VReflect: Evaluating the Impact of Perspectives, Mirrors and Avatars in Virtual Reality Movement Training

Dennis Dietz
LMU Munich
Munich, Germany
dennis.dietz@ifi.lmu.de

Fabian Berger
LMU Munich
Munich, Germany
fabianberger2@gmx.de

Changkun Ou
LMU Munich
Munich, Germany
research@changkun.de

Francesco Chiossi
LMU Munich
Munich, Germany
francesco.chiossi@ifi.lmu.de

Giancarlo Graeber
Catholic University of Eichstätt-Ingolstadt
Eichstätt-Ingolstadt
Germany
gigraeber@gmail.com

Andreas Martin Butz
LMU Munich
Munich, Germany
butz@ifi.lmu.de

Matthias Hoppe
Keio University Graduate School of Media Design
Yokohama, Japan
matthias.hoppe@kmd.keio.ac.jp

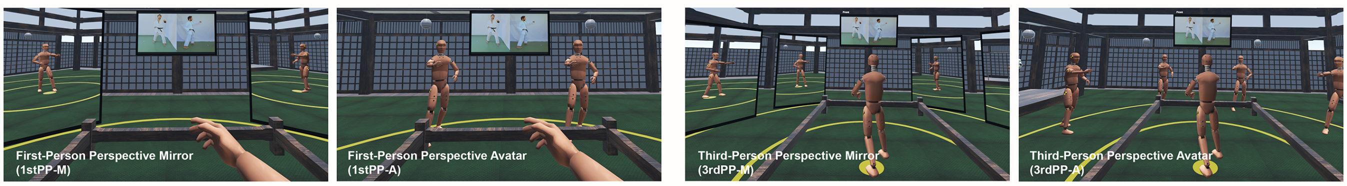

Figure 1: The 4 training settings: A screen presents karate moves demonstrated by an expert from two different perspectives. In the two left images, users experience the environment in the 1stPP perspective and see themselves in additional mirrors surrounding them. In the second image on the left, users see themselves represented by avatars that mimic their movements. The two right images present the same scenarios using the 3rdPP perspective [9].

Abstract

Virtual reality training systems require the careful design of content presentation, user embodiment, and overall user experience. We explore the impact of different perspectives (first-person and third-person) and virtual self-visualization techniques (VSVTs: mirrors and external avatars) on user embodiment, performance and experience. In a study with 28 participants learning karate movements, we tested four combinations of these factors. Results indicate that perspective influences visual focus and embodiment, while VSVTs affect movement execution, particularly in the third-person avatar condition. Measurements of physiological activity, workload, presence, and enjoyment found no significant overall advantages for any of the conditions. Interviews revealed that most participants preferred the familiar first-person mirror combination, although participants in third-person perspective focused more on their own body and noted the helpfulness of this viewpoint. The study demonstrates that alternative perspectives and visualization techniques offer valuable training options, as these conditions did not produce significant differences in measured cognitive load when compared with each other. Future VR training systems should incorporate interactive feedback and customization options to accommodate individual preferences and optimize learning experiences.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; *User studies*;
• **Applied computing** → Interactive learning environments.

Keywords

virtual reality, motor learning, perspectives, self-observation, embodiment

ACM Reference Format:
Dennis Dietz, Fabian Berger, Changkun Ou, Francesco Chiossi, Giancarlo Graeber, Andreas Martin Butz, and Matthias Hoppe. 2025. VReflect: Evaluating the Impact of Perspectives, Mirrors and Avatars in Virtual Reality Movement Training. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766021

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/25/11
https://doi.org/10.1145/3756884.3766021

1 Introduction

Virtual reality (VR) is increasingly utilized for motor skill training in domains like sports and rehabilitation, offering powerful tools to improve training outcomes [53, 55]. However, for these systems to be truly effective, they must adapt to individual learner needs, a challenge that requires optimizing visual feedback mechanisms [60]. Two fundamental, yet underexplored, factors that shape this feedback are the user's visual perspective and the method of self-observation [38]. The interaction between these elements is critical, but their combined impact on motor learning remains poorly understood.

Users typically experience VR from a first-person perspective (1stPP), which promotes high immersion, whereas a third-person perspective (3rdPP) provides a detached viewpoint that can improve overall body awareness [17]. Rather than being two polarities, these viewpoints exist on a continuum in VR, offering a rich design space [22]. To leverage these perspectives for self-observation, VR systems employ what we term Virtual Self-Visualisation Techniques (VSVTs), which are virtual objects that visualise the user's avatar and movements. We identify two psychologically distinct types: **virtual mirrors**, which leverage powerful principles of self-recognition through a familiar, laterally inverted reflection [25, 34]; and **external avatars**, which act as independent, anatomically correct (non-mirrored) representations of the user, facing the user.

When, e.g. an **external avatar** is combined with a **3rdPP**, it offers an additional viewpoint for analysing complex movements. However, this specific combination can introduce unique coordination challenges precisely because the user must mentally map their actions onto a separate, non-mirrored body in the scene [48]. A direct comparison of how these psychologically distinct VSVTs (mirror vs. external avatar) function across different perspectives is missing, limiting our ability to design truly adaptive training systems.

To address this gap, we present a detailed analysis from a user study that systematically investigates these factors. Our work is guided by the following research questions:

- **RQ1:** How do the interactions between perspectives (1stPP vs. 3rdPP) and VSVTs (mirrors vs. external avatars) comparatively affect movement accuracy and visual attention during a motor learning task?
- **RQ2:** What are the subjective effects of these visual conditions in terms of embodiment, cognitive workload, presence, and enjoyment?
- **RQ3:** Which combination of perspective and VSVT do users qualitatively prefer for virtual movement training, and what reasons underlie these preferences?

Our findings reveal a crucial trade-off between visual conditions rather than a single optimal setup. While no condition offered a universal performance advantage, we found that perspective significantly influenced the sense of embodiment (SoE) and visual focus. A key result was a notable performance decrease in the 3rdPP-avatar condition, where participants struggled with rotational accuracy due to losing visual contact with their avatar. Furthermore, qualitative feedback underscored the importance of moving beyond passive observation, highlighting a clear user desire for interactive feedback and customizable systems to accommodate individual learning preferences. Our contribution is threefold: we provide empirical evidence on how perspectives and VSVTs shape performance, embodiment, and user experience in VR training, showing that alternatives such as third-person avatars can enhance body awareness without reducing effectiveness (I); we demonstrate that these alternatives broaden training options without resulting in a higher measured cognitive load relative to the other tested setups, supported by motion, eye-tracking, and interview data (II); and we propose design recommendations for adaptive VR systems, including perspective switching and interactive feedback to align with individual learning preferences and improve skill acquisition (III).

