CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/190449/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Ghanbari, Milad, Huang, Yujie, Zhou, Wei, Le Callet, Patrick, Timmerer, Christian and Amirpour, Hadi 2026. QoE and depth perception in spatial augmented reality darts. Presented at: MM '26: The 34th ACM International Conference on Multimedia, Rio de Janeiro, Brazil, 10-14 November 2026. Published in: Li, J., Gao, X., Le Callet, P., Janowski, L., Lu, W., Yang, J., Wang, J. and Zhang, J. eds. QoEVMA '26: Proceedings of the 4th International ACM Workshop on Quality of Experience in Visual Multimedia Applications. IEEE, pp. 1-9. 10.1145/3840476.3840508

Publishers page: https://doi.org/10.1145/3840476.3840508

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

QoE and Depth Perception in Spatial Augmented Reality Darts

Milad Ghanbari
Christian Doppler Laboratory ATHENA
Institute of Information Technology
University of Klagenfurt
Klagenfurt, Austria
milad.ghanbari@aau.at

Yujie Huang
Nantes Université
École Centrale Nantes
Nantes, France
yujie.huang@etu.univ-nantes.fr

Wei Zhou
Cardiff University
Cardiff, UK
zhouW26@cardiff.ac.uk

Patrick Le Callet
Nantes Université
École Centrale Nantes
Nantes, France
patrick.le-callet@univ-nantes.fr

Christian Timmerer
Christian Doppler Laboratory ATHENA
Institute of Information Technology
University of Klagenfurt
Klagenfurt, Austria
christian.timmerer@aau.at

Hadi Amirpour
Christian Doppler Laboratory ATHENA
Institute of Information Technology
University of Klagenfurt
Klagenfurt, Austria
hadi.amirpour@aau.at

Abstract

Accurate depth perception is critical for precision-based interaction in augmented reality applications, particularly in tasks that rely on ballistic motion such as throwing. While recent spatial computing devices support high-fidelity passthrough rendering and controller-free hand tracking, empirical evidence on how users perceive and act upon virtual depth in such environments remains limited. This paper investigates the effect of virtual target distance and time pressure on user performance and quality of experience in a physics-based augmented reality dart-throwing task implemented on the Apple Vision Pro. A controlled, repeated-measures user study with 22 participants evaluated performance at three target distances (near, standard, and far) and under a time-limited condition. Objective measures included score, hit accuracy, and throw timing, while subjective measures captured perceived depth, immersion, control, enjoyment, comfort, progress, time pressure, and adaptation. Results show a clear decrease in throwing accuracy as target distance increased, indicating limits in actionable depth perception despite visually stable spatial rendering. Under time pressure, participants significantly increased throwing speed without a corresponding loss in accuracy, suggesting robust motor execution once the interaction model was learned. Subjective ratings of control and depth understanding remained high across conditions, even when objective performance varied. These findings provide quantitative evidence that modern passthrough augmented reality systems can support reliable controller-free precision interaction, while also clarifying how depth and cognitive demand shape user performance and experience.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; **User studies**.

Keywords

Quality of Experience, Depth Perception, Spatial Computing, Augmented Reality, Hand-Tracking, Interactivity

ACM Reference Format:
Milad Ghanbari, Yujie Huang, Wei Zhou, Patrick Le Callet, Christian Timmerer, and Hadi Amirpour. 2026. QoE and Depth Perception in Spatial Augmented Reality Darts. In *4th Workshop on Quality of Experience (QoE) in Visual Multimedia Applications (QoEVMA '26), November 10–14, 2026, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3840476.3840508

1 Introduction

Recent advances in spatial computing have enabled augmented reality (AR) systems that combine high-fidelity passthrough rendering with controller-free hand tracking [14, 20, 38]. These capabilities make it possible to place interactive virtual objects into the user's physical environment with a level of visual stability and spatial coherence that was difficult to achieve with earlier consumer devices [2, 21, 32]. As a result, AR is increasingly being used for applications that require direct physical interaction, including games, training systems, and skill-based simulations [9, 10, 25]. Despite these technical improvements, the human factors that determine performance and experience in such interactions remain insufficiently quantified in precision-based spatial interaction tasks [36].

One of the central challenges in interactive AR is depth perception [7, 27]. Precision-based tasks often depend on the user's ability to accurately judge distance and translate that perception into a well-controlled motor action [3]. Even small perceptual errors can lead to large performance differences when tasks involve ballistic motion, such as throwing [11]. While depth perception has been extensively studied in virtual reality [19, 23], fewer empirical studies have examined how users perceive and act upon virtual depth in high-quality passthrough AR environments, particularly when interaction is performed without physical controllers [8, 29].

Controller-free hand tracking aims to increase the naturalness of the interaction by allowing users to act directly with their hands [28, 33]. However, removing physical controllers also removes haptic feedback and a stable proprioceptive reference, which can affect motor accuracy and the sense of control [5, 37]. For dynamic tasks that require precise timing and force estimation, such as throwing a dart, it remains unclear whether users can

maintain precision and consistency when all feedback is provided through visual cues alone [13, 26]. This uncertainty is especially relevant for modern AR head-mounted displays, where visual realism may create high user expectations regarding physical correctness and responsiveness.

In addition to perceptual accuracy, cognitive factors, such as time pressure, influence motor behavior and Quality of Experience (QoE) [14, 30]. Increased time constraints are commonly associated with speed-accuracy trade-offs, where faster execution leads to reduced precision. In throwing tasks, time pressure may reduce aiming calibration time, potentially amplifying depth-related errors [1]. Understanding whether such trade-offs occur in controller-free AR interaction is important for the design of interactive experiences that combine challenge, engagement, and performance stability [12, 34].

This paper examines these issues through a controlled user study using a physics-based AR dart-throwing task on the Apple Vision Pro (AVP). Specifically, we examine how virtual target distance influences motor precision, temporal execution, and inter-participant consistency in controller-free AR, and whether time pressure induces a measurable speed–accuracy trade-off. Objective performance measures are analyzed alongside subjective ratings to provide a comprehensive view of interaction quality [16, 17]. By focusing on a precision-based task in a high-fidelity passthrough AR environment, this work aims to provide empirical evidence that informs the design and evaluation of controller-free spatial interactions [4].

