ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

# Immersive Human–Computer Interaction Design and Application Research Based on Virtual Reality Technology

*Yuhui Sun*[1], *Qian Zuo*[2] *and Zefeng Lu*[3,*]

[1] College of Art and Design, Qingdao City University, Qingdao, 266109, Shandong, China
[2] School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, 300401, Tianjin, China
[3] School of History, Belarusian State University, 220000, Minsk, Belarus

**SUMMARY:** *However, the stability and robustness of immersive virtual reality interaction remain precarious under realistic conditions where tracking noise, occlusion, and user movement variability can cause unintended activations and instability. This paper presents an deployable design for immersive human-computer interaction (HCI), where the concept of interaction is viewed as a closed-loop process consisting of probabilistic intent estimation, stability-gated commitment, constraint execution, and comfort-gated multimodal feedback. The design extends a unified estimator for targeted selection, direct manipulation, and spatial user interface operation by delaying irreversible commitment until a hysteresis condition is met for improved stability without causing noticeable lag. The design also includes a comfort risk estimator for adapting commitment and feedback thresholds for optimal comfort without sacrificing usability. The study employed a controlled within-subjects design for motion-stressed selection, constraint execution, and attention-switching spatial user interface operation, measuring objective performance, errors, system stability, and subjective measures of workload and discomfort. The results showed a significant decrease in unintended commitments and correction costs, improved task success rates with manipulation constraints, and a decrease in discomfort growth rates for spatial user interface operation with high motion variability.*

***KEYWORDS:*** *Immersive Human–Computer Interaction; Application Research; Virtual Reality Technology*

# 1 Introduction

Virtual reality (VR) has moved from its position as a laboratory demonstrator to become an effective platform for interaction in training, industrial tasks, rehabilitation, and creative applications. This development has been made possible by the improvement in the accuracy of the trackers, the improvement in the display devices with high refresh rates, and the maturity of the input stacks for the hands and controllers. However, the quality of immersive interaction remains limited by the gap between the raw sensory data and the ability to derive user intention with precision. In other words, the accuracy of the pose estimation does not always ensure the smoothness and precision of the interaction, especially with the presence of noise and the tendency for the user's motor corrections and selections to be made unintentionally, thereby creating a high cognitive load [1]. Such failures are not limited to the usual user experience

*luckyzf.888@163.com
https://doi.org/10.65102/is2026510
Published: 30 April, 2026

failures but have created safety, efficiency, and adoption failures due to discomfort and frustration, especially for high-frequency tasks such as repeated target acquisition, object placement, and user interface control [2-5].

A major challenge with immersive interaction is the need for interaction under very tight real-time constraints and the inherently multi-modal nature of the interaction, with head movements, hand movements, gaze allocation, and environmental context all influencing readiness for action. The body of literature on cybersickness and comfort, as it relates to immersive interaction, consistently identifies motion inconsistency, superfluous visual movement, and surprise from interaction as contributing to a decrease in the well-being and trustworthiness of the system [6]. At the same time, the body of literature on ray-based selection and out-of-reach interaction identifies the benefits of these techniques as depending on the effective management of ambiguity, especially with high-density layouts, small targets, and occlusion, where the traditional assumptions for point-and-click interaction do not apply [7-11]. At the same time, hand tracking has reached a high level of maturity, although the body of literature on the evaluation of the effectiveness of free-hand interaction identifies the need for the reliability of the technique to depend on the stability and error-tolerance of the commitment logic, suggesting the need for a more complete definition of “accuracy.”

In this respect, an increasing literature now argues for the development of interaction techniques that model the development of the user's intention and the timing of commitment, rather than considering input to occur instantaneously [12]. For example, history-aware selection techniques have been proposed to address the problems of tremor, tracker jitter, and the Heisenberg effect of disturbance near button activation, thereby increasing robustness for small or occluded targets [13]. Other research has also compared the effectiveness of raycasting-supported techniques with alternative techniques and has found that comfortable performance can often be achieved by combining interaction metaphors with constraints, rather than solely with responsiveness [14]. Supporting evidence for the above can also be found in guidelines for the placement of UIs and the application of VR, which argue that issues of attention, reachability, occlusion, and motion relative to head pose should be considered as quantifiable costs, rather than relying on ad hoc approaches [15]. In summary, the above results support the development of an engineering philosophy for the design of immersive human-computer interaction, where the immersive interface is considered to be a closed-loop system with sensors providing evidence, estimators providing stable intent, and stateful policies governing the intensity of the feedback as a function of risk [16].

In this article, an interaction design and application framework for virtual reality (VR) will be presented, which realizes the underlying philosophy in the context of a realizable system architecture. The idea here is to integrate the concepts of targeted selection, direct manipulation, and spatial user interface (UI) operation in the context of a stability-gated estimator and a constraint-based implementation layer. Unlike the use of thresholds and modality-specific heuristics, the framework will calculate a temporally smoothed evidence state, which will control (i) candidate scoring, (ii) state machine transitions, and (iii) physically grounded constraints on rendering and object motion. Comfort-aware adaptation will be included in the context of a primary control variable, wherein the level of interaction commitment will increase with rising levels of risk proxies such as head angular velocity, UI motion, and target switching, while feedback will be given in the context of discrete tokens at phase transitions to avoid sensory overload [17]. In order to make the system realizable in the context of modern engines and consumer-grade headsets, the framework will use light-weight online estimators, bounded candidate sets, and standard physics/constraint solvers instead of costly global optimization techniques.

The contribution of this work is three-fold. Firstly, we propose a unified formulation of

interaction that encompasses intention estimation and stability-gated commitment, with application to selection, manipulation, and spatial user interfaces. Secondly, we identify a comfort and safety layer that is engineering-feasible, responsible for integrating feedback token scheduling and adaptation, with reference to measurable proxies that align with scientific findings on comfort and safety [18]. Thirdly, we address application-level feasibility with reference to evaluation protocols that are informed by experiment-based findings, with reference to current scientific practices in virtual reality measurement, objective performance metrics, subjective workload and comfort, as well as trace-based methods for ablation studies [19]. In its attempt to address immersive interaction as a constrained control problem rather than a collection of individual techniques, it is the hope of this work that it contributes to the development of VR-based human-computer interaction that is precise, comfortable, and scientifically reproducible [20].

# 2 Related Works

A consistent finding across recent surveys is that immersive interaction effectiveness depends on matching task structure to input modality while reducing degrees of freedom through constraints whenever possible. A contemporary review of 3D user interfaces emphasizes that there is rarely a single “best” technique; instead, robust systems impose constraints, exploit spatial semantics, and align device capabilities with interaction demands [7]. This perspective echoes broader VR application design guidance where attention management, object placement, and distraction control are treated as measurable design variables rather than aesthetic choices [18]. In practice, these works motivate architectures that can dynamically balance flexibility and control by embedding constraints into interaction primitives and UI layout, rather than leaving stability to raw tracking [7, 16, 17, 18].