2 Related Work

2.1 Perspectives in Games and Virtual Reality

Camera placement in video games is a decisive factor for the game's look, and further impacts its playing style and interactions, resulting in many genres. The two most ordinary perspectives in desktop video games, according to Denisova and Cairns [8], are 1stPP and 3rdPP, each having their respective advantages and disadvantages. In recent years, research projects have used different perspectives in VR for design goals such as enabling locomotion, improving task performance, and as control schemes [10, 14, 42]. Therefore, we see that VR perspectives offer novel movement learning qualities that have not been extensively explored yet. A small body of work focuses on the use of different perspectives for movement learning [13, 21, 59] We see 1stPP and 3rdPP as opportunities to provide a training scenario in a new perspective. While these perspectives are well-established in traditional gaming, their implementation and perception in VR differ significantly

Recently, developers have also brought the 3rdPP to VR games. While offering novel game mechanics, they remain a rarity as VR games and experiences are usually designed as 1stPP due to the nature of VR hardware. However, Hoppe et al. [22] presented that VR perspectives are less distinct than their traditional desktop equivalents. While a VR experience is designed with what they term an *Out-of-Character view* (3rdPP), the subjective experience differs from user to user. Some might feel like a puppeteer controlling the character from the outside (*Out-of-Character embodiment*), while others still feel embodied in the character they are observing (*Within-Character embodiment*). Because this nuanced terminology by Hoppe et al. [22] is not yet widely adopted, we will stick to the more familiar technical distinction of 1stPP and 3rdPP in this paper, while being mindful of these perceptual differences.

2.2 Virtual Self-Visualization Techniques

Mirrors have long been studied in psychology for their effects on self-perception and body ownership [26, 27], providing basic principles that inform mirror use in virtual environments [50].

Building on these psychological insights, research has increasingly examined mirrors in VR contexts. A meta-analysis by Mottelson et al. [38] of 111 studies on Body Ownership Illusions (BOIs) shows ongoing research on embodiment, but highlights that combinations of 3rdPP with mirror reflections remain underexplored, with most studies focusing on 1stPP. This underscores methodological gaps and the need for comparative studies on feedback mechanisms. HCI research has applied mirrors in various VR functions, from interaction metaphors [6, 33] to self-perception tools in social contexts [15] and movement feedback [4, 23, 52]. Notably, Elsayed et al. [13] explored one of the rare examples where 3rdPP perspective was combined with mirrors for movement guidance. Shifting to avatars, this approach provides direct digital representations of users, viewable from various perspectives. Unlike reflective mirrors, avatars significantly impact embodiment [51], task performance [32], and self-perception [31]. This can lead to behavioral changes via the Proteus effect [18, 57], offering opportunities for movement training beyond mirror-based methods. Building on this distinction, we use our term VSVTs to collectively analyze mirrors and external avatars. We address the existing gap by systematically comparing their effects in movement training, particularly the underexplored third-person mirror combination, to examine impacts on learning outcomes and user experience.

2.3 Movement Training Approaches

Movement learning is fundamentally rooted in observation and imitation, as established by Bandura's social learning theory [3]. Digital training systems have translated this principle into two primary feedback approaches: imitating an expert and direct self-observation. Due to the scarcity of human trainers, research has explored digital solutions for movement training. Anderson et al. [2] developed a system using Microsoft Kinect™ to record expert movements, providing guidance and feedback through an augmented reality mirror. Ikeda et al. [24] created a real-time golf training system where learners adjust their posture to match an expert's silhouette. Sun et al. [49] introduced a machine-learning model for basketball training that adjusts goals based on user performance. Chua et al. [7] developed a wireless Tai Chi training application, allowing more freedom of movement. Han et al. [19] demonstrated an augmented learning tool for Tai-Chi Chuang, where learners can view different angles of movements and see themselves through a drone's perspective. Hoang et al. [21] presented a system where users see themselves from a first-person perspective with an expert's skeleton overlay. Embodying the user in different avatars and realities was the focus of the research done by Pastel et al. [40] in a karate context. They explored the effects of different embodiments on the learning ability of complex movements. While much research has focused on transferring teacher information to the user, more work is needed on utilising the environment to view one's own posture and movements. Our research addresses this gap by systematically comparing the effects of different perspectives (1stPP and 3rdPP) and VSVTs on movement training, embodiment, and overall user experience in VR training environments, thereby exploring how these design choices influence performance outcomes and users' visual focus and perceived workload during skill acquisition.

3 Concept & User Study

To investigate perceptual effects and user performance across different perspectives and VSVTs in virtual movement training, we developed VReflect. This system and the study design build upon and significantly extend our preliminary work presented in [9]. The design of this VR training system was grounded in a dual approach: insights from previous literature and requirements discussions with experts from martial arts, yoga, and dance [14, 19, 40]. Specifically tailored for novices, VReflect consists of a virtual dojo, training material, and four distinct setups combining perspectives and VSVTs. We chose karate as the training task because its techniques involve complex, full-body movements that are sensitive to visual feedback and are divisible into distinct, measurable blocks, allowing for standardized performance analysis [11, 40].

The scenarios implemented in VReflect leverage established learning principles. We use **perspectives** to control the user's viewpoint, where the first-person perspective (1stPP) is known to enhance immersion and agency [17], while the third-person perspective (3rdPP) can provide a better overview for self-correction and body awareness [45]. For **VSVTs**, we employed virtual mirrors, which facilitate real-time self-assessment [23], and external avatars, which can increase embodiment and influence behavior through effects like the Proteus effect [51, 57]. We conducted a 2x2 within-participants experimental design (see Figure 4). The independent variables were Perspective (with levels: 1stPP and 3rdPP) and VSVT (with levels: mirror and external avatar). To mitigate potential learning effects, we used a Latin Square randomization to determine the order of the four conditions for each participant group.

The first independent variable, Perspective, controlled the user's viewpoint. We included the first-person perspective (1stPP) as it is the standard for immersive VR and known to enhance agency and immersion [17]. In contrast, the third-person perspective (3rdPP) was implemented as a fixed view positioned 1.5 meters behind and slightly above the user. This specific placement was empirically determined for our setup. It was chosen to provide a clear, unobstructed view of the avatar's full body for self-assessment while keeping the instructional screens visible. As the literature indicates, the ideal camera distance is highly task-dependent and varies significantly (e.g., 1.2m to 3m) based on hardware and interaction goals [14, 16, 17].

The second independent variable, VSVT, determined the method of self-observation. We compared virtual mirrors, which provide an intuitive but laterally-inverted view for self-correction, with external avatars, which offer a non-mirrored, anatomically correct depiction of the user's movements from multiple simultaneous viewpoints.