2 Related Work

Depth perception has been a long-standing topic in immersive systems research, as errors in distance judgment can directly affect task performance [7]. Studies in virtual reality have consistently shown that users tend to underestimate egocentric distances, even when stereoscopic rendering and motion parallax are available [22]. These perceptual distortions have been linked to display characteristics such as field of view, rendering scale, and latency [35]. While similar issues are expected in AR, fewer empirical studies have quantified how depth perception influences motor performance in high-quality passthrough AR, where real-world visual context is preserved, but virtual objects must still be interpreted in depth [8].

Input modality plays a critical role in spatial interaction [3]. Hand-held controllers provide stable haptic feedback and a reliable proprioceptive reference, which has been shown to support precise manipulation and selection tasks [24]. In contrast, controller-free hand tracking offers a more direct and intuitive interaction metaphor but lacks physical feedback [28, 29]. Prior comparative studies indicate that controller-free input is often preferred for natural, direct manipulation tasks, while controllers remain advantageous for precision-oriented or abstract interactions [31]. However, most existing work focuses on selection and manipulation tasks rather than dynamic, ballistic actions such as throwing, where timing, velocity estimation, and release control are critical [11].

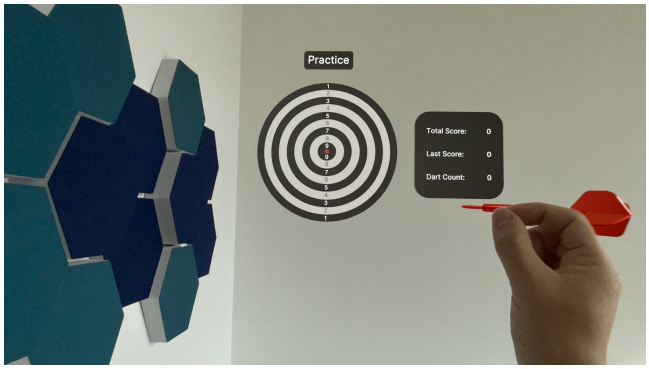


Figure 1: Captured screenshot of the SDART user interface, visualizing a dartboard, experiment condition on top, scores panel on the right, and the user holding a dart pin.

Physics-based interaction has been identified as a key factor for maintaining plausibility and user engagement in immersive environments [15]. Rather than scripting specific behaviors, physics-driven simulations allow object motion to emerge from general physical rules, which can support more flexible interaction [13]. Research has also shown that users adapt their motor behavior to altered virtual physics, developing internal models that compensate for perceptual or physical inconsistencies [9, 25]. While such adaptation can improve short-term performance in virtual environments, it may also mask underlying perceptual errors, making objective evaluation of depth-related performance particularly important [12].

QoE in immersive systems is typically assessed using a combination of objective performance measures and subjective self-reports [16, 19]. Objective metrics such as accuracy, score, and task completion time provide insight into effectiveness and efficiency, while subjective measures capture perceived depth, immersion, control, enjoyment, comfort, progress, time pressure, and adaptation [18]. Previous research has reported that objective performance and subjective experience do not always align, highlighting the need for mixed-methods evaluation [12]. In the context of controller-free AR interaction, understanding how subjective confidence relates to measurable performance is particularly relevant [10].

The reviewed literature indicates a lack of empirical studies that jointly examine depth perception, physics-based interaction, and QoE in controller-free passthrough AR [8]. In particular, there is limited quantitative evidence on how users perform precision throwing tasks at different virtual distances and under varying cognitive demands [11]. The present work addresses this gap by providing a controlled evaluation of depth-dependent performance and subjective experience in a physics-based AR dart-throwing task.

3 System Overview

The system was implemented on the AVP, a standalone, high-resolution passthrough AR head-mounted display (HMD) that supports real-time environment capture and spatial tracking through its visionOS (v26.4). The device provides stereoscopic

color passthrough with low latency, allowing virtual content to be visually aligned with the physical environment [6]. Head and hand motion were continuously tracked, enabling six-degrees-of-freedom interaction without physical controllers. The application was developed using the Unity 6 game engine (v6000.0.23f1) and the platform's native spatial computing framework. This setup allows virtual objects to be placed in stable world coordinates and rendered with consistent scale and depth relative to the physical environment. All rendering and interaction logic were executed locally on the device. The starting scene, including the dartboard, user-interface panels, and a dart pin held by the user, is shown in Figure 1.

The dart-throwing task was designed to rely exclusively on controller-free hand tracking. Users grabbed a virtual dart using a pinch gesture and initiated a throw by releasing the gesture. The position and instantaneous velocity of the hand at the moment of release were sampled and used to initialize the dart's motion. This approach mirrors common interaction techniques for throwing in immersive environments while avoiding explicit button-based input. Visual feedback was provided throughout the interaction. The dart, target, and hit location were rendered continuously, and the target remained fixed in world space at a predefined distance from the user. No artificial aids, such as trajectory previews or aim assistance, were included, ensuring that performance depended solely on the user's perception and motor control.

Dart motion was managed by a physics-based simulation rather than predefined animations. At release, the dart was assigned an initial velocity derived from the recent hand movement. Standard rigid-body physics were applied to update position and orientation over time. Collision detection was used to determine impact with the dartboard, and scores were computed based on the final hit location. Using a physics-driven model allowed throwing behavior to emerge from the user's motion and supported natural variation across throws. This design choice also enabled the analysis of learning and adaptation effects, as users adjusted their motor behavior to the system's response.

All interactions were logged automatically during the experiment. For each throw, the system recorded the resulting score, the two-dimensional hit position on the dartboard, and the time between consecutive throws. These data were stored for offline analysis and used to compute performance metrics across conditions. In addition to objective performance data, subjective responses were collected separately after each experimental condition using standardized questionnaires. This separation reduced interference between task execution and self-reporting while allowing objective and subjective measures to be analyzed jointly.

4 Experimental Design

Twenty-two participants were recruited for the study. The sample included individuals with varying levels of experience with augmented and virtual reality systems, as well as different degrees of familiarity with real-world dart throwing; both were collected via a brief self-report during onboarding (novice, occasional, or regular use/play) but were not used to stratify recruitment or balance conditions. All participants reported normal or corrected-to-normal vision. Prior to the experiment, participants provided informed consent and were introduced to the general procedure.

