior of virtual citizens through procedural animation. In *Proceedings of the IEEE International Conference on Virtual Reality and Visualization*, 2019. doi: 10.1109/ICVRV47840.2019.00057.

A. Aristidou, J. Lasenby, Y. Chrysanthou, and A. Shamir. Inverse kinematics techniques in computer graphics: A survey. *Computer Graphics Forum*, 37(6):35–58, 2017. doi: 10.1111/cgf.13310.

A. Berni and Y. Borgianni. Applications of virtual reality in engineering and product design: Why, what, how, when and where. *Electronics*, 9(7):1064, 2020. doi: 10.3390/electronics9071064.

Blender Foundation. Blender — free and open 3d creation software, 2023. URL https://www.blender.org/. Accessed: 2023-11-14.

Ł. Burdka and P. Rohleder. Applied inverse kinematics for bipedal characters moving on diverse terrain. *Machine Graphics and Vision*, 26(1/4):3–11, 2017. doi: 10.22630/mgv.2017.26.1.1.

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

E. Chang, H. T. Kim, and B. Yoo. Virtual reality sickness: A review of causes and measurements. *International Journal of Human-Computer Interaction*, 36(17):1658–1682, 2020. doi: 10.1080/10447318.2020.1778351.

C. Curtis, S. O. Adalgeirsson, H. S. Ciurdar, W. B. Knox, et al. Toward believable acting for autonomous animated characters. In *Proceedings of the ACM SIGGRAPH Conference on Motion, Interaction and Games*, 2022. doi: 10.1145/3561975.3562941.

Epic Games. Unreal engine, 2023. URL https://www.unrealengine.com/.

T. M. Franke, T. Ho, and C. A. Christie. The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation*, 33(3):448–458, 2011. doi: 10.1177/1098214011426594.

GNU Project. Pspp statistical analysis software, 2023. URL https://www.gnu.org/software/pspp/.

D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson. Immersive virtual reality as a pedagogical tool in education. *Journal of Computers in Education*, 8(1):1–32, 2021. doi: 10.1007/s40692-020-00169-2.

V. Kelkkanen, D. Lindero, M. Fiedler, and H. Zepernick. Hand-controller latency and aiming accuracy in 6-dof vr. *Advances in Human-Computer Interaction*, 2023:1–18, 2023. doi: 10.1155/2023/1563506.

P. Liu, Y. Zhou, X. Wei, Q. Su, W. Song, Q. Kou, and X. Jin. Character hit reaction animations using physics and inverse kinematics. *Computer Animation and Virtual Worlds*, 34(3–4), 2023. doi: 10.1002/cav.2170.

C. Malleson, A. Gilbert, M. Trumble, J. Collomosse, A. Hilton, and M. Volino. Real-time full-body motion capture from video and imus. In *International Conference on 3D Vision (3DV)*, 2017. doi: 10.1109/3dv.2017.00058.

R. H. Riffenburgh. *Tests on Ranked Data*. Elsevier, 2012. doi: 10.1016/b978-0-12-384864-2.00011-1.

S. L. Rogers, R. Broadbent, J. Brown, A. Fraser, and C. P. Speelman. Realistic motion avatars are the future for social interaction in virtual reality. *Frontiers in Virtual Reality*, 2, 2022. doi: 10.3389/frvir.2021.750729.

R. Salama and M. Elsayed. A live comparison between unity and unreal game engines. *Global Journal of Information Technology*, 11(1):1–7, 2021. doi: 10.18844/gjit.v11i1.5288.

A. Thampan, A. Razak, and S. Krishna. Evolution of augmented reality (ar) and virtual reality (vr). *International Journal of Research Publication and Reviews*, 4, 2023. URL https://ijrpr.com/uploads/V4ISSUE4/IJRPR12239.pdf.

Unity Technologies. Unity game engine, 2019. URL https://unity.com/.

J. van Gumster. *Blender For Dummies*. John Wiley & Sons, 2020.

Y. Wu, Y. Wang, S. Jung, S. Hoermann, and R. W. Lindeman. Using a fully expressive avatar to collaborate in virtual reality. *Frontiers in Virtual Reality*, 2, 2021. doi: 10.3389/frvir.2021.641296.

J. Yang, T. Chen, F. Qin, M. S. Lam, and J. A. Landay. Hybridtrak: Adding full-body tracking to vr using an off-the-shelf webcam. In *CHI Conference on Human Factors in Computing Systems*, 2022. doi: 10.1145/3491102.3502045.

M. Żuk, M. Wojtków, M. Popek, J. Mazur, and K. Bulińska. Three-dimensional gait analysis using a virtual reality tracking system. *Measurement*, 188:110627, 2022. doi: 10.1016/j.measurement.2021.110627.