As the baseline condition for our comparison, we chose the 1stPP-Mirror (1stPP-M) configuration. This decision was grounded in its ecological validity, as it closely replicates real-world training environments, such as dojos or fitness studios, where mirrors are a standard tool for self-observation and movement correction. Therefore, this condition represents the digital equivalent of a typical training scenario, serving as a practical reference point against which we evaluated the alternative perspectives and VSVTs [4, 40]. This resulted in four conditions (see Figure 1):

1stPP-Mirror (1stPP-M) The standard 1stPP perspective combined with eight virtual mirrors reflecting the self-avatar's movements.

1stPP-Avatar (1stPP-A) The 1stPP perspective, where mirrors were replaced by eight external avatars mimicking the user's movements, always oriented towards the user's self-avatar.

3rdPP-Mirror (3rdPP-M) A fixed 3rdPP perspective surrounded by eight virtual mirrors reflecting the avatar's actions.

3rdPP-Avatar (3rdPP-A) The 3rdPP perspective, with eight external avatars providing multiple, non-mirrored views, always oriented towards the user's self-avatar.

3.1 Apparatus

3.1.1 Hardware. Our setup utilized a dual-PC configuration to ensure high-performance data recording and rendering. The main PC (Intel Core i9-12900, 32GB RAM, Nvidia RTX 3070) ran the Unity scene and managed the HTC Vive Pro Eye headset. A second PC (Intel Core i7-4790, 16GB RAM, Nvidia GTX 750 Ti) was dedicated to managing the optical tracking system. In addition to motion and eye-tracking data, our primary PC also recorded ECG data from a Polar H10 heart rate chest band.

3.1.2 Motion Tracking. We used a hybrid system to capture ten tracking points (head, hands, hip, elbows, knees, and feet). Head tracking was provided by the Vive Pro Eye headset itself. Hand and hip movements were captured using Valve Index controllers and a single HTC Vive Tracker 3.0. The remaining tracking points (elbows, knees, and feet) were tracked using a 16-camera OptiTrack system operating at 100Hz. For the elbows and knees, we used Velcro straps with custom marker sets (six for each elbow, five for each knee), repositioning the pivot points in the Motive Studio software to the respective joint's tip for accuracy. For the feet, we used OptiTrack's integrated marker shoes. All positional data from the OptiTrack system was streamed over the network to the main PC and integrated with the Vive Index data.

3.1.3 Avatars. This data was used to animate the participant's self-avatar in real time using the *Final IK* toolkit [44]. To prioritize inclusivity and minimize potential bias, this self-avatar was intentionally designed as a gender-neutral, abstract puppet-like figure. We opted for this minimalistic representation to ensure participants focused on their movements rather than the avatar's appearance, thereby preventing distractions and fostering a more universal training experience. The same model was used for the external avatars to maintain visual consistency.

3.1.4 Virtual Dojo. The virtual environment was based on a commercially licensed model of a traditional Japanese dojo [41], which we adapted and extended to create an authentic and immersive training atmosphere. The room featured a clearly defined front and back, each equipped with a large display screen (8x4.5 meters) showing the expert videos. The dojo itself had wooden floors and shoji screen walls. The central training area was designed around the participant, whose starting position was always at the center of the room. This area was defined by the arrangement of the VSVTs: depending on the condition, eight virtual mirrors or eight external avatars were positioned in an octagonal pattern around this central point Figure 2. To ensure an unobstructed view of both screens, the VSVTs at the front and back were placed adjacent to the displays. This setup was designed to provide a comprehensive, 360-degree view for self-observation while maintaining clear visibility of the instructional material.

3.1.5 Teaching Material. The training content consisted of four different video sequences, each demonstrating three karate movements performed by an expert. While the sequences differed between conditions, they were carefully selected in consultation with our expert to be of comparable difficulty for novices, thus minimizing task complexity as a confounding factor. Each sequence was exclusively assigned to one of the four experimental conditions. The

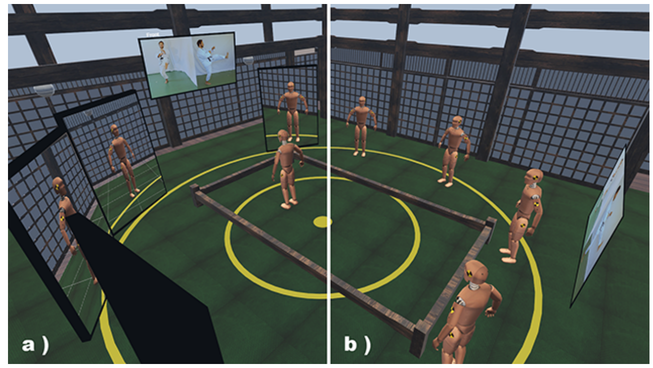

Figure 2: A bird's-eye view of the virtual dojo, illustrating in a split image the two VSVT conditions: (a) training area surrounded by virtual mirrors, and (b) by external avatars.

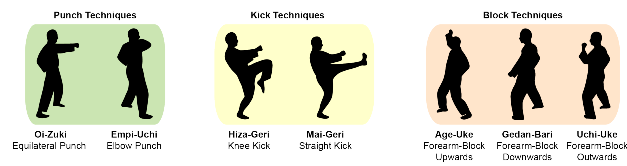

Figure 3: Karate moves for the virtual training environment. One move was chosen from each block to ensure equal difficulty across the different training conditions.

sequences were: **Elbow-Kick-BlockMid (Empi-Geri-Chudan-Uke)** (in the 1stPP-M condition), **Punch-Knee-BlockDown (Tsuki-Hiza-Geri-Gedan-Barai)** (in the 1stPP-A condition), **Knee-Elbow-BlockMid (Hiza-Geri-Empi-Chudan-Uke)** (in the 3rdPP-M condition), and **Kick-Punch-BlockUp (Geri-Tsuki-Age-Uke)** (in the 3rdPP-A condition) (see Figure 3). The videos were recorded from a frontal and side perspective and displayed on the large screens, side by side (split screen), in the virtual dojo.

3.2 Participants

We recruited 28 volunteers (15 female, 13 male; M=26.88 years, SD=6.29) for this study, which received approval from the local ethics committee. An inclusion criterion was having less than two years of prior martial arts experience to ensure a novice skill level appropriate for the selected training movements. All participants were right-handed and had normal or corrected-to-normal vision. Only two participants reported having no previous experience with VR. The entire session lasted approximately two hours, and participants received monetary compensation (15 Euros).

Due to corrupt or missing motion data from four individuals, we excluded their data from our analysis. Consequently, all quantitative and qualitative results presented in this paper are based on a final sample of 24 participants.

3.3 Procedure

The study procedure consisted of three main phases: preparation, experimental trials, and post-study debriefing.