This study employed a within-subjects design with two main factors: target distance and time pressure. Target distance referred to the distance between the participant and a dartboard positioned at a fixed standard height (1.73 m) directly in front of the participant. Time pressure was manipulated by allocating limited (60 s) or unlimited time. Four experimental conditions were designed to examine the effects of these two factors, as shown in Table 1. Under the unlimited-time condition, three target distances were tested: near (1.87 m), standard (2.37 m), and far (2.87 m), with participants instructed to throw 15, 40, and 15 darts, respectively. Under the 60 s limited-time condition, participants were positioned at the standard distance and asked to throw as many darts as possible within the allotted time while maintaining accuracy. To avoid learning effects, the order of using the four conditions was counterbalanced across participants, meaning that half of the users started from near distance to far distance, and the other half in reverse order, while the time-pressure mode was consistently the last stage. However, for data analysis, the results of all users were re-ordered by target distance as **Stage 1: Near**, **Stage 2: Standard**, and **Stage 3: Far**, so that stage comparisons reflect distance rather than each participant's original presentation order.

At the beginning of the session, participants completed a short practice phase to become familiar with the interaction technique and the physics model. During each experimental condition, participants performed a series of throws while standing at a fixed position. Short breaks were provided between conditions to reduce fatigue. The entire session lasted approximately between 5 min to 15 min for all participants, and no performance feedback beyond the in-task scoring was provided between conditions. Several objective measures were recorded to assess the performance, including the score achieved for each throw, the spatial distribution of hit locations on the dartboard, and the time between consecutive throws. These measures were used to analyze accuracy, consistency, and efficiency.

We note two design choices that constrain the interpretation of the results and that we discuss further in Section 5.9. First, the standard-distance condition was assigned 40 throws, more than double the 15 throws used for the near and far conditions, because it also served as the baseline for the intra-stage (early vs. late trials) and time-pressure comparisons; this asymmetry means that reported differences between the standard condition and the near/far conditions are potentially confounded with unequal practice at each distance rather than distance alone. Second, the standard distance was always administered as the second stage relative to near and far (whose order was counterbalanced), and the time-limited condition was always administered last so that participants had already learned the interaction model before performing under time pressure; consequently, the standard-vs. time-limited comparison in Section 5 cannot fully separate the effect of time pressure from order effects such as continued learning or fatigue.

After completing each condition, participants filled out a questionnaire assessing their subjective experience. Items were custom-designed for this study, following common Likert-scale QoE questionnaire practice in immersive systems research [4, 16, 17], and rated on a 5-point scale from *strongly disagree* to *strongly agree*; each of the nine constructs below was assessed with two items. For

Table 1: Four experimental conditions used in this study.

Condition	Target distance	Time constraint	Number of throws
1	Near	Unlimited	15
2	Standard	Unlimited	40
3	Far	Unlimited	15
4	Standard	Limited (60 s)	As many as possible

example, the two **comfort** items were phrased as "I did not experience physical strain or fatigue while playing" and "The gesture-based interaction felt intuitive and effortless." The evaluation highlights the primary aspects of spatial gaming, including *(1)* **depth perception**: the ability to judge distances and the realism of object positioning in space; *(2)* **immersion**: the focus on the task and the sensation of physically throwing darts; *(3)* **self-perception**: the perceived control over body movement and the translation of personal skills into the simulation; *(4)* **enjoyment**: the engagement levels and satisfaction derived from hitting targets; *(5)* **comfort**: the lack of physical strain and the intuitiveness of the interaction model; *(6)* **comparison**: the perceived viability of the simulation as a practical alternative or training tool compared to real-world darts; *(7)* **progress**: the awareness of performance improvement through the stages; *(8)* **timer**: the perceived pressure and performance impact of the temporal constraints; and *(9)* **adaptation**: the ability to learn repeatable aiming routines and adjust to game physics.

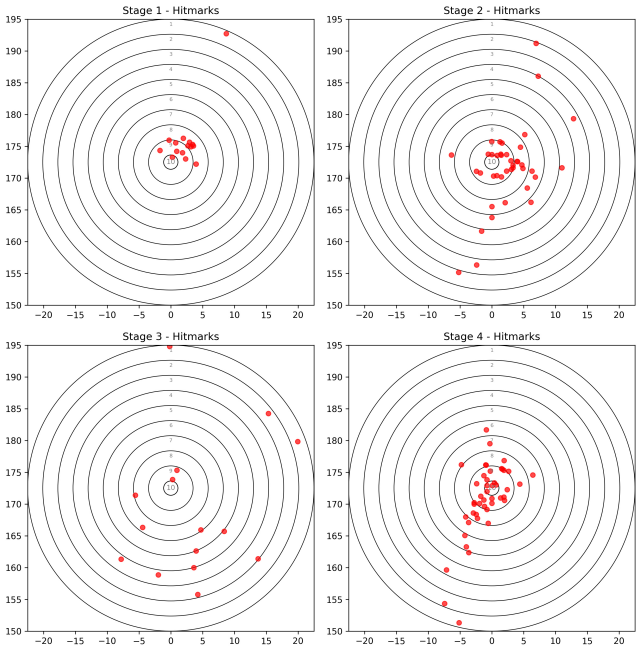


Figure 2: A sample of the spatial distribution of dart hit locations on the dartboard for each condition. Stage 1: Near, Stage 2: Standard, Stage 3: Far, Stage 4: Standard (limited-time).

5 Results

This section provides the results across various dimensions as follows: reporting spatial hit distributions and distance-dependent accuracy, followed by performance, consistency, and temporal metrics across stages (including standard vs. time-limited comparisons and intra-stage progression), and concluding with subjective results and regression analyses linking subjective measures to objective outcomes.

5.1 Spatial Distribution of Hit Locations

The distance to the target affected the distribution of dart impacts on the dartboard. With increasing distance, the dart hits were more dispersed and less clustered around the center, reflecting reduced throwing precision at farther ranges. Figure 2 depicts an example of the spatial distribution of dart hit locations on the dartboard for each condition recorded from one user. The near condition shows tighter clustering around the center, whereas the far condition exhibits a wider dispersion. This pattern suggests that increasing virtual distance imposes greater perceptual and motor demands. While users maintained general spatial awareness, fine-grained ballistic precision decreased as target distance increased. As will be shown in the following sections, this pattern was observed consistently across all participants, indicating difficulty in distance-dependent motor calibration.