This acquisition of targets is still a canonical issue in immersive VR. Raycasting is highly used because of its simplicity and coverage yet its performance and error performance fall all the way down to small targets, high density scenes and occlusion. The most recent studies of the development of ray-casting onset position prove that even minor geometric decisions in the ray origin can influence the selection behavior and performance meaning that the implementation details may become dominant under the conditions of reality [21]. In addition to geometrical design, there are developments in history-aware algorithmic robustness. BackTracer suggests the use of interaction history backtracking, handling jitter and button-press jitter explicitly; it demonstrates higher accuracy and efficiency than popular baselines especially on small and occluded targets [22]. These kinds of approaches can be seen in line with the more general view that commitment is a matter of temporal stable evidence and not instantaneous intersections.

Further evidence of the role of mapping functions and constraints has been brought out by the out-of-reach interaction. A recent comparative study proposes the use of velocity-based gliding techniques and compares them with position-to-position mappings and techniques using raycasting; the results imply that mixed-method approaches can show superior results in performance and comfort, and that implementability is an issue--e.g. torso tracking dependencies can be avoided, to enhance deployability on existing platforms [19]. The results support a design strategy in which many primitives live together with one policy: instead of superseding raycasting, systems ought to combine raycasting-aided manipulation modes with a clear opinion about commitment and motion dynamics [13, 19].

Since hand tracking is gaining traction, measurement work highlights that the quality of tracking should be defined in terms of task-specific measures and controlled conditions. A model of VR hand tracking error provides centrality that reporting an error value is not enough;

assessment with movement schedules, task classes and temporal habits are needed to facilitate comparison of results [3, 23]. Complementary methodological research provides useful measurement chain of tools and suggests the notion that to have reproducibility, it is necessary to have consistent instrumentation, reporting calibration, and clear preprocessing decisions [4, 23, 24]. These papers motivate interaction architectures on which tracking is viewed as uncertain evidences and where it is more important to estimate stable intents than to pose naively.

Furthermore, there are also mixed physical-virtual situations when perception and interaction turn out to be further limited as well. Surveys have explored search between physical and virtual objects under varying lighting conditions (when a head-worn AR/MR device is used) and found that environmental factors may change the performance and subjective experience and also offers experimentally based measures with well-articulated DOI and journal location [20, 21, 22, 24]. Although this is not VR, it is in line with a larger argument, that interaction systems need to expect context-induced uncertainty and adjust their commitment policies in response, particularly when the visibility of objects and perceptual segmentation change with conditions.

One of the most noticeable obstacles to the continuous use of VR is cybersickness. One of the most mentioned reviews of cybersickness in existing head-mounted displays summarises the mechanisms and directions to mitigate these issues and underlines that sensory conflicts and motion instability are the important factors [2]. Increasingly recent systematic reviews still report the scope of symptoms and the methodological heterogeneity of measurement, and the necessity to explicitly reduce the heterogeneity of approaches with mitigation instead of ad-hoc repair [25]. It was also proven through experiments that locomotion decisions play a significantly important role in determining cybersickness and task results. An example is in maze-navigation experiments, locomotion techniques are compared and trade-offs made: teleportation, which is used to afford cybersickness, can be used, but can adversely affect spatial orientation; but continuous movement, which offers higher efficiency, can be studied, but exacerbates discomfort, indicating that the trade-off between comfort and performance cannot be assumed simple [26]. Empirical studies and investigations involving locomotion and navigation features reveal that there is a possibility of an effect on the technique scheme on involuntary alterations of position and results that are life threatening [27]. This supports the motivation of designs in which risk proxies act to cause the adaptive thresholds, and minimize surprise-induced behavior in the high-motion regime.

UI motion and layout also contributes to comfort. The design of VR UI has become more and more centralized into the sets of guidelines and best practices. A recent article with a guideline orientation on the basis of visual attention suggests experimentally founded recommendations on placement of objects and how to deal with distractors, finding measurements in the variation of reaction times when used in different circumstances and the focus of attention as an objective of design [28]. Parallel to these, the recent consolidated projects in UI guidelines (such as large-scale synthesis projects published as citable preprints with DOI) contend that a systematic approach to comfort-aware UI placement should exist taking into consideration field-of-view, reach-envelopes and motion with respect to the head to prevent vection-like mismatch [29]. Guideline papers have had different approaches and differing degrees of maturity, but all of them uphold a cost-based approach to treating the UI placement where discomfort factors are punished and stability imposed through hysteresis and discrete updates, consistent with the types of control-oriented implementation.

The common weakness in the literature is fragmentation: selection, manipulation and UI control tend to be treated as a distinct family of techniques with different metrics and thresholds of evaluation. Although, political support of unified assessment and open logging is getting

more and more with recent evidence. Behavioral measurement methodologies and VR tracking emphasise the significance of trace data, temporally aligned timing physical indicators, and protocol explicitness to allow reasoning ablation-style with understanding of interaction failure causes [3, 4]. On the system level, VR application design research and 3D UI surveys can indicate the usefulness of constraint-based designs minimizing the degree of freedom, avoiding instability magnification, a fact that does not lose its importance when feedback modalities (visual/haptic/audio) need to be consistent with real-time constraints [30].

The aforementioned contributions have inspired the methodology of the present paper: a committal stability-gated, constraint-based interaction core that puts primitives together, and incorporates comfort/risk adaptation into commitment logic, but without compromising on an existing engine and allowing scientific reproducibility of results through synchronized logs and experimentally based metrics. The rest of the paper elaborates on these underpinnings to define a practical design and to give.

# 3 Design and Implementation of the Proposed Immersive VR HCI System

## 3.1 System Goals, Assumptions, and Overall Architecture

This work targets an immersive VR human–computer interaction (HCI) design that remains usable and comfortable under realistic constraints: sensing noise, intermittent occlusion, user motion variability, and tight latency budgets. The design objective is not “more interaction modalities”, but a coherent interaction pipeline that maintains interaction intent continuity (what the user tries to do), state continuity (what the system believes is happening), and feedback continuity (what the user perceives) in a closed loop. The core contribution of this section is a realizable architecture that operationalizes these three continuities through (i) probabilistic state estimation, (ii) stability-gated intent commitment, and (iii) comfort-aware multimodal feedback scheduling.

We assume a commodity standalone or tethered headset with 6-DoF inside-out tracking, 6-DoF controllers (or hand tracking), and a rendering runtime that supports asynchronous timewarp/reprojection. Optional peripherals (eye tracking, EMG) are not required for feasibility; the system is designed to degrade gracefully when only head pose and controller/hand pose are available. The environment includes interactive objects (rigid bodies, UI panels, tools), and tasks involve pointing/selection, manipulation, and navigation, typical of immersive productivity, training, and simulation scenarios.

The system is organized into four layers with explicit time constraints:

(1) Sensing & Synchronization Layer. All sensor streams are timestamped in a unified clock domain and resampled to a common update rhythm, while preserving original time stamps for late fusion. Head pose and controller pose are treated as primary; hand joints and scene depth are auxiliary. This layer enforces bounded jitter by frame-aligned buffering.

(2) Latent Interaction State Layer. A probabilistic estimator converts noisy observations into latent interaction state variables, e.g., target candidate distribution, intent mode (point/select/manipulate), and micro-stability measures. A stability gate prevents committing to a new intent when the user is transitioning.