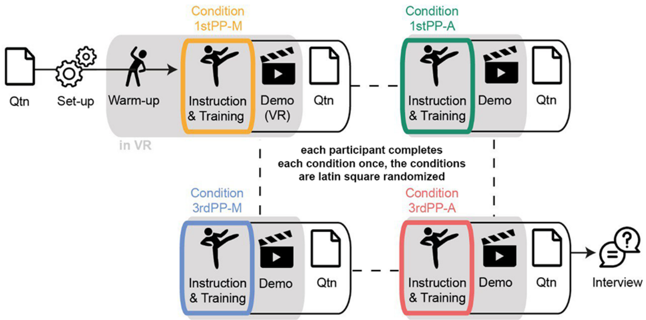

Figure 4: After a demographic questionnaire and a consent form, introduction, setup and warm-up in VR, participants experienced four conditions with all possible combinations of First- vs. Third-Person and Mirror vs. Avatar (1stPP-M, 1stPP-A, 3rdPP-M, 3rdPP-A) in a counterbalanced order. Each condition included an 'Instruction & Training' block with a 10-minute session where participants watched an instruction video and trained on three moves. In the 'Demo' session, participants demonstrated the moves they learned. After each demonstration, they answered a set of questionnaires. Finally, a concluding semi-structured interview was conducted.

Preparation. Upon arrival, participants were briefed on the study's purpose, duration, and procedures before providing written informed consent. They then completed a survey covering demographics, sports background, and prior VR experience. Afterwards, the experimenter assisted in attaching the heart rate sensor and calibrating the motion capture system to the participant's body. Finally, participants put on the HMD, performed an eye-tracking calibration, and familiarized themselves with the controllers.

Experimental Trials. The main experiment consisted of four trials, one for each experimental condition, with the order counterbalanced across participants using a Latin Square randomization. The entire in-VR session was guided by a standardized, computer-generated voice to ensure consistent instructions. The session began with a brief avatar calibration and a guided warm-up exercise. Throughout the trials, we periodically monitored for discomfort using the Fast Motion Sickness Scale [30]. Each of the four trials followed an identical structure:

(1) **Instruction:** The voice explained the current condition's setup. Participants then watched an instructional video of the three-move karate sequence assigned to that condition. The full sequence was shown twice, followed by each of the three moves shown individually twice. Finally, participants were asked to replicate the movements while the video played two more times.
(2) **Training:** A 10-minute training session followed, in which the instructional video looped continuously. Participants could end the session earlier if they felt they had mastered the movements.
(3) **Demonstration:** After the training, all VSVTs were removed from the scene, and the view was reset to the 1stPP perspective for all participants. They were then asked to perform the learned three-move sequence from memory.
(4) **Questionnaires:** After each demonstration, participants removed the HMD to complete a set of questionnaires (raw NASA-TLX [20], IPQ [46], Embodiment Questionnaire (EG) [17], Physical Activity Enjoyment Scale (PACES) [29]) via the survey on the computer. This break was intentionally designed to reduce potential fatigue from prolonged VR use. The time taken to complete the questionnaires provided a standardized rest period between conditions, which typically lasted between 5 to 7 minutes.

Post-Study Debriefing. After the final trial and questionnaires, we conducted a semi-structured interview with each participant to gather qualitative feedback on their experience, preferences, and the perceived effectiveness of the different conditions.

3.4 Measurements and Data Preprocessing

To answer our research questions, we collected and analyzed a combination of objective and subjective data. The dependent variables and their corresponding research questions are as follows:

- To answer **RQ1 (Performance & Visual Focus)**, we analyzed motion data for movement accuracy, eye-tracking data for visual attention patterns, physiological data (ECG) for physical exertion, and NASA-TLX scores for cognitive workload.
- To answer **RQ2 (Subj. Experience)**, we used questionnaire data on embodiment (EQ), presence (IPQ), and enjoyment (PACES).
- To answer **RQ3 (Qualitative Preferences)**, we analyzed qualitative feedback from semi-structured interviews.

3.4.1 Motion Data. We recorded participants' motion data as described in Section 3.1. To quantitatively assess movement accuracy, we compared participants' motion data from the demonstration phase against pre-recorded reference movements from our karate expert. For this comparison, we employed Dynamic Time Warping (DTW), a standard and widely used method in HCI for analyzing time-series data such as gestures and full-body movements [1, 61]. DTW calculates an optimal alignment between two temporal sequences and returns a distance score, where a lower score indicates a higher similarity between the participant's and the expert's movement. We specifically chose DTW for its robustness in handling temporal misalignments, which are expected when comparing novice and expert performances [28]. During the demonstration session for each condition, participants performed the movement sequence once. We cleaned this data by visually reviewing it and removing irrelevant segments. Due to data loss for the elbow and knee joints, we excluded them from the analysis and focused on the head, waist, hands, and feet. Following established procedures, the data was standardized using z-scores across six dimensions (position and rotation) [47]. We report positional differences in centimeters and rotational differences in degrees.

3.4.2 Eye-Tracking Data. We recorded eye-movement data using the built-in 120Hz eye tracker of the HTC Vive Pro Eye via the Tobii XR SDK. This allowed us to capture eye-gaze vectors and analyze participants' visual attention. We focused on two metrics: the total time participants looked at specific virtual objects (self-avatar, surrounding avatars, mirrors) and how frequently they shifted their gaze between the front and back instruction screens. The instructional videos themselves were excluded from this analysis, as viewing them was a necessary part of the task. Both metrics provide insights into how visual focus was affected by the conditions.

3.4.3 Physiological Data (ECG). During all sessions, we measured participants' heart rate (HR) using a Polar H10 chest band (130 Hz). The raw ECG data was processed using the Neurokit Python Toolbox [35], which involved filtering and QRS complex identification to extract mean HR values. This served as an objective indicator of physical engagement and exertion [43]. While we did not normalize HR changes against a resting baseline for each individual trial, our counterbalanced, within-subjects design allows for a direct comparison of the mean HR values across the four conditions to identify relative differences in physical exertion. This approach is common in studies where the primary goal is to compare exertion between experimental conditions rather than quantifying absolute workload [56, 58].

3.4.4 Questionnaire Data. After each condition, participants completed a set of validated questionnaires on a computer. We assessed cognitive workload using the raw NASA-TLX [20]. To measure subjective experience, we used the IPQ [46] for sense of presence, the PACES [29] for enjoyment, and the EQ by Gorisse et al. [17] for the SoE.

3.4.5 Qualitative Interview Data. We supplemented the quantitative with qualitative data from semi-structured interviews conducted at the end of the study. Our interview protocol was based on questions developed by Elsayed et al. [13] regarding perception and experience with VR movement guidance systems. We adapted and extended these questions to specifically address our research focus on perspectives and VSVTs. The full list of questions that guided the interviews can be found in the supplemental materials. The interviews were conducted with each of the 24 participants and had an average duration of 9:10 minutes (min: 04:55, max: 14:39). Audio recordings were transcribed and thematically coded to analyze user experiences, feature preferences, and comparisons to other training methods.