This increased dispersion at greater distances is consistent with depth-perception limitations, but it is also the pattern one would expect from real-world ballistic throwing, where the same angular or velocity error at release translates into a proportionally larger positional error on the target as distance increases, independent of any perceptual misjudgment. Our design cannot fully disentangle these two explanations, since we did not manipulate release-error magnitude independently of distance or include a condition that isolates motor-execution noise from depth misjudgment (e.g., a task with equivalent target angular size at every distance). We consider this an important direction for future work and discuss it further as a limitation in Section 5.9.

5.2 Performance Across Distances

Throwing performance varied systematically with the target distance and was evaluated based on the score achieved for each throw. A significant overall difference in performance was found and was consistent across participants (Friedman test: χ^2=27.20, p<0.001, Kendall's W=0.62), indicating a robust effect of virtual distance on precision performance. A paired Wilcoxon signed-rank test revealed a statistically significant decrease in performance from Stage 1 (Near) to Stage 2 (Standard) (W=21.0, p<0.001). A second Wilcoxon test between Stage 2 (Standard) and Stage 3 (Far) also showed a significant performance decline (W=15.0, p<0.001), indicating a consistent reduction in throwing accuracy as target distance increased.

Figure 3 shows the dart throwing performance across experimental stages grouped by distance, with the corresponding numerical values presented directly in the figure. The mean score was highest in the near-distance condition and decreased progressively in the standard and far-distance conditions. We also observed that

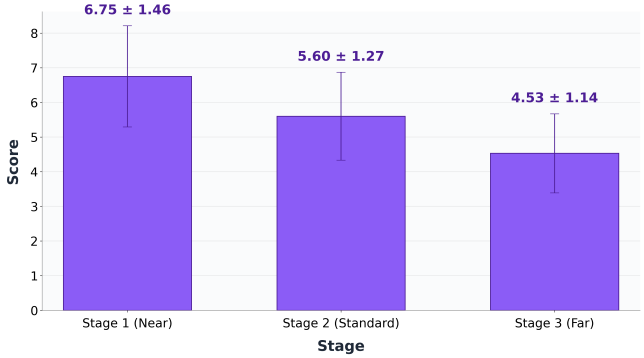


Figure 3: Dart throwing performance across experimental stages, grouped by target distance. Only conditions under unlimited time were used.

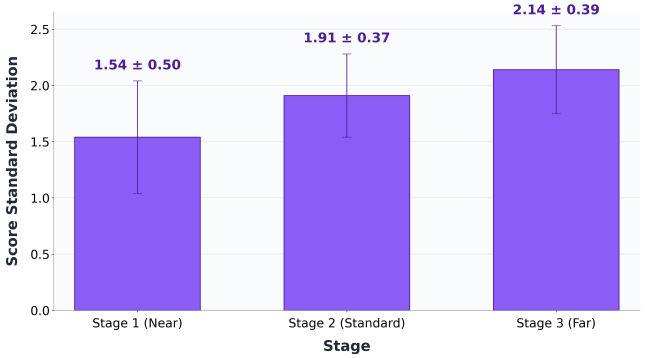


Figure 4: Mean of inter-participant standard deviations of scores across experimental stages. Higher values mean lower throwing consistency.

throwing time increased with target distance across the three conditions (see Section 5.4). This increase suggests that participants required more time to execute throws at greater distances. However, despite this temporal adjustment, throwing accuracy still decreased with distance, suggesting that increased execution time alone did not compensate for the higher task demands at greater distances.

5.3 Inter-Participant Consistency

Inter-participant consistency refers to the score variation among participants. It is represented by the standard deviation of the scores achieved by each subject. A lower standard deviation means higher performance consistency, since the data points are closer to the mean value. Figure 4 shows the inter-participant consistency across experimental stages and the concrete consistency metric values. With increasing target distance, the consistency metric (i.e., the standard deviation) increased, indicating that consistency itself decreased and that dart throws became less repeatable as perceived depth increased. This reduced consistency at larger distances suggests that participants were less able to reproduce similar motor patterns reliably, further supporting the interpretation that depth uncertainty affected both accuracy and motor stability.

A Friedman test was conducted to examine differences across Stage 1, Stage 2, and Stage 3. The test indicated a significant effect of stage on the measured metric ($\chi^2(2)$=9.72, p=0.0078). Post-hoc pairwise comparisons using Wilcoxon signed-rank tests showed a significant change from Stage 1 to Stage 2 (W=45.0, p=9.12×10^{-4}), while the difference between Stage 2 and Stage 3 reached marginal significance (W=89.5, p=0.0839). Moreover, a direct comparison between Stage 1 and Stage 3 shows a significant difference (W=4.5, p=1.287×10^{-5}). Together, these results indicate that performance significantly decreases as throwing distance increases.

5.4 Temporal Metrics Across Stages

Throwing time (also referred to as time per throw) was measured as the interval between the release of one dart and the release of the next, i.e., the time elapsed between two consecutive pinch-release events; it therefore captures the time participants spent

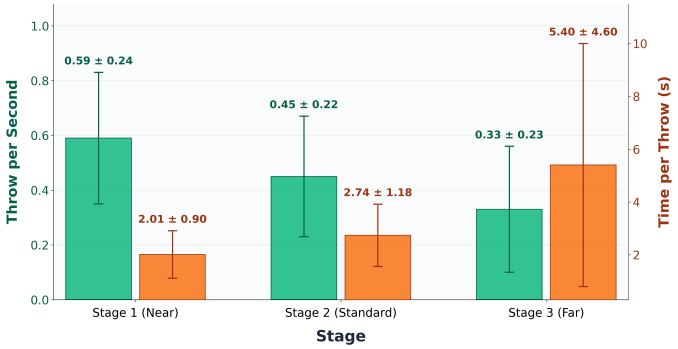


Figure 5: Temporal metrics across experimental stages.