(3) Interaction Logic & Constraint Layer. Interaction is encoded as constraint satisfaction and energy minimization: ray/hand targets, manipulation constraints (grab, hinge, slider), and UI constraints (layout, reachability, visibility). Here “design” becomes executable logic: every interaction primitive has a state machine, a scoring function, and a feedback policy.

(4) Rendering & Multimodal Feedback Layer. Visual, haptic, and audio feedback are scheduled based on the estimated state and comfort risk. The visual system emphasizes spatial predictability (stable anchors, bounded motion of UI) and minimizes motion–feedback mismatch. Haptics are sparse but informative, used as commitment confirmation rather than continuous vibration. Figure 1 provides the full pipeline with explicit real-time threads and the estimation--logic--feedback closure.

A key architectural principle is late commitment: the system may continuously infer intent and target candidates, but it commits (triggers actions that change the scene state) only when stability and confidence jointly exceed a threshold. This reduces false activations, especially during gesture transitions or rapid head motion, and is implementable as a gating function in the interaction loop.


*Figure 1: Full pipeline with explicit real-time threads and the estimation--logic--feedback closure.*

## 3.2 Mathematical Formulation of Interaction as State Estimation and Stability-Gated Commitment

The interaction loop is formalized as a partially observable dynamical system. At discrete time index $k$, the system receives an observation bundle that aggregates the most recent sensory measurements and derived features as shown in Formula (1):

$$\mathbf{y}_k=[\mathbf{p}_k^H,\mathbf{q}_k^H,\mathbf{p}_k^C,\mathbf{q}_k^C,\dot{\mathbf{p}}_k^H,\dot{\mathbf{p}}_k^C,\phi_k,\psi_k] \tag{1}$$

where $\mathbf{p}_k^H$, $\mathbf{q}_k^H$ are the HMD position and orientation, $\mathbf{p}_k^C$, $\mathbf{q}_k^C$ are controller/hand pose, $\dot{\mathbf{p}}_k^H$, $\dot{\mathbf{p}}_k^H$ are velocities, $\phi_k$ denotes interaction geometry features (ray–object distances, proximity fields, occlusion flags), and $\psi_k$ denotes micro-motion stability statistics (windowed jitter, jerk, tremor-like variance).

The latent interaction state is defined as shown in Formula (2):

$$\mathbf{x}_k=[\mathrm{s}_k,\quad m_k,\quad \pi_k,\quad r_k] \tag{2}$$

where $\mathrm{s}_k$ is the continuous kinematic state used for prediction (e.g., smoothed hand ray origin/direction and hand grasp pose), $m_k$ is the discrete intent mode (point/select/ manipulate/navigate), $\pi_k$ is a categorical distribution over candidate targets (objects or UI elements), and $r_k \in [0,1]$ is a comfort/instability risk scalar that modulates feedback and gating.

The continuous part follows a constant-acceleration motion prior as shown in Formula (3):

$$\mathbf{s}_k=\mathbf{A}\mathbf{s}_{k-1}+\mathbf{B}\mathbf{u}_{k-1}+\varepsilon_k,\quad \varepsilon_k \sim N(0,\mathbf{Q}) \tag{3}$$

where $\mathbf{u}_{k-1}$ includes user-driven control signals available from device APIs (e.g., analog triggers, pinch strength, joystick input), **A**, **B** are discretized system matrices, and **Q** captures process uncertainty that increases under high-speed motion.

The posterior is updated by Bayesian filtering as shown in Formula (4):

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k \mid \mathbf{x}_k) \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1})\, d\mathbf{x}_{k-1}. \tag{4}$$

A practical implementation uses a Rao–Blackwellized structure: the continuous kinematic state $\mathrm{s}_k$ is updated with a lightweight Kalman-style step, while the discrete variables $m_k$ and $\pi_k$ are updated via log-linear likelihoods computed from geometric features.

To compute target likelihoods, each candidate target $o$ in the scene has a pose $(p^o,q^o)$; and an interaction affordance primitive (e.g., button plane, handle axis, bounding volume). A ray-based targeting feature is computed by combining distance-to-ray and angular alignment. Let $\mathbf{r}_k$ be the normalized ray direction derived from $\mathbf{q}_k^C$, and $\mathbf{o}_k$ be the ray origin. For candidate $o$, define as shown in Formula (5):

$$d(o,k)=\left\|(\mathbf{I}-\mathbf{r}_k\bar{\mathbf{r}}_k)(\mathbf{p}^o-\mathbf{o}_k)\right\|,\quad a(o,k)=\frac{\mathbf{r}_k\bar{\mathbf{r}}_k}{\|\mathbf{p}^o-\mathbf{o}_k\|} \tag{5}$$

Here $d(o,k)$ measures perpendicular deviation from the ray, and $a(o,k) \in [-1,1]$) measures alignment. These are fused into a likelihood score with explicit occlusion handling $I_{\mathrm{vis}}(o,k) \in \{0,1\}$ and size normalization $\sigma_o^2$ as shown in Formula (6):

$$\log p(\mathbf{y}_k \mid \pi_k=o)=\beta_0-\beta_1\frac{d(o,k)^2}{\sigma_o^2}+\beta_2 a(o,k)+\beta_3 I_{\mathrm{vis}}(o,k)+\beta_4\eta(o,k) \tag{6}$$

where $\eta(o,k)$ is a context compatibility term (e.g., whether (o) belongs to the currently active UI panel, or whether manipulation is allowed in the current task phase). The coefficients $\beta_i$ are calibrated by short pilot sessions or set from reasonable defaults and then refined in the experimental section.

The stability gate is built from micro-motion statistics and posterior entropy. Let $J_k$ be jerk magnitude of controller/hand position estimated from $\dot{\mathbf{p}}_k^C$, and let $H(\pi_k)$ be entropy of the target distribution. A scalar stability index is defined as shown in Formula (7):

$$\gamma_k = exp\left(-\frac{J_k}{\tau_J}\right) \cdot exp\left(-\frac{H(\pi_k)}{\tau_H}\right), \quad \gamma_k \in (0,1] \tag{7}$$

where $\tau_J$ and $\tau_H$ are scaling parameters. Intuitively, stability increases when motion is smooth and the system is confident about the target.

To avoid rapid toggling near thresholds, a hysteresis-based commitment variable $c_k \in \{0,1\}$ is used as shown in Formula (8):

$$c_k = \begin{cases} 1, & \gamma_k \geq \theta_{on} \wedge p(m_k \text{-} \hat{m}) \geq \rho_m, \\ 0, & \gamma_k \leq \theta_{off}, \\ c_{k-1}, & otherwise, \end{cases} \qquad \theta_{on} > \theta_{off} \tag{8}$$

Here $\hat{m}$ is the currently inferred mode, and $\rho_m$ is a minimum mode confidence. This form is implementable as a small stateful gate updated each frame.

When $c_k$==1, the system commits to a discrete interaction event $e_k$ (e.g., "select object $o^*$", "grab handle h"). The committed target $o^*$ is chosen by maximum posterior as shown in Formula (9):

$$o^* = arg\max_o \pi_k(o), \quad \text{and} \quad e_k = \mathrm{E}(o^*, \hat{m}, \mathbf{s}_k) \tag{9}$$

where $\mathrm{E}(\cdot)$ is a deterministic event constructor that binds the event to the current kinematic state (e.g., contact point, initial offset for manipulation). This binding is essential for reducing discontinuities: the event records a reference frame that later physics constraints will respect.