4 Results

Our results demonstrate a holistic approach from quantitative measures from questionnaires, objective measures, and qualitative measures from our interviews. Our participants repeated four sessions per person, for which we used a two-by-two mixed method design, with the participant as a random effect. To validate the normality, we used the Shapiro-Wilk test [37]. We further examined our data using ART ANOVAs for the significant, not normal, distribution [54] (with formula `measure ~ VSVTs×perspective+(1|participant)`). Otherwise, we used the two-way ANOVA. Furthermore, we continued with the ART-C post hoc test for our ART ANOVA results. We make our data and preprocessing scripts openly available on the Open Science Framework at this link [1].

4.1 Motion

Since the standardized DTW-D distance on position and rotation violated the normality assumption ($W = .972$, $p < .001$), we conducted an ART ANOVA to analyze main effects of two independent variables, namely VSVT and Perspective, as well as their interaction, on the movement rotation (see Figure 5). Results in Table 1 showed a significant main effect of VSVT on movement rotation. However, the main effect of Perspective was not significant. There was a significant interaction between VSVT and Perspective. Post-hoc pairwise comparisons used Bonferroni's method to explore the significant interaction effect further.

Table 1: Overview of results analyzed using ART ANOVAs. Significant results are highlighted in bold.

	VSVT			Perspective			V × P		
	df	F	p	df	F	p	df	F	p
Position DTW-D	1	1.39	.240	1	3.20	.074	1	2.90	.089
Rotation DTW-D	1	9.836	**.002**	1	.50	.480	1	6.307	**.012**
ET Total Watch Count	1	.86	.356	1	1.48	.228	1	.51	.478
ET Rotation Count	1	.01	.928	1	5.02	**.029**	1	1.02	.316
HR	1	.69	.410	1	.65	.424	1	3.28	.075
NASA-TLX	1	.98	.324	1	.02	.900	1	1.28	.258
IPQ	1	.05	.824	1	.00	.990	1	.86	.354
Embodiment	1	.17	.681	1	4.20	**.041**	1	.00	.989
PACES	1	.38	.537	1	2.76	.097	1	3.69	.055

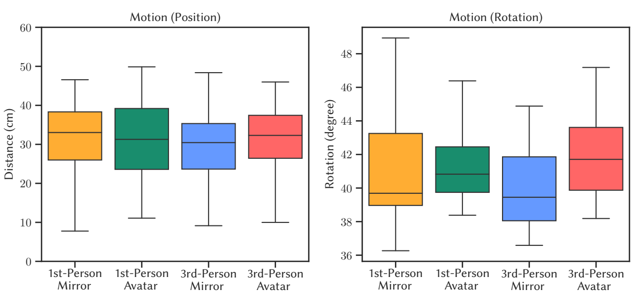

Figure 5: Normalized DTW distance for position (left, cm) and rotation (right, degrees), where lower values mean higher accuracy compared to the expert. A significant interaction effect was found for rotational accuracy. Note: The y-axis on the rotation plot is truncated to better visualize the significant effect, as a full scale from zero would obscure the differences.

Participants in the 1stPP-A condition provided numerically lower scores compared to those in the 3rdPP-A condition, although this difference was not statistically significant ($t = -1.27$, $p = .412$). Similarly, no significant difference was observed between the 1stPP-A and VSVT conditions ($t = 0.47$, $p = .64$).

In the 3rdPP-A condition, participants showed significantly larger rotation deviations from the expert compared to the 1stPP-A condition ($t = 4.18$, $p < 0.001$) and the 3rdPP-M condition ($t = 2.91$, $p = .019$). This suggests that the combination of 3rdPP with an external avatar impairs the accuracy of rotation movements, possibly because participants temporarily lose sight of their avatar during rotations and thus receive less visual feedback. No significant differences were found between the 1stPP conditions (mirror vs. externa-avatar), suggesting that the type of self-observation in 1stPP has less influence on movement accuracy.

[1]https://osf.io/zchx8/

4.2 Eye Tracking

4.2.1 Total Watch Count. An ANOVA was conducted to examine the effects of VSVTs and Perspective on Eye Tracking total watch count as data were normally distributed. The main effects of VSVTs ($F(1, 69) = 0.86$, $p = 0.356$) and Perspective ($F(1, 69) = 1.48$, $p = 0.23$), as well as their interaction ($F(1, 69) = 0.51$, $p = 0.48$), were not statistically significant ($p > 0.05$). These results suggest that neither VSVTs nor Perspective had a significant impact on Eye Tracking total watch count (see Figure 6).

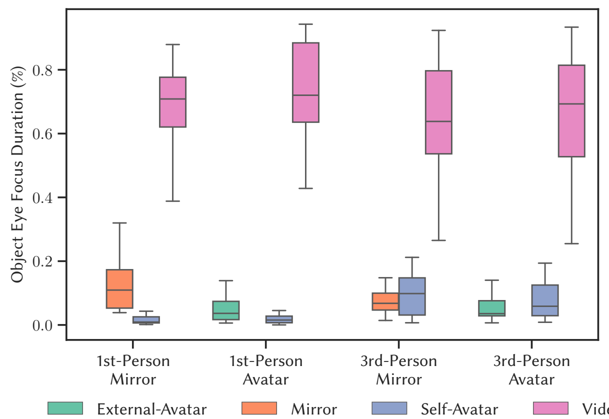

Figure 6: Percentage of focus duration on specific objects over the time of each session. We excluded the significantly higher duration for the video from further calculations.

4.2.2 Rotation Count. An ANOVA was conducted to examine the effects of VSVTs and Perspective on rotation count. The main effect of VSVTs ($F(1, 57) = 0.01$, $p = .928$) and the interaction between VSVTs and Perspective ($F(1, 57) = 1.02$, $p = .32$) were not statistically significant ($p > 0.05$). However, there was a significant main effect of Perspective ($F(1, 57) = 5.02$, $p = 0.03$ with Bonferroni correction in ART contract), indicating that Perspective had a significant impact on turn count. These results suggest that while VSVTs did not significantly influence rotation count, Perspective significantly affected this measure (see Figure 7).

4.3 Electrocardiogram

We conducted an ART ANOVA to examine the effects of VSVTs and Perspective on heart rate (HR). The results revealed no significant main effects or interactions (all $p > .05$), indicating that neither VSVTs nor Perspective significantly impacted participants' physiological responses during the training sessions.