picking up the next dart, aiming, and calibrating their throw, rather than the flight time of the dart itself. A Friedman test was conducted to examine differences in time-per-throw across the three stages (Stage 1, Stage 2, Stage 3). The test revealed a significant effect of stage ($\chi^2(2)$=25.04, p=3.65×10^{-6}), indicating that temporal performance differed across distances. Post-hoc pairwise comparisons using Wilcoxon signed-rank tests (two-sided) showed a significant increase in throw time from Stage 1 to Stage 2 (W=20.0, p=2.21×10^{-5}) and a further significant increase from Stage 2 to Stage 3 (W=46.0, p=1.03×10^{-3}). Together, these results indicate that participants required progressively more time to execute throws as the target distance increased. Figure 5 shows temporal performance metrics including numerical values across stages. Mean throw time increased steadily with distance (2.01 s, 2.74 s, 5.40 s), and variability was largest at Stage 3 (SD=4.60 s), suggesting greater inconsistency in execution timing at the farthest distance.

5.5 Standard vs. Time-Limited Conditions

Under the time-limited condition, participants significantly reduced the time between throws. The Wilcoxon signed-rank test revealed a significant difference in time per throw between Stage 2 and Stage 4 (W=54.0, p=0.0025). Despite this increase in speed, the mean score and consistency remained comparable to the standard

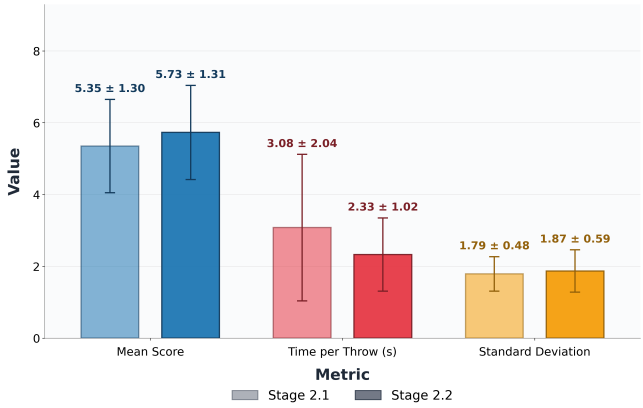


Figure 6: Comparison of performance and timing between standard and time-limited conditions.

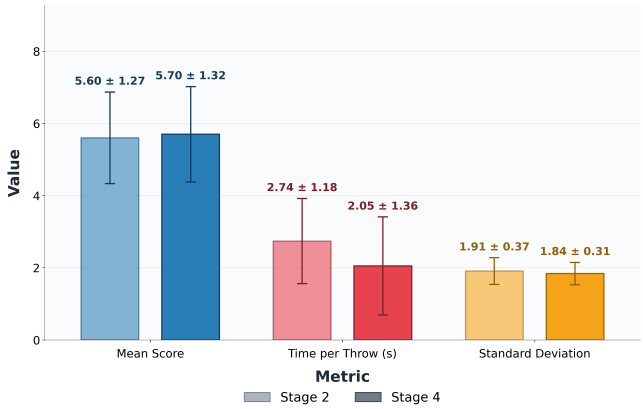


Figure 7: Intra-stage performance progression between early and late trials within the standard-distance stage.

condition. Figure 6 presents a comparison of performance and timing between conditions, with the relevant numerical values. The absence of a speed-accuracy trade-off suggests that, once adapted to the physics-based interaction, participants could execute learned motor patterns more rapidly without loss of precision, which indicates robust motor control under increased cognitive demand.

5.6 Intra-Stage Progression

Performance within the standard-distance stage was further analyzed by comparing early and late trials, meaning the first 15 (i.e., Stage 2.1) and last 15 throws (i.e., Stage 2.2) of the stage. Figure 7 illustrates intra-stage progression, and reports the numerical comparison. While the statistical analysis showed marginally significant difference for throw time (W=95.0, p=0.071) and for mean score (W=56.5, p=0.070), the slight improvement observed in both score and consistency suggests short-term motor adaptation to the interaction and physics model, potentially reflecting participants' growing familiarity with the system's dynamics and their ability to refine throwing techniques over repeated trials.

5.7 Subjective Results

The subjective evaluation (N=22) as visualized in Figure 8 indicates an overall positive reception of the SDART system, particularly regarding visual realism, enjoyment, and comfort. Within the **Depth** items, the statement on realistic object positioning received the highest rating of the questionnaire (M=4.55, SD=0.60), and the ease of judging distances was also rated highly (M=4.23, SD=0.92). **Enjoyment** was strongly endorsed, with a mean of 4.41 (SD=1.05) for gameplay engagement and high satisfaction when hitting the target (M=4.50, SD=1.01). Overall, **comfort** remained high despite the use of a head-mounted display (M=4.23, SD=1.07), while gesture-based interactions were perceived as intuitive and effortless (M=4.05, SD=1.09). Regarding **self-perception**, participants felt in control of their body movements (M=4.09, SD=1.11) and believed their personal skills translated effectively into the game (M=3.91, SD=1.27). **Immersion** items received more moderate ratings, with the feeling of *actually throwing darts* at $M = 3.64$ ($SD = 1.33$) and a focus leading to unawareness of the physical room at M=3.45 (SD=1.37). In terms of **comparison**, the system was viewed as a strong alternative to real darts (M=3.55, SD=0.96), though perceived training practicality was more divided (M=3.50, SD=1.22). Participants reported a clear sense of **progress**, noting performance improvement across stages (M=4.09, SD=1.19) alongside increasing fatigue (M=4.22, SD=1.02). In the **timer** category, perceived pressure during the timed condition was moderate (M=3.27, SD=1.32), with a perceived performance improvement during that stage rated at M=3.50 (SD=1.33). Finally, **adaptation** ratings confirmed that users could adapt to game physics (M=4.04, SD=0.99) and develop a repeatable aiming routine (M=4.18, SD=0.96).

The MOS reported above are aggregated across all experimental stages, since our questionnaire design collected **depth**, **immersion**, **self-perception**, **enjoyment**, and **comfort** ratings once per session rather than separately after every distance condition, which prevents a stage-by-stage statistical comparison of these constructs. The **progress**, **timer**, **comparison**, and **adaptation** constructs were, by design, tied to specific stages (e.g., timer to the time-limited condition) and are reported accordingly above rather than contrasted across all four conditions. Consequently, and notably, we did not directly test whether depth perception or other QoE dimensions themselves varied with target distance, even though objective performance did; we return to this gap in Section 5.9 and consider it a priority for follow-up analysis, since the paper's motivating question concerns precisely how distance shapes both performance and experience.