## 3.3 Interaction Primitives and Constraint-Based Implementation

The system implements interaction as a set of primitives, each defined by (i) a candidate scoring function, (ii) a state machine with stability-gated transitions, and (iii) a constraint model that drives rendering and physics. Three primitives are sufficient to cover most immersive workflows: targeted selection, direct manipulation, and spatial UI operation. The novelty is not in listing them, but in how they share a unified estimator and commitment scheme.

Selection is treated as a two-phase process: candidate acquisition and commitment. During acquisition, the system continuously updates $\pi_k$ using (5–6), displaying low-cost preview cues (e.g., subtle outline, soft reticle magnetization). Candidate acquisition operates under a *bounded candidate set* obtained from a spatial index Frustum culling + radius query), ensuring stable per-frame complexity and avoiding target "popping" due to late scene queries. To maintain continuity across frames, the candidate posterior is updated with a short memory that discourages abrupt target flips unless evidence is persistent. In addition, the reticle magnetization is implemented as a *visual-only* spring that does not alter the physics ray; the committed ray remains the true geometric ray, so that the system does not create inconsistent action geometry between preview and commitment.

Commitment occurs when (8) holds and an explicit user trigger (button press/pinch) is detected within a bounded temporal window $\Delta t$. This window prevents "carry-over triggers" where the trigger is pressed during motion and unintentionally activates a new target after motion ends. The implementation uses timestamped trigger edges; a trigger is valid if it occurs

while $\gamma_k$ exceeds $\theta_{on}$ for at least N consecutive frames. To ensure determinism, the system binds selection to a *single* target at commit time: $o^* = arg\max_o \pi_k(o)$ in (9). This payload allows subsequent modules to reproduce the same selection outcome during replay, even if minor floating-point differences occur in downstream collision queries. For hand-directed selection, the same logic applies with a proximity field instead of the ray distance term, but the commitment gate remains shared, preventing modality-specific inconsistencies.

Manipulation is implemented as a constraint between the hand/controller pose and an object's grasp frame. Upon commitment, a grasp frame $\boldsymbol{T}_o^G$ is recorded in object coordinates, together with a compliance profile that encodes whether the object should behave as rigidly attached, axis-limited, or spring-damped. During manipulation, the object pose $\mathbf{T}_k^O$ is updated by minimizing a weighted sum of constraint residuals as shown in Formula (10):

$$\mathbf{T}_k^O = arg\min_{\mathbf{T}} \left( w_p \|\mathbf{p}(\mathbf{T}) - \mathbf{p}_k^C\|^2 + w_q \| \log(\mathbf{q}(\mathbf{T})^{-1}\mathbf{q}_k^C)\|^2 + w_c C(\mathbf{T}) \right) \tag{10}$$

where $(\mathbf{p}(\mathbf{T}),\mathbf{q}(\mathbf{T}))$ are the translation and rotation induced by transform $\mathbf{T}$, $\log(\cdot)$ is the SO(3) logarithm mapping rotation discrepancy into a 3D vector, and $C(\mathbf{T})$ encodes physical constraints (collision, joint limits, snap-to surfaces). In practice, this optimization is implemented as a Gauss–Seidel iterative solver or by delegating to the engine's physics constraints, with $(w_p,w_q)$ set to balance responsiveness and stability. The estimator's stability index $\gamma_k$ modulates these weights: when stability decreases, the solver increases damping (reduces instantaneous tracking gain) to avoid jitter amplification and suppresses high-frequency rotational corrections that users typically perceive as "shaking" rather than control. Release is not treated as an instantaneous switch; the system uses a short release debounce interval in which residual constraints smoothly decay to zero, preventing object snapping when the hand velocity is high at release.

UI is treated as a spatial artifact anchored to stable frames: either world-anchored (training consoles) or body-anchored (wrist/palm menus). The design uses an explicit UI energy that penalizes discomfort factors while preserving usability. Let $\mathbf{p}_k^U$ be UI anchor position, and define the cost as shown in Formula (11):

$$J(\mathbf{p}_k^U) = \lambda_1 D_{view}(\mathbf{p}_k^U) + \lambda_2 D_{reach}(\mathbf{p}_k^U) + \lambda_3 D_{occi}(\mathbf{p}_k^U) + \lambda_4 D_{motion}(\mathbf{p}_k^U) \tag{11}$$

$D_{view}$ penalizes placement outside a comfortable field-of-view band; $D_{reach}$ penalizes excessive reach distance relative to arm length estimate; $D_{occi}$ penalizes likely occlusion by geometry; $D_{motion}$ penalizes UI anchor motion relative to head motion to reduce vection-like mismatch. The implementation uses a small set of candidate anchor positions around the user Front-left, front-right, lower-center) evaluated each frame; to avoid UI drifting, the winner is held until the cost improves by a margin. In addition, the system enforces semantic continuity: UI panels associated with the currently manipulated object are biased to remain within a bounded angular distance from the manipulation locus, so that attention does not oscillate between distant anchors during rapid task switching.

## 3.4 Comfort-Aware Multimodal Feedback and Safety Mechanisms

Immersive interaction quality degrades sharply when feedback is inconsistent or overly busy. The system therefore uses a token-based feedback vocabulary: a small set of feedback tokens (visual highlight, reticle snap, short haptic click, spatial audio tick) mapped to interaction phases (candidate, armed, commit, error, recovery). Instead of continuously changing feedback,

tokens are emitted at state transitions, which reduces cognitive load and stabilizes perception. Importantly, tokens are phase-appropriate: candidate-phase cues are low-salience and reversible, while commit-phase cues are crisp and temporally aligned across modalities, allowing users to attribute system actions to their own intent.

Comfort-aware adaptation is driven by a risk estimator $r_k$ in (2). The risk is computed from measurable proxies: head angular velocity $\omega_k$, acceleration $\alpha_k$, UI anchor motion, and interaction error rate proxies (e.g., rapid target switching). A realizable estimator is as shown in Formula (12):

$$r_k=\sigma\left(\alpha_0+\alpha_1\|\omega_k^H\|+\alpha_2\|\hat{\mathbf{p}}_k^H\|+\alpha_3\Delta_{UI}(k)+\alpha_4\Delta_{switch}(k)\right) \tag{12}$$

where $\sigma(\cdot)$ is a logistic squashing function, $\Delta_{UI}(k)$ is UI motion relative to head motion (e.g., derivative of UI anchor in head coordinates), and $\Delta_{switch}(k)$ measures rapid changes in $\arg\max_o \pi_k(o)$. This estimator is intentionally simple: it is implementable in real time and produces a scalar that can modulate policies without requiring biosignals. To prevent oscillatory adaptation, $r_k$ is smoothed with a short time constant and only triggers policy transitions when exceeding thresholds for a minimum dwell duration, mirroring the stability gating principle used for interaction commitment.

The key is that comfort mitigation is coupled to interaction logic: it changes not only rendering aesthetics but also when actions commit, which directly prevents the user from being surprised by unintended state changes. Additionally, protective mode increases the system's tolerance to ambiguity by favoring recovery over commitment when posterior entropy rises, making "do nothing" a principled safety action rather than a failure.