4.4 Workload, Presence, Enjoyment

For the subjective measures collected via questionnaires, we conducted ART ANOVAs to examine the effects of VSVTs and Perspective. Neither NASA-TLX (workload), IPQ (presence), nor PACES (enjoyment) showed significant main effects or interactions (all $p > .05$), suggesting that neither perspective nor visualization technique significantly impacted participants' subjective experience of workload, presence, or physical activity enjoyment.

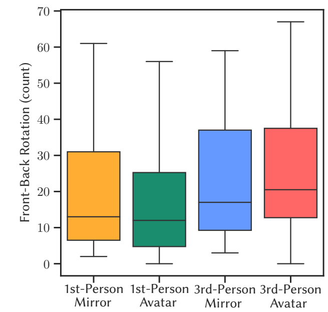

Figure 7: We calculated how often participants would rotate between front and back screen, based on the eye tracking (ET) data. We see significant differences for the perspective, showing less rotations in 1stPP.

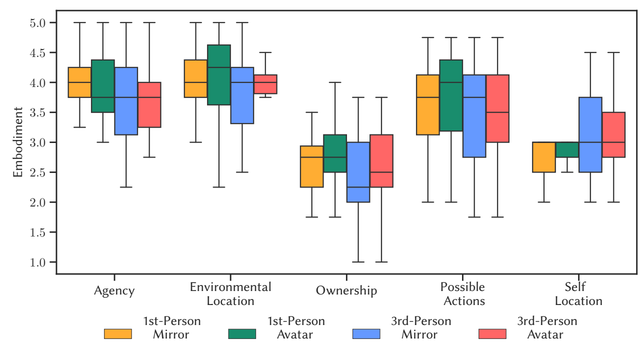

Figure 8: Participants rated the feeling of embodiment on a Likert scale in different categories (results shown over all sessions). Agency is significantly impacted by perspective, as is overall embodiment [9].

4.5 Embodiment

The embodiment questionnaire revealed significant differences for perspective, with higher agency scores in the 1stPP condition compared to 3rdPP ($F(1, 298) = 13.10$, $p = 0.001$; 1stPP = 3.98, 3rdPP = 3.66). Similarly, the total embodiment scale showed significantly higher scores for 1stPP (see Figure 8 and Table 1).

4.6 Qualitative Findings

Our qualitative data was transcribed and analyzed following the thematic analysis approach by Blandford et al. [5]. First, four transcriptions were open-coded and aligned to one codebook. Afterward, all interviews were coded and categorized, resulting in four themes derived from the participants' feedback on VR for movement training. The results highlight both strengths and areas for improvement in the current VR setup.

4.6.1 General Reception of VR for Movement Training. The use of VR for movement training was generally well-received, with 25 participants expressing positive feelings about the concept. Many described the system as "*interesting*" and "*effective*." For example, P22 found VR more engaging than real-life scenarios, stating it felt "*more* *fun*" and "*I didn't feel as tired as I think I would be eal life or as bored.*" Several participants (6) highlighted specific advantages of VR. P15 appreciated the ability to "*observe your body from different angles,*" which could be beneficial for those who are "*a little bit shy*" (P1) or prefer not to socialize (P2). VR was also seen as a viable option for those unable to attend real classes, and P18 suggested it could be particularly useful for beginners, people with limited mobility, and potentially elderly individuals. However, some participants (5) expressed reservations. P24 noted that the effectiveness of VR "*depends on how long the training session lasts*" and requires robust feedback. P10 pointed out the challenges of using VR for martial arts, which demand high precision, while P5 found the headset cumbersome and lacking a clear advantage unless gamified, stating, "*It is boring just as it is.*"

4.6.2 Situating VR: Comparison with Traditional Training Methods. Despite the positive feedback, when asked to choose, a majority of participants (20) indicated a preference for real classes over VR or TV applications. The primary reason cited was the desire for "*professional feedback*" (7 participants) from a human instructor. The "*social aspect*" of real classes was also a valued factor for six participants. Conversely, five participants favored the VR application because it lacks the social pressures of a real class, which could make them feel "*nervous*" or "*embarrassed*" (P11, P13). These participants appreciated the privacy offered by VR. P16 was particularly enthusiastic, stating, "*I would prefer the VR system. I think I cannot have this environment to practice using a TV and a real class. I think VR covered everything that a real class can do.*" A few participants remained undecided, suggesting a preference for "*a combination of all*" methods (P7) or noting that it "*depends on the sport*" (P1).

4.6.3 Condition-Specific Feedback: Perspectives and Visualization. Feedback on the instructional videos was predominantly positive, with 13 participants praising their helpfulness and features like looping the video or the split screen. P13 commented: "*The video instructions were very helpful. They helped me to learn the movements piece by piece.*" Criticism was minor, focusing on suggestions like adjustable playback speed.

Regarding the four training conditions, preferences varied significantly, revealing key insights into the user experience.

The 1stPP-M condition received the most positive feedback (14 participants), who appreciated its familiarity and "natural" feel.

The 1stPP-A condition garnered positive responses from nine participants who found the external avatars helpful for whole-body visualization, though some reported initial confusion with the non-mirrored movements. The 3rdPP conditions received mixed feedback. Six participants valued the comprehensive, objective view it provided, while others described feeling disconnected from their body. The combination of a 3rdPP and mirrors was often described as "*confusing*" (P6) or "*strange*" (P19).

4.6.4 Pathways to Improvement: User-Suggested Enhancements. Participants provided extensive constructive feedback. Fourteen commented on the technical setup, highlighting common issues such as headset weight, restrictive cables, and tracking inaccuracies. P19 emphasized the need for a "*standalone headset*" to enable more freedom of movement. Furthermore, twenty participants offered concrete suggestions to enhance the VR training experience. These can be grouped into two main categories:

Performance Feedback Mechanisms: There was a strong desire for real-time, actionable feedback. This included scoring systems (P2, P4), metrics on movement speed (P11), and visual error identification, such as highlighting an incorrectly positioned limb (P6).

Advanced Instructional and Motivational Features: Participants suggested more direct guidance tools, such as floor projections for foot placement (P3), a "ghost" of the expert's ideal movement to follow (P12), or a "*virtual sensei*" (P15). Others expressed interest in multiplayer functionality with customizable avatars (P6), gamification elements (P5), and applications for other sports (P22).

5 Discussion

In this section, we integrate our quantitative and qualitative data to discuss our findings, structured around our three research questions: the impact of visual conditions on movement accuracy and visual attention (RQ1), their effects on the subjective experience (RQ2), and finally, user preferences and the resulting design implications (RQ3).

5.1 Performance and Focus Depend on the Visual Condition.

In addressing RQ1, our study found that no single condition was universally superior for overall movement accuracy. Positional deviation from the expert, perceived cognitive workload (NASA-TLX), and physical exertion (ECG) were comparable across all conditions. This is a crucial finding: it demonstrates that alternative setups, such as using a third-person perspective or external avatars, are viable training options that can be introduced to provide different feedback without impairing general performance or imposing a higher cognitive load than the other tested conditions.