5.8 Subjective Influence on Objective Measures

Standardized multiple linear regression was applied across distinct experimental stages to determine the influence of nine subjective items on mean score and throw time. Data were transformed into Z-scores prior to analysis, so that regression coefficients (β) represented standardized effect sizes for direct comparison. Results indicated that Adaptation was the most consistent predictor of execution speed, particularly in the near (β=−0.94) and standard (β=−0.83) distance stages. Negative regression coefficients in the time per throw model indicate an inverse relationship between

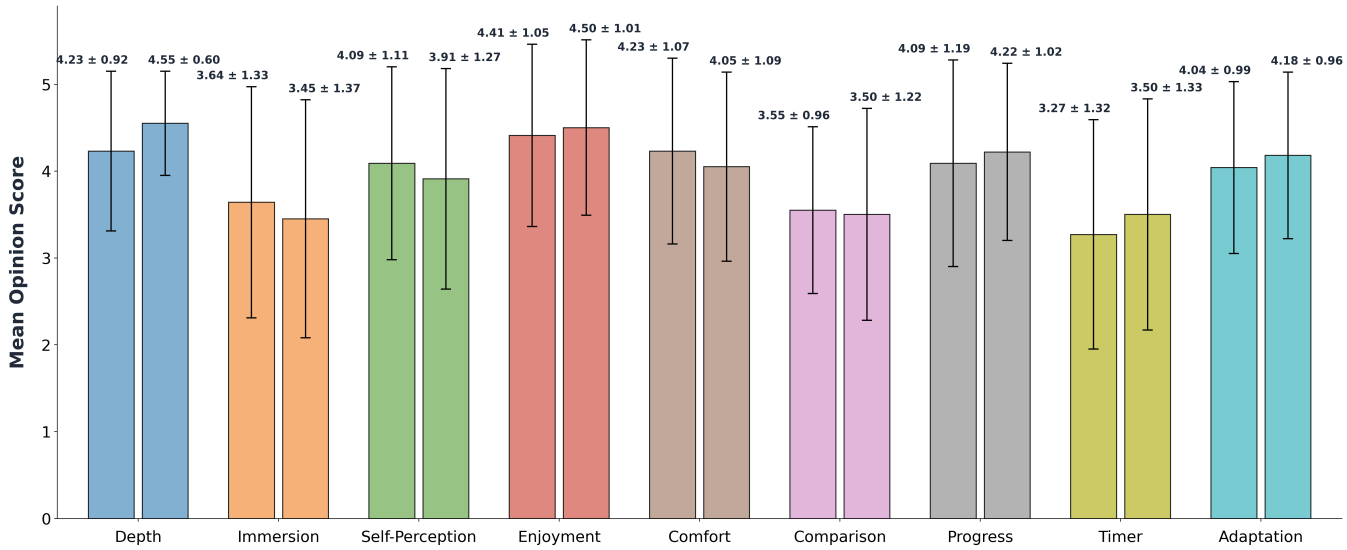


Figure 8: Mean Opinion Scores (MOS) per construct (two items each). Each bar represents mean values across participants (N=22), and error bars indicate standard deviation. Within each construct, the left bar corresponds to the first item and the right bar to the second item listed for that construct in Section 4.

Table 2: Stage-based standardized multiple regression results predicting mean score and mean time per throw.

Stage	Mean Score Model			Time per Throw Model		
	Strongest Predictor	β	R^2	Strongest Predictor	β	R^2
Stage 1 (Near)	Progress	0.40	0.523	Adaptation	**-0.94**	0.579
Stage 2.1 (Early Learning)	Adaptation	**0.80**	0.259	Adaptation	**-0.81**	0.566
Stage 2 (Standard)	Adaptation	0.42	0.206	Adaptation	**-0.83**	0.448
Stage 2.2 (Late Learning)	Self-Perception	0.52	0.229	Enjoyment	1.47	0.567
Stage 3 (Far)	Comfort	0.45	0.355	Comfort	−1.12	0.507
Stage 4 (Time Pressure)	Self-Perception	0.60	0.362	Adaptation	−0.58	0.374

subjective ratings and execution time. Because the dependent variable measures the number of seconds taken for each throw, a lower numerical value represents higher efficiency and faster gameplay. Scoring performance was initially driven by perceived Progress (β=0.40), but self-perception became the dominant factor as learning stabilized (β=0.52) and when participants performed under time pressure (β=0.60). Accuracy at the furthest distance was uniquely predicted by comfort (β=0.45), suggesting that physical ease is critical when motor demands are high. These findings demonstrate that objective success in the simulation is primarily governed by motor calibration and procedural adjustment rather than affective components such as immersion or enjoyment.

A notable pattern in our results is that objective accuracy dropped substantially with distance (Section 5) while depth-related and overall subjective ratings remained high (Section 5, Figure 8). We interpret this dissociation through three, non-exclusive explanations. First, our depth and immersion items were collected once per session rather than per distance condition (see above), so the reported MOS reflects participants' overall impression of the system rather than their perception specifically at the far distance, where performance was weakest; a stage-specific depth rating might reveal a decline that our current design cannot capture. Second, the items we used ask participants to judge whether object positioning *felt* realistic and whether distances were *easy to judge*, which are appraisals of perceptual plausibility and interaction fluency rather than direct measures of throwing accuracy; participants can reasonably perceive a stable, realistic environment even while their motor execution in that environment is imprecise, particularly if they attribute missed throws to task difficulty (e.g., "darts are just hard to throw accurately from far away") rather than to a flaw in the system's depth cues. Third, self-report ratings of enjoyment, comfort, and control are known in the QoE literature to be comparatively insensitive to moderate performance variation, especially in gamified tasks where players may reframe lower scores as part of the intended challenge rather than as a system shortcoming [12]. Distinguishing between these explanations would require condition-specific subjective measures and is a priority for follow-up work.