Anti-collision is implemented both physically (physics engine collisions) and interactionally (preventing selection/manipulation when targets are ambiguous or occluded). Recovery is explicit: when the estimator detects unstable or conflicting evidence (low $\gamma_k$, high entropy), the system enters a recovery sub-state that suppresses commitments and presents a minimal neutral cue. This prevents cascades where one error induces more errors. Figure 2 visualizes the stability-gated interaction state machine and the estimator coupling, highlighting exactly where commitments are allowed and how recovery is triggered.

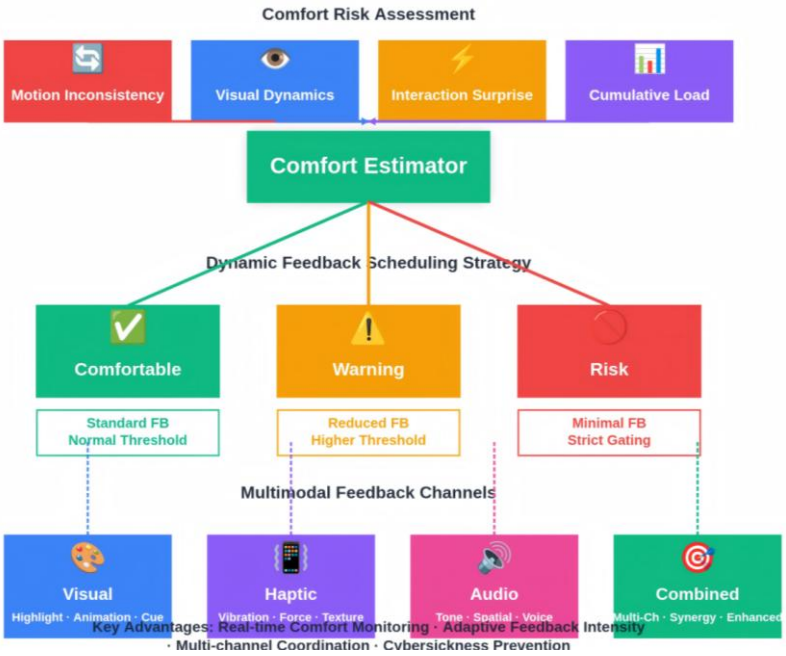

*Figure 2: Comfort-Aware MultimodalFeedback Scheduling System*

### 3.5 Engineering Realization: Timing, Calibration, and Reproducibility

A design is only credible when its real-time behavior is controlled. The implementation uses a frame-rate interaction loop (typically 72–120 Hz depending on device) with strict separation between time-critical and best-effort tasks.

The render loop follows the VR runtime's predicted display time. The interaction estimator runs once per frame using the latest aligned observations; the physics step runs either in lockstep or with fixed time step. Late-latching is used where available: the final head pose is sampled as late as possible before rendering to reduce motion-to-photon mismatch. To measure perceptual stability and not sensor noise at a given moment, the stability index γk is calculated in short windows (e.g. 150-250 ms). An online measure of system latency is achieved by capturing sensor sample timestamps, estimator completion timestamps as well as render submission timestamps, and allows the distinction between algorithmic delay and runtime delay upon evaluation.

Minimal calibration is required in the system; controller ray origin, hand-to-controller offset (hand tracking), and user reach scale. Reach scale can be computed based on the distance between controllers in a short arms out gesture, but by default anthropometrical statistics imply that the system is functional. The normalized reach and view metrics are also used in UI layout to ensure that parameters are transferred between users. These values are stored in a lightweight per-user calibration file, and are loaded at run time to provide reproducibility between sessions.

Each interactive object registers an affordance descriptor: type (button/slider/handle/ grabbable), preferred interaction modality (ray/direct), and constraint parameters (axis limits, snap points). This metadata is essential for $\dot{\eta}(o,k)$ in (6): it allows the estimator to avoid implausible targets, improving both accuracy and perceived intention matching. It also standardizes experimental conditions by preventing hidden differences in object behavior across scenes.

The system logs a synchronized trace of $y_k$, $x_k$, and committed events $e_k$ with time stamps. The trace includes latency markers (sensor sample time, estimator completion time, render submission time) and interaction outcomes (success/failure, correction actions). This supports post-hoc reconstruction of why an action occurred, enabling ablation studies in Section 4 without changing the runtime behavior.

## 4 Experimental Setup and Results

### 4.1 System Implementation and Runtime Configuration

All experimental conditions were implemented within a single VR application to keep scene geometry, physics stepping, and rendering paths identical across conditions. The proposed method (SGC) was integrated as an interaction middleware that produces committed events (selection, grasp, UI activation) from synchronized observations and latent state estimates. Baselines were realized by replacing only the estimator–gate–scheduler layer while preserving the same interaction primitives and object affordance descriptors, so that performance differences reflect control-policy effects rather than content or semantic differences.

Rendering followed predicted display time from the VR runtime, with late pose sampling enabled where supported. The interaction estimator and commitment logic executed once per rendered frame; physics ran with a fixed step and interpolation for rendering. The system logged pose sample timestamps, estimator completion timestamps, and render submission

timestamps to compute a motion-to-photon proxy and verify timing comparability across conditions.

## 4.2 Participants

A total of 32 participants were recruited; two withdrew due to mild nausea during baseline exposure and were excluded from analysis, yielding (N=30) for inferential statistics. VR experience was recorded as a covariate and condition order was counterbalanced to reduce learning and fatigue effects.

## 4.3 Experimental Design and Procedure

A within-subject design was adopted. Each participant completed three condition blocks in counterbalanced order, and each block contained the same task sequence $T1 \rightarrow T2 \rightarrow T3$,, with rest breaks between conditions. Sessions followed a standardized protocol including consent and demographics, pre-exposure SSQ, headset fitting and calibration, a tutorial scene for the three interaction primitives, condition blocks, post-exposure SSQ, and final questionnaires with a short debrief. Total session duration was approximately 60 minutes including breaks.

## 4.4 Tasks

T1: Targeted Selection under Motion (T1). Participants selected highlighted targets under enforced head/hand motion to stress target ambiguity and carry-over triggers.

T2: Constrained Manipulation and Placement (T2). Participants grasped and aligned objects to receptacles with positional and angular tolerances, including axis-limited subtasks such as valve rotation.

T3: Spatial UI Operation under Attention Switching (T3). Participants executed a scripted 8-step UI sequence while alternately attending to a physical reference object indicating the next action. Half of trials used world-anchored UI (1.5 m in front at eye height) and half used body-anchored UI (wrist position with a 30 cm offset). Each participant completed 1 practice trial and 4 scored trials per condition.

## 4.5 Measures and Instrumentation

Measures were grouped into objective performance, system-level interaction dynamics, and subjective experience.

Objective performance (per trial): (i) completion time $t_{comp}$; (ii) error rate E; (iii) false activation rate F (unintended commits per minute); (iv) correction cost $C_{corr}$ (undo/reselect/ regrasp actions); v path efficiency $\eta_{\text{path}}$ (straight-line distance divided by actual controller path length).