However, we identified a critical exception for rotational movements. The significantly higher rotational deviation in the 3rdPP-A condition is a key result. Our qualitative data provides an explanation, as participants reported losing visual contact with their self-avatar during turns, with one describing the combination of 3rdPP and mirrors as “confusing” (P6). We theorize the 3rdPP-A setup created a visual discontinuity during turns, unlike the constant reference provided by mirrors, leading to disorientation. This aligns with findings from Medeiros et al. [36], who also noted performance issues in 3rdPP when the connection to the avatar is weakened [36]. While our workload metrics showed no significant differences, the non-mirrored view in the 3rdPP-A setup may have introduced a mental mapping challenge that affected rotational accuracy. Comparing these two feedback types is a clear direction for future work.

Eye-tracking data further illuminated differing behavioral strategies. Participants in 3rdPP conditions looked at their self-avatar more frequently, suggesting this perspective effectively promotes body awareness, a potential advantage for novices learning correct posture. In contrast, participants in 1stPP conditions rotated their viewpoint between the front and back screens less often, indicating they remained more grounded and embodied within their virtual body to guide their movements.

5.2 Perspective Shapes Subjective Experience.

Addressing RQ2, our findings on subjective experience highlight a distinct trade-off between perspectives. In line with previous work [4], the 1stPP yielded a significantly stronger sense of agency and higher overall embodiment scores. This was confirmed by our interviews, where participants described the 1stPP-M condition as the most "*natural*" (P1) and feeling "*most like it was me*" (P10). This sense of direct control and body ownership is often considered crucial for motivation and skill internalization.

Interestingly, while mirrors are frequently cited as tools to enhance embodiment [4, 12, 39], we found no significant main effect of the VSVT on our embodiment scores, and despite the clear impact of perspective on embodiment, we observed no corresponding effect on the sense of presence (IPQ). This surprising finding might suggest that in a goal-oriented training task, the high general immersion of VR and the intense focus on the movements can overshadow the specific influence of a visual condition on presence.

5.3 Split Preferences Require Flexible Systems.

Our qualitative findings provide a clear answer to RQ3: there is no one-size-fits-all solution. User preferences were split and reflected the trade-offs identified in our other results. The 1stPP-M condition was most popular (14 users), valued for its familiarity and the immersive sense of control. This reveals a notable discrepancy between objective user performance and subjective preference, suggesting that comfort and intuitiveness is a key driver for user acceptance. However, other users saw clear benefits in alternative setups. The 3rdPP was praised for its objective viewpoint for posture correction and its engaging, "*game-like*" quality (P21). External avatars, while sometimes confusing, were appreciated by some for providing a non-mirrored, 360-degree view.

This diversity in feedback strongly advocates for our primary design recommendation: the **development of flexible and adaptive VR training systems.** A system that forces a single paradigm will likely fail to meet the needs of all users. Instead, future applications should empower learners by **allowing dynamic perspective switching, offering a choice of VSVTs, and integrating the interactive feedback mechanisms** which participants strongly desired (e.g., scoring, error highlighting, virtual sensei). This call for more direct guidance aligns with a broader trend in technology-enhanced training, which moves towards real-time, corrective feedback such as skeleton overlays or expert silhouettes [13, 24]. By creating systems that can strike a "*balance between play and training*" (P14), we can better support diverse learning styles and unlock the full potential of VR for skill acquisition. A promising future direction is to analyse the learning curve during training to provide real-time interventions, an approach that was beyond the scope of our post-training evaluation.

5.4 Limitations and Future Work

Our study's limitations inform future research. We chose the 1stPP-M condition as a natural valid baseline, mirroring common real-world training setups. A limitation of this approach is that we did not include a control condition consisting of only one factor (i.e., a 1stPPview without any VSVT). Consequently, our design does not allow us to fully isolate the specific effects of adding a virtual mirror or avatars on performance and cognitive load compared to a complete absence of self-observation. Another limitation is the static pairing of karate sequences to conditions. Although an expert deemed them to be of comparable difficulty, we cannot entirely rule out that inherent movement differences acted as a confounding variable. For instance, the higher rotational deviation observed in the 3rdPP-A condition might be partially attributable to the specific Kick-Punch-BlockUp sequence used, rather than solely to the visual setup.

Future work should address these points by incorporating a "no-VSVT" baseline and counterbalancing or standardizing movement sequences. Furthermore, while VR hardware is advancing, the reliability and ease of setup for full-body tracking remain a challenge. Looking ahead, we plan to conduct longitudinal studies to explore how perception and performance evolve with prolonged use. We also aim to enhance the system's interactivity, enabling users to control perspectives and feedback as they suggested. Finally, we intend to adapt our findings for new domains like yoga, dance, or rehabilitation, which will require tailoring feedback to the unique demands of each application.

6 Conclusion

This paper presented a comprehensive evaluation of how visual perspective (1stPP vs. 3rdPP) and VSVTs (mirrors vs. external avatars) impact VR movement training. Our mixed-methods user study reveals that there is no single optimal configuration for performance. Instead, we uncovered a crucial trade-off: the 1stPP enhances the sense of agency and immersion, while the 3rdPP perspective offers valuable opportunities for self-correction by increasing body awareness. Crucially, our results indicate these benefits can be achieved without imposing a significantly higher cognitive load when compared to the familiar 1stPP-M setup.

A key finding was the significant performance decrease in rotational accuracy in the 3rdPP-A condition, likely caused by users losing visual contact with their avatar, which is a critical insight for designers of training systems. While most users preferred the familiar 1stPP-M setup, the varied feedback underscores our primary contribution: a one-size-fits-all approach is suboptimal for VR movement training.

Our findings strongly support for the development of flexible and adaptive VR training systems. We offer practical design recommendations, such as enabling dynamic perspective switching and providing choices of self-observation, to create more effective and user-centered learning experiences. The future of VR training lies not in finding one perfect view, but in creating intelligent systems that can be tailored to individual needs, tasks, and preferences, ultimately maximizing the potential of VR for skill acquisition.

Acknowledgments

Francesco Chiossi was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within "AI Motive: Multimodal Intent Communication of Autonomous Systems" (SCHM 1751/13-1) and Project ID 251654672 TRR 161. The authors acknowledge the use of Gemini 2.5 Pro for language proofreading and polishing of this manuscript. All ideas and content remain the authors'.