Relatedly, participant familiarity with real-world dart throwing was recorded only as a coarse self-report category (Section 4) and

was not included as a covariate in our statistical models; we did not observe an obvious qualitative separation between more and less experienced participants during data collection, but we did not test this formally, and we cannot rule out that dart-throwing experience contributed to between-participant variability in accuracy, consistency, or subjective confidence. We flag this as an open question below.

5.9 Limitations

Several design choices constrain interpretation. First, the standard-distance condition had far more throws (40) than near/far (15 each), since it also served as baseline for intra-stage and time-pressure comparisons; effects involving it may partly reflect unequal practice rather than distance. Second, near/far order was counterbalanced, but standard distance always came second and the time-limited condition always came last — so time-pressure effects are confounded with order (learning, fatigue) and can't be attributed to time pressure alone. Third, subjective ratings (depth, immersion, self-perception, enjoyment, comfort) were collected once per session rather than per distance condition, preventing stage-by-stage comparisons against target distance despite their relevance to our research question; we plan to fix this with a per-condition questionnaire in follow-up work. Fourth, we lacked a manipulation isolating depth misjudgment from ordinary motor noise (e.g., holding angular target size constant across distances), so greater hit dispersion at longer distances could reflect amplified execution error rather than depth perception per se. Fifth, prior dart-throwing and AR/VR experience were only coarse self-report categories, not analyzed as covariates, so their influence on performance and ratings is untested. Finally, our sample of 22 limits power for exploratory subgroup analyses and the generalizability of the regression results in Table 2.

6 Conclusion

This paper investigated depth perception, motor performance, and QoE in a controller-free AR dart throwing task. Using a high-fidelity passthrough AR system, a controlled user study with 22 participants evaluated performance across varying target distances and under time-limited conditions. Results showed that throwing accuracy systematically decreased as target distance increased, indicating that actionable depth perception is limited even when stereoscopic rendering is stable. A comparison of early and late trials in Stage 2 revealed a slight improvement in both score and consistency, suggesting short-term motor adaptation to the interaction model. Under time pressure, participants increased throwing speed without a loss of accuracy, suggesting that learned motor patterns can be executed robustly in dynamic conditions. Subjective ratings of depth, immersion, comfort, and control remained consistently high across the study even as objective accuracy declined with distance, indicating that perceived quality of experience is not a reliable proxy for objective performance in controller-free AR interaction and underscoring the value of combining both types of measures when evaluating such systems. These findings demonstrate that modern passthrough AR systems can support reliable controller-free precision interaction, and provide empirical evidence for the design of depth-sensitive AR tasks, highlighting the need to consider both perceptual limitations and user experience when evaluating performance.

Acknowledgments

The financial support of the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology and Development, and the Christian Doppler Research Association is gratefully acknowledged. Christian Doppler Laboratory ATHENA: https://athena.itec.aau.at/.

References

[1] Sian L. Beilock and Thomas H. Carr. 2001. On the Fragility of Skilled Performance: What Governs Choking Under Pressure? *Journal of Experimental Psychology: General* 130, 4 (2001), 701–725. doi:10.1037/0096-3445.130.4.701

[2] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A survey of Augmented Reality. *Found. Trends® Hum.–Comput. Interact.* 8, 2-3 (March 2015), 73–272.

[3] Doug A Bowman, Jian Chen, Chadwick A Wingrave, John Lucas, Andrew Ray, Nicholas F Polys, Qing Li, Yonca Haciahmetoglu, Ji-Sun Kim, Seonho Kim, Robert Boehringer, and Tao Ni. 2006. New Directions in 3D User Interfaces. *Int. J. Virtual Real.* 5, 2 (Jan. 2006), 3–14.

[4] Kjell Brunnström, Katrien De Moor, Ann Dooms, Sebastian Egger-Lampl, Marie-Neige Garcia, Tobias Hossfeld, Satu Jumisko-Pyykkö, Christian Keimel, Chaker Larabi, Bob Lawlor, Patrick Le Callet, Sebastian Möller, Fernando Pereira, Manuela Pereira, Andrew Perkis, Antonio Pinheiro, Ulrich Reiter, Peter Reichl, Raimund Schatz, and Andrej Zgank. 2013. *Qualinet White Paper on Definitions of Quality of Experience.*

[5] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019.* ACM, New York, NY, USA.

[6] Ruizhi Cheng, Nan Wu, Matteo Varvello, Eugene Chai, Songqing Chen, and Bo Han. 2024. A First Look at Immersive Telepresence on Apple Vision Pro. In *Proceedings of the 2024 ACM on Internet Measurement Conference.* ACM, Madrid Spain, 555–562. doi:10.1145/3646547.3689006

[7] Sarah H Creem-Regehr, Jeanine K Stefanucci, and Bobby Bodenheimer. 2023. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 378, 1869 (Jan. 2023), 20210456.

[8] Alejandro Delgado, Hetvi Shah, Md Mehedi Hasan Jibon, Anzhelika Kurnikova, Fareeza Rahman, and Julia Woodward. 2025. "Because we are human, we use hands": Understanding People's Expectations and Mental Models of Virtual Object Interaction in AR Headsets. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, Vol. 1. ACM, New York, NY, USA, 1–8.

[9] Martin Egiziano, Loïc Chomienne, Valentin Hervet, Nicolas Mascret, Richard Kulpa, and Gilles Montagne. 2025. Virtual reality as a perceptual-motor training tool: Validity and fidelity assessments of a 4 × 100 m relay simulator. *Appl. Sci. (Basel)* 15, 6 (March 2025), 3224.

[10] Milad Ghanbari, Wei Zhou, Cosmin Stejerean, Christian Timmerer, and Hadi Amirpour. 2025. SDART: Spatial dart AR simulation with hand-tracked input. In *Proceedings of the 33rd ACM International Conference on Multimedia.* ACM, New York, NY, USA, 13543–13545.

[11] Amirpouya Ghasemaghaei, Mykola Maslych, Yahya Hmaiti, Esteban Segarra Martinez, and Joseph J LaViola. 2024. Towards better throwing: A comparison of performance and preferences across point of release mechanics in virtual reality. In *Graphics Interface*, Vol. 1. ACM, New York, NY, USA, 1–11.