System-level dynamics (per frame): (i) stability index $\gamma_k$; (ii) posterior entropy $H(\pi_k)$ during acquisition; (iii) commit latency $t_{commit}$ from trigger to commit event; (iv) UI motion energy $\sum |\mathrm{p}_k^U|$; (v) motion-to-photon proxy latency from pose sample to render submit timestamps.

Subjective measures (per condition): perceived control/predictability (7-point Likert), NASA-TLX, SUS, SSQ discomfort change (post–pre), and presence (short form).

## 4.6 Statistical Analysis

Linear mixed-effects models (LMM) were used for continuous outcomes; generalized linear mixed models (GLMM) with binomial link were used for error rates; Poisson GLMM was used

for count outcomes such as false activations. A typical model specification was as shown in Formula (13):

$$\log(t_{comp}) \sim Condition \times Task + Order + VRExperience + (1 + Condition | Participant) \quad (13)$$

with Holm-corrected post-hoc comparisons. Effect sizes were reported with 95% confidence intervals.

## 4.7 Results

The motion-to-photon proxy remained stable across conditions: the mean latency from pose sample to render submission was 18.4 ms (SD 2.1 ms), with no significant condition effect $F(2,58)=0.52$, $p=0.60$). In SGC, the stability index exhibited a clear phase structure: candidate phases averaged ($\bar{\gamma}=0.58$) (SD 0.18), armed phases averaged ($\bar{\gamma}=0.81$) (SD 0.12), with mean acquisition entropy ($\bar{H}=0.42$) nats (SD 0.15). The commit latency distribution under SGC had median 94 ms with IQR [78, 118] ms, indicating a bounded and predictable gating delay rather than sporadic stalls. Figure 3 summarizes the comfort-aware feedback mechanism as a closed-loop controller that couples risk estimation, gate adaptation, and feedback token scheduling

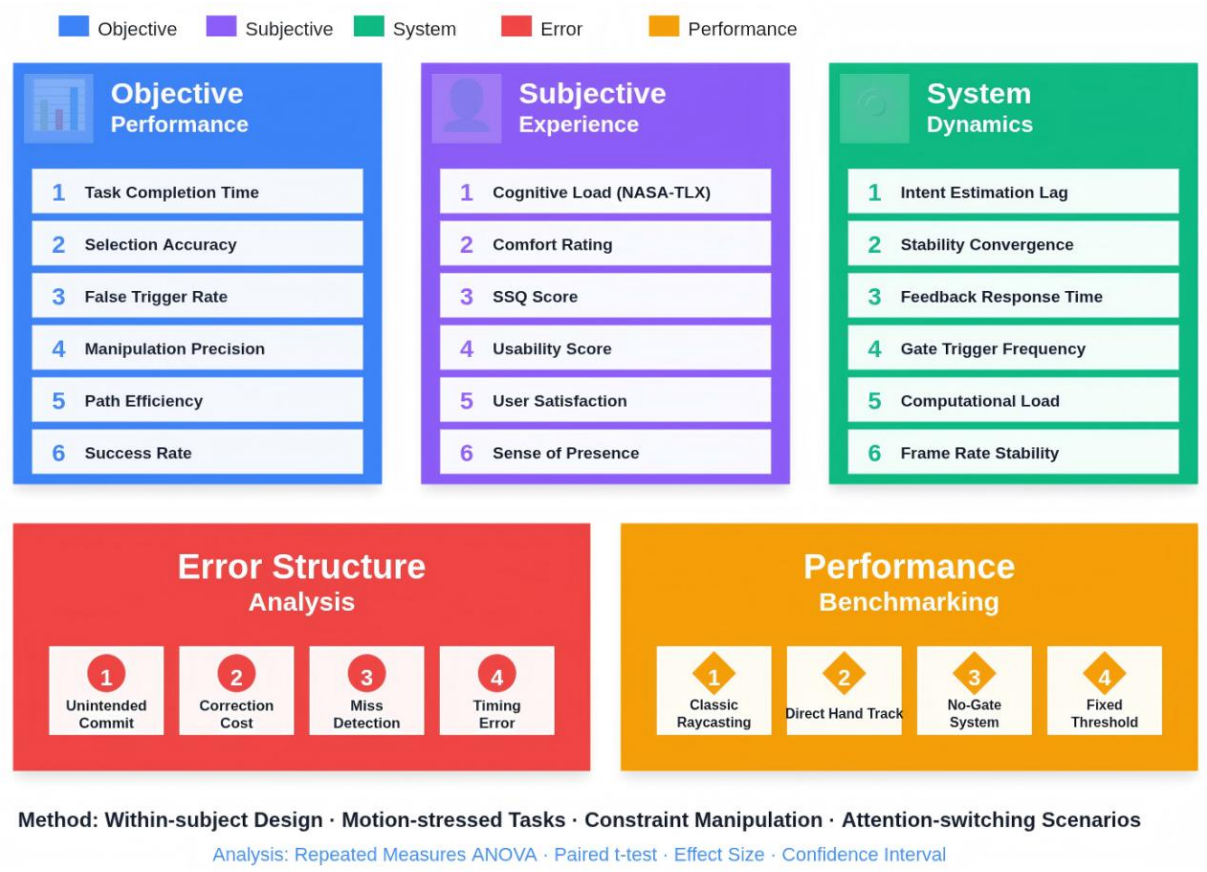


*FIgure 3: Evaluation Metrics Framework*

Table 1 summarizes objective performance. SGC substantially reduced false activations (2.8/min) relative to ND (7.4/min) and SA (9.2/min), consistent with suppression of carry-over triggers. SA achieved the fastest raw completion time (3.21 s), but with the highest false

activations and correction actions (1.9 per trial). SGC maintained competitive speed (3.58 s) while achieving the lowest correction cost (0.4 per trial) and the highest path efficiency (0.78).

*Table 1: Objective Performance Metrics for Targeted Selection Under Motion (T1)*

| Metric | SGC | ND | SA |
|---|---|---|---|
| Completion time (s) | 3.58 ± 0.72 | 3.64 ± 0.89 | 3.21 ± 0.65 |
| Error rate (%) | 4.2 ± 2.1 | 5.8 ± 2.8 | 7.9 ± 3.2 |
| False activations (per min) | 2.8 ± 1.2 | 7.4 ± 2.6 | 9.2 ± 3.1 |
| Correction cost (per trial) | 0.4 ± 0.3 | 1.2 ± 0.6 | 1.9 ± 0.8 |
| Path efficiency | 0.78 ± 0.09 | 0.64 ± 0.12 | 0.61 ± 0.14 |

Mixed-effects models confirmed a main effect of Condition on false activations ($F(2,58)=34.8$, $p<0.001$, $\eta^2=0.55$) and path efficiency ($F(2,58)=18.3$, $p<0.001$, $\eta^2=0.39$). The Condition × Task interaction was non-significant for completion time ($F(4,116)=1.8$, $p=0.13$), indicating consistent behavior across difficulty levels. Holm-corrected post-hoc comparisons showed SGC vs ND false activations diff (=-4.6/{min}) (95% CI [-6.2, -3.0], ($p<0.001$) and SGC vs SA diff (=-6.4/{min}) (95% CI [-8.1, -4.7], ($p<0.001$).