[36] Daniel Medeiros, Rafael K. Dos Anjos, Daniel Mendes, João Madeiras Pereira, Alberto Raposo, and Joaquim Jorge. 2018. Keep My Head on My Shoulders!: Why Third-Person Is Bad for Navigation in VR. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, Tokyo Japan, 1–10. doi:10.1145/3281505.3281511

[37] Nornadiah Mohd Razali and Bee Yap. 2011. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics* 2 (Jan. 2011).

[38] Aske Mottelson, Andreea Muresan, Kasper Hornbæk, and Guido Makransky. 2023. A Systematic Review and Meta-Analysis of the Effectiveness of Body Ownership Illusions in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 30, 5 (Sept. 2023), 76:1–76:42. doi:10.1145/3590767

[39] Sofia Adelaide Osimo, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater. 2015. Conversations between Self and Self as Sigmund Freud—A Virtual Body Ownership Paradigm for Self Counselling. *Scientific Reports* 5, 1 (Sept. 2015), 13899. doi:10.1038/srep13899

[40] Stefan Pastel, K. Petri, C. H. Chen, Ana Milena Wiegand Cáceres, M. Stirnatis, C. Nübel, L. Schlotter, and K. Witte. 2023. Training in Virtual Reality Enables Learning of a Complex Sports Movement. *Virtual Reality* 27, 2 (June 2023), 523–540. doi:10.1007/s10055-022-00679-7

[41] pmonasteriop. 2021. Dojo Matrix Drunken Wrestlers. https://www.fab.com/listings/840f0363-d436-4fd0-8fa9-71bebc2ef4f2 Accessed: 2025-09-15.

[42] Aniruddha Prithul, Isayas Berhe Adhanom, and Eelke Folmer. 2021. Embodied Third-Person Virtual Locomotion Using a Single Depth Camera. In *Proceedings of Graphics Interface 2021 (Gi 2021)*. Canadian Information Processing Society, 210–219. doi:10.20380/GI2021.24

[43] M Richter, GHE Gendolla, and RA Wright. 2016. Three Decades of Research on Motivational Intensity Theory: What We Have Learned about Effort and What We Still Don't Know. *Advances in motivation science* 3 (2016), 149–186.

[44] RootMotion. 2025. Final IK toolkit. https://assetstore.unity.com/packages/tools/animation/final-ik-14290 Accessed: 2025-09-15.

[45] Patrick Salamin, Tej Tadi, Olaf Blanke, Frederic Vexo, and Daniel Thalmann. 2010. Quantifying Effects of Exposure to the Third and First-Person Perspectives in Virtual-Reality-Based Training. *Learning Technologies, IEEE Transactions on* 3 (Oct. 2010), 272–276. doi:10.1109/TLT.2010.13

[46] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (June 2001), 266–281. doi:10.1162/105474601300343603

[47] Mohammad Shokoohi-Yekta, Bing Hu, Hongxia Jin, Jun Wang, and Eamonn Keogh. 2017. Generalizing DTW to the Multi-Dimensional Case Requires an Adaptive Approach. *Data Mining and Knowledge Discovery* 31, 1 (Jan. 2017), 1–31. doi:10.1007/s10618-016-0455-0

[48] James Smith, Xinyun Cao, Adolfo G. Ramirez-Aristizabal, and Bjoern Hartmann. 2023. Dual Body Bimanual Coordination in Immersive Environments. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)*. Association for Computing Machinery, New York, NY, USA, 230–243. doi:10.1145/3563657.3596082

[49] Siqi Sun, Yongqing Sun, Mitsuhiro Goto, Shigekuni Kondo, Dan Mikami, and Susumu Yamamoto. 2022. Motor Learning Based on Presentation of a Tentative Goal. In *Proceedings of the 2022 International Conference on Multimedia Retrieval (ICMR '22)*. Association for Computing Machinery, New York, NY, USA, 601–607. doi:10.1145/3512527.3531413

[50] Thomas Waltemate. 2018. *Creating a Virtual Mirror for Motor Learning in Virtual Reality.* Ph. D. Dissertation. Universität Bielefeld. https://pub.uni-bielefeld.de/record/2932705 Accessed: 2025-09-15.

[51] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1643–1652. doi:10.1109/TVCG.2018.2794629

[52] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The Impact of Latency on Perceptual Judgments and Motor Performance in Closed-Loop Interaction in Virtual Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (VRST '16)*. Association for Computing Machinery, New York, NY, USA, 27–35. doi:10.1145/2993369.2993381

[53] Kerstin Witte, Dan Bürger, and Stefan Pastel. 2025. Sports Training in Virtual Reality with a Focus on Visual Perception: A Systematic Review. *Frontiers in Sports and Active Living* 7 (March 2025). doi:10.3389/fspor.2025.1530948

[54] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[55] Biao Xie, Huimin Liu, Rawan Alghofaili, Yongqi Zhang, Yeling Jiang, Flavio Destri Lobo, Changyang Li, Wanwan Li, Haikun Huang, Mesut Akdere, Christos Mousas, and Lap-Fai Yu. 2021. A Review on Virtual Reality Skill Training Applications. *Frontiers in Virtual Reality* 2 (April 2021). doi:10.3389/frvir.2021.645153

[56] Wenge Xu, Hai-Ning Liang, Qiuyu He, Xiang Li, Kangyou Yu, and Yuzheng Chen. 2020. Results and Guidelines From a Repeated-Measures Design Experiment Comparing Standing and Seated Full-Body Gesture-Based Immersive Virtual Reality Exergames: Within-Subjects Evaluation. *JMIR serious games* 8, 3 (July 2020), e17972. doi:10.2196/17972

[57] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. 2009. The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research* 36, 2 (April 2009), 285–312. doi:10.1177/0093650208330254

[58] Kangyou Yu, Shaoyue Wen, Wenge Xu, Maurizio Caon, Nilufar Baghaei, and Hai-Ning Liang. 2023. Cheer for Me: Effect of Non-Player Character Audience Feedback on Older Adult Users of Virtual Reality Exergames. *Virtual Reality* 27, 3 (Sept. 2023), 1887–1903. doi:10.1007/s10055-023-00780-5

[59] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective Matters: Design Implications for Motion Guidance in Mixed Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 577–587. doi:10.1109/ISMAR50242.2020.00085

[60] Maryam Zahabi and Ashiq Mohammed Abdul Razak. 2020. Adaptive Virtual Reality-Based Training: A Systematic Literature Review and Framework. *Virtual Reality* 24, 4 (Dec. 2020), 725–752. doi:10.1007/s10055-020-00434-w

[61] Yang Zou, Yanguang Wan, Yonghao Zhang, Xiao Zhou, Jian Cheng, Cuixia Ma, Chun Yu, Xiaoming Deng, and Hongan Wang. 2025. Gesture Builder: Flexible Gesture Customization and Efficient Recognition on VR Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 2 (June 2025), 67:1–67:37. doi:10.1145/3729484