[12] S N Gieser, Caleb Gentry, James LePage, and Fillia Makedon. 2016. Comparing objective and subjective metrics between physical and virtual tasks. In *Lecture Notes in Computer Science.* Springer International Publishing, Cham, 3–13.

[13] Simon Greenwold. 2003. *In form.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[14] Asim Hameed, Sebastian Möller, and Andrew Perkis. 2024. A holistic quality taxonomy for virtual reality experiences. *Front. Virtual Real.* 5 (Aug. 2024).

[15] Markus Holl, Markus Oberweger, Clemens Arth, and Vincent Lepetit. 2018. Efficient physics-based implementation for realistic hand-object interaction in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE.

[16] Tobias Hossfeld and Christian Timmerer. [n. d.]. QoE – ACM SIGMM Records. https://records.sigmm.org/tag/qoe/. Accessed: 2025-12-16.

[17] Yujie Huang, Alexandre Bruckert, and Patrick Le Callet. 2025. Functional Vision Assessment in Virtual Reality: An overview. (2025).

[18] Yujie Huang, Audrey Crozet, Toinon Vigier, Alexandre Bruckert, Patrick Le Callet, and Pierre Lebranchu. 2025. Orientation and mobility test in virtual reality,

a tool for quantitative assessment of functional vision: dataset and evaluation in healthy subjects. *arXiv [cs.HC]* (April 2025).

[19] Yujie Huang, Pierre Lebreton, Alexandre Bruckert, and Patrick Le Callet. 2025. Behavior Metric for Evaluating the Impact of Visual Distortions on User Viewport in a VR Navigation Experience. (Oct. 2025).

[20] Mohammadhossein Izadimehr, Milad Ghanbari, Guodong Chen, Wei Zhou, Xiaoshuai Hao, Mallesham Dasari, Christian Timmerer, and Hadi Amirpour. 2025. SVD: Spatial Video Dataset. In *Proceedings of the 33rd ACM International Conference on Multimedia*. ACM, New York, NY, USA, 12988–12994.

[21] H Kato and M Billinghurst. 2003. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE Comput. Soc.

[22] Chaowanan Khundam, Naparat Sukkriang, and Frédéric Noël. 2021. No difference in learning outcomes and usability between using controllers and hand tracking during a virtual reality endotracheal intubation training for medical students in Thailand. *J. Educ. Eval. Health Prof.* 18 (Aug. 2021), 22.

[23] Chaowanan Khundam, Varunyu Vorachart, Patibut Preeyawongsakul, Witthaya Hosap, and Frédéric Noël. 2021. A comparative study of interaction time and usability of using controllers and hand tracking in virtual Reality training. *Informatics (MDPI)* 8, 3 (Sept. 2021), 60.

[24] Francisco Lopez Luro and Veronica Sundstedt. 2019. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, New York, NY, USA.

[25] Stefan C Michalski, Ancret Szpak, and Tobias Loetscher. 2019. Using virtual environments to improve real-world motor skills in sports: A systematic review. *Front. Psychol.* 10 (Sept. 2019), 2159.

[26] Kiran Nasim and Young J Kim. 2016. Physics-based interactive virtual grasping. In *HCI Korea 2016*. The HCI Society of Korea.

[27] Cristian Pamparau. 2023. A review of Milgram and Kishino's reality-virtuality continuum and a mathematical formalization for combining multiple reality-virtuality continua. *Int. J. Adv. Comput. Sci. Appl.* 14, 4 (2023).

[28] Ken Pfeuffer. 2024. Design Principles & Issues for Gaze and Pinch Interaction. *arXiv [cs.HC]* (Jan. 2024).

[29] Matti Pouke, Elmeri Uotila, Evan G. Center, Kalle G Timperi, Alexis P Chambers, Timo Ojala, and Steven M LaValle. 2024. Adaptation to simulated hypergravity in a Virtual Reality throwing task. *ACM Trans. Appl. Percept.* (Feb. 2024).

[30] Unnikrishnan Radhakrishnan, Lisheng Kuang, Konstantinos Koumaditis, Francesco Chinello, and Claudio Pacchierotti. 2024. Haptic feedback, performance and arousal: A comparison study in an immersive VR motor skill training task. *IEEE Trans. Haptics* 17, 2 (April 2024), 249–262.

[31] Hanna-Riikka Rantamaa, Jari Kangas, Sriram Kishore Kumar, Helena Mehtonen, Jorma Järnstedt, and Roope Raisamo. 2023. Comparison of a VR stylus with a controller, hand tracking, and a mouse for object manipulation and medical marking tasks in virtual reality. *Appl. Sci. (Basel)* 13, 4 (Feb. 2023), 2251.

[32] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented reality*. Addison-Wesley Educational, Boston, MA.

[33] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. Comparing controller with the Hand gestures pinch and grab for picking up and placing virtual objects. *arXiv [cs.HC]* (2022).

[34] Pooya Soltani and Antoine H P Morice. 2023. A multi-scale analysis of basketball throw in virtual reality for tracking perceptual-motor expertise. *Scand. J. Med. Sci. Sports* 33, 2 (Feb. 2023), 178–188.

[35] Evropi Stefanidi, Marit Bentvelzen, Paweł W Woźniak, Thomas Kosch, Mikołaj P Woźniak, Thomas Mildner, Stefan Schneegass, Heiko Müller, and Jasmin Niess. 2023. Literature Reviews in HCI: A Review of Reviews. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Vol. 13. ACM, New York, NY, USA, 1–24.

[36] Sara Vlahovic, Mirko Suznjevic, and Lea Skorin-Kapov. 2022. A survey of challenges and methods for Quality of Experience assessment of interactive VR applications. *J. Multimodal User Interfaces* 16, 3 (Sept. 2022), 257–291.

[37] Apostolos Vrontos, Verena Nitsch, and Christopher Brandl. 2024. Electrical muscle stimulation for kinesthetic feedback in AR/VR: A systematic literature review. *Multimodal Technol. Interact.* 8, 2 (Jan. 2024), 7.

[38] Yibo Wang, Yuhan Luo, Janghee Cho, and Junnan Yu. 2025. What is "spatial" about spatial computing? *arXiv [cs.HC]* (Aug. 2025).