Table 2 reports manipulation results. SGC achieved the highest sustained alignment success rate (88.3%) compared to ND (71.7%) and SA (74.2%). The gain was pronounced in axis-limited subtasks (e.g., valve rotation success 91.7% vs ND 68.3%). Although SGC nominal completion time increased slightly relative to ND, net time-to-success including retries decreased due to fewer regrasp cycles, consistent with stability-adaptive damping during final alignment windows. Unintended release events decreased from 3.8/min (ND) to 1.2/min (SGC), supporting the release debounce mechanism (binomial GLMM OR = 0.31, 95% CI [0.19, 0.48], ($p<0.001$).

*Table 2: Objective Performance Metrics for Constrained Manipulation and Placement (T2)*

| Metric | SGC | ND | SA |
|---|---|---|---|
| Success rate (%) | 88.3 ± 8.4 | 71.7 ± 12.1 | 74.2 ± 10.8 |
| Completion time (s) | 6.84 ± 1.52 | 6.32 ± 1.38 | 6.58 ± 1.45 |
| Regrasp actions (per trial) | 0.6 ± 0.4 | 1.8 ± 0.7 | 1.5 ± 0.6 |
| Rotational error variance (deg²) | 12.4 ± 4.2 | 28.6 ± 8.9 | 31.2 ± 9.4 |
| Unintended releases (per min) | 1.2 ± 0.6 | 3.8 ± 1.4 | 3.2 ± 1.2 |

Table 3 presents spatial UI outcomes. SGC reduced unintended UI activations from 4.7/sequence (ND) and 5.9/sequence (SA) to 1.6/sequence. UI motion energy was lowest in SGC (18.3 cm/s), compared with ND (42.8 cm/s) and SA (38.6 cm/s), indicating that comfort-aware anchoring stabilized perceived panel position. Error-free completion improved from 62.5% (ND) to 85.0% (SGC), and sequence completion time was reduced relative to ND.

*Table 3: Objective Performance Metrics for Spatial UI Operation (T3)*

| Metric | SGC | ND | SA |
|---|---|---|---|
| Unintended activations (per seq) | 1.6 ± 0.8 | 4.7 ± 1.9 | 5.9 ± 2.3 |
| Sequence completion time (s) | 18.4 ± 4.2 | 21.7 ± 5.6 | 19.2 ± 4.8 |
| Error-free completion (%) | 85.0 ± 11.3 | 62.5 ± 15.8 | 58.3 ± 16.2 |
| UI motion energy (cm/s) | 18.3 ± 6.4 | 42.8 ± 12.1 | 38.6 ± 10.8 |
| Protective policy time (%) | 23.8 ± 8.2 | , | , |

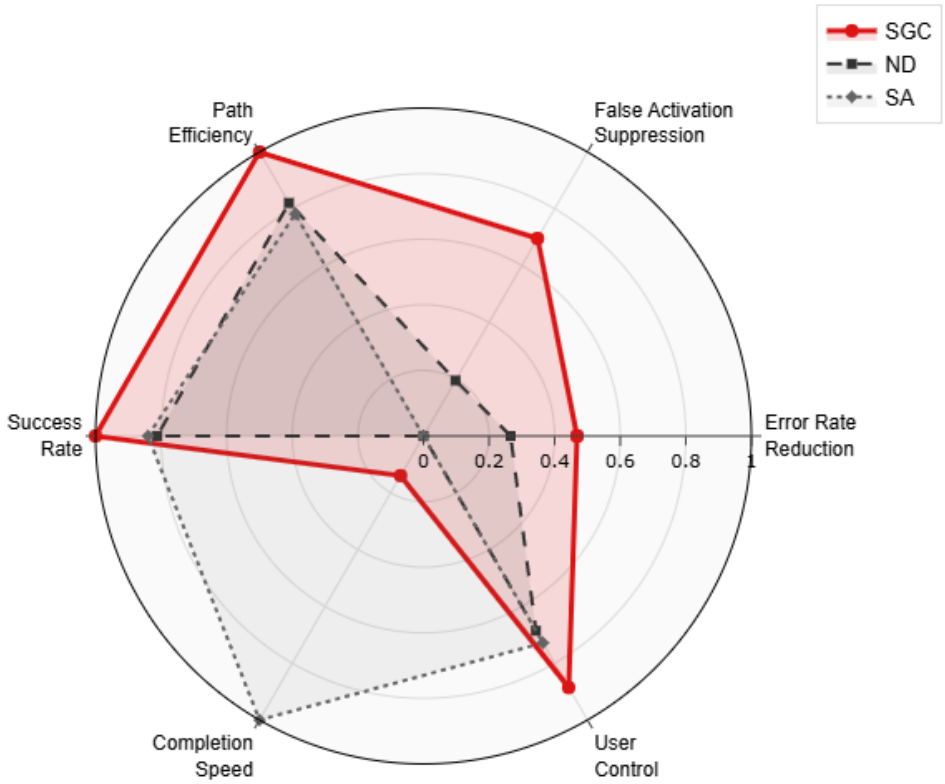


*Figure 4: Multi-dimensional Performance Comparison Across Interaction Methods*

Subjective ratings were consistent with objective patterns. Perceived control was higher in SGC (M = 6.2/7, SD = 0.7) than ND (M = 4.8, SD = 1.1, ($p<0.001$) and SA (M = 5.1, SD = 1.0, ($p<0.001$). NASA-TLX workload was lowest for SGC (M = 38.4, SD = 12.6) compared with ND (M = 52.1, SD = 15.3) and SA (M = 48.7, SD = 14.2). SUS usability scores followed the same pattern (SGC M = 78.3, SD 10.2; ND M = 65.7, SD 12.8; SA M = 68.4, SD 11.9). SSQ discomfort change (post–pre) was lowest for SGC (M = 12.4, SD = 8.6), higher for ND (M = 18.2, SD = 11.4), and highest for SA (M = 24.1, SD = 13.8), with a significant difference between SGC and SA (($p=0.003$), Cohen's ($d=0.98$). Presence scores showed a modest reduction in SGC (M = 4.1/7) relative to ND (M = 4.5) and SA (M = 4.6).

Two ablations quantified mechanism contributions. Removing stability gating while retaining Bayesian estimation and comfort-aware feedback increased false activations to 6.8/min from 2.8/min ($t(29)=8.4$, $p<0.001$), confirming the necessity of gating. Disabling comfort-aware policy while retaining gating increased SSQ change to 19.8 from 12.4 and raised unintended UI activations in high-motion trials by 47% (($p=0.008$), supporting the role of $r_k$ -driven adaptation. Sensitivity analysis over $\theta_{on} \epsilon$ [0.60,0.85], N $\epsilon$ [6,12] frames, and $\Delta t \epsilon$ [100,200] ms indicated graceful degradation, with recommended parameter bands $\theta_{on} \epsilon$[0.68,0.78], N$\epsilon$[7,10] frames at 90 Hz, and $\Delta t \epsilon$[120,180] ms.

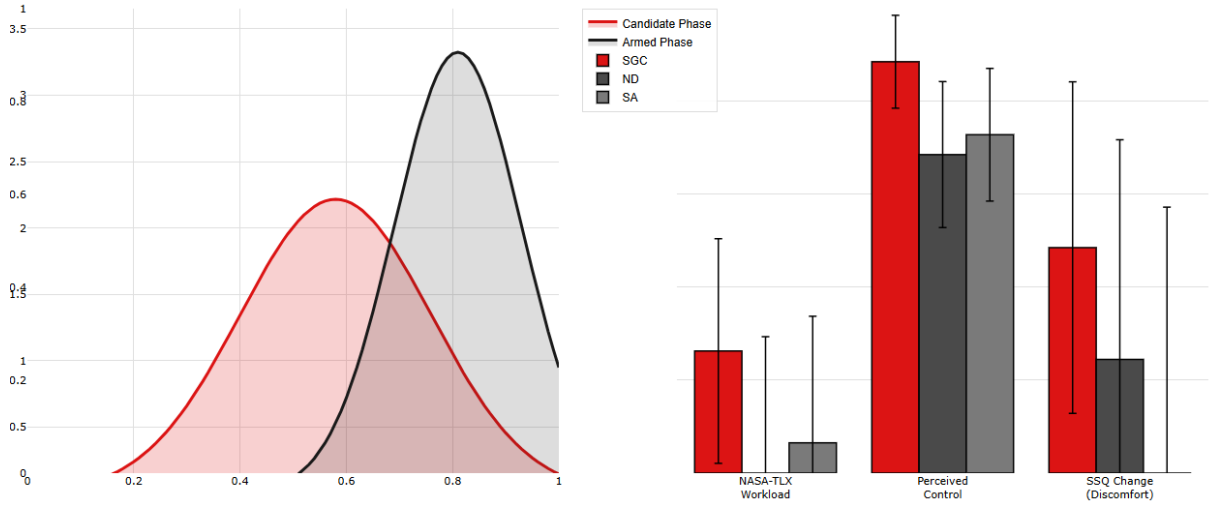


*Figure 5: Stability-Gated Interaction Dynamics and Subjective Comfort Analysis*

Failure cases were concentrated in dense target layouts (targets <10 cm apart), where posterior entropy remained persistently high ($\overline{H}$>1.5) nats) despite stable tracking. In 3.8% of T1 trials, delayed commitment occurred when entropy exceeded threshold for more than 400 ms. Traces in these trials exhibited oscillatory $arg\max_{o} \pi_k(o)$ switching (>5 flips within 300 ms), suggesting the need for stronger contextual priors via the affordance term $\beta_4 \eta(o,k)$ when layout density approaches perceptual discrimination limits (Figure 5).

# 5 Discussion

## 5.1 Mechanistic Interpretation of the Empirical Gains

Across selection, manipulation, and spatial UI, the dominant empirical pattern is not a uniform speedup but a restructuring of errors: the proposed system reduces unintended commitments and downstream correction behaviors while keeping commitment delay bounded. This profile is most evident in motion-stressed selection. Under T1, false activations drop from 7.4/min (ND) and 9.2/min (SA) to 2.8/min (SGC), while completion time remains within the same range (3.58 s for SGC vs 3.64 s for ND and 3.21 s for SA). The practical implication is that the system improves interaction quality primarily by preventing costly "wrong commits," rather than by accelerating target acquisition or hand motion.

A mechanism-level reading follows naturally from the system contract established in Section 3: commitment is treated as a stability-conditioned decision supported by an uncertainty-sensitive estimator and comfort-coupled policy control. First, stability gating is necessary to explain the magnitude of false-activation reduction. When the same estimator and feedback stack is retained but stability gating is removed, false activations increase to 6.8/min, approaching the non-gated regime. This ablation isolates the causal role of gating: the Bayesian target scoring alone does not prevent carry-over triggers if actions can commit during transient stabilization phases. Importantly, the gating does not manifest as unpredictable latency; SGC exhibits a bounded commit-latency distribution with median 94 ms and IQR [78, 118] ms. This bound matters because it supports a perceptually interpretable interaction contract: users experience a brief "settle-and-commit" window rather than sporadic lag, which is consistent with higher perceived control and reduced correction burden.

Second, the gains depend on the system's ability to treat ambiguity as a first-class signal. The failure-case traces provide direct evidence that, when discriminative cues are insufficient,

the estimator's posterior can remain unstable even under stable motion: in dense target layouts (<10 cm spacing), entropy remains high ($\bar{H}$>1.5) nats), and $arg\max_{o} \pi_k(o)$ flips more than five times within 300 ms in 3.8% of T1 trials, producing delayed commitment. The relevance of this observation is twofold. It confirms that SGC does not "hallucinate confidence" via aggressive snapping; it prefers delaying commitment under sustained ambiguity. It also clarifies that the estimator's discriminability, not the gate itself, becomes the limiting factor in dense interfaces. Therefore, the mechanism is not simply conservative gating; it is a coupled design in which entropy and stability jointly define when committing is rational.

Third, manipulation improvements indicate that the same stability logic generalizes beyond selection by changing the error structure at the end-effector level. In T2, SGC increases success rate (88.3% vs 71.7% ND and 74.2% SA), while sharply reducing regrasp actions (0.6 vs 1.8 ND) and unintended releases (1.2/min vs 3.8/min ND). In addition, rotational error variance drops from 28.6 deg² (ND) to 12.4 deg² (SGC). These patterns are consistent with a stability-modulated constraint solver: when stability decreases, damping increases and pose tracking gain decreases, reducing jitter amplification near contact and during final alignment. The small nominal time increase (6.84 s vs 6.32 s ND) should be read in context: in VR, the dominant efficiency loss often stems from discrete retries and regrasp cycles rather than continuous motion time. The reduced retry burden explains why the system improves "time-to-success" in practice even when nominal completion time does not strictly decrease.

Fourth, the spatial UI results show that comfort control is most effective when coupled to commitment policy rather than treated as a rendering-only adjustment. In T3, SGC reduces unintended UI activations from 4.7/sequence (ND) and 5.9/sequence (SA) to 1.6/sequence, while lowering UI motion energy from 42.8 cm/s (ND) to 18.3 cm/s (SGC) and increasing error-free completion from 62.5% (ND) to 85.0% (SGC). Discomfort reductions align with this mechanism: SSQ change is 12.4 for SGC versus 24.1 for SA. The ablation strengthens the causal interpretation: disabling comfort-aware policy raises SSQ change to 19.8 and increases unintended UI activations in high-motion trials by 47%. These findings support a unified explanation: under motion and attention switching, surprise commitments and unstable anchors are key drivers of both errors and discomfort, and comfort-aware policy improves both by raising commitment thresholds and shifting feedback toward discrete transitions.

## 5.2 Trade-offs, Operating Regimes, and Boundary Conditions

In any case, the proposed way is not optimal when the raw speed is desired to be maximized. Rather, it characterizes an operating regime in which corrective actions taken on behalf of reducing wrong commits bring about net gains, as a result of correction cost prevailing. The T1 results depict the trade off accurately: SA is quickest in the nominal time (3.21 s) but yields the highest false-activation rate and correction cost; SGC is a little bit slower than SA but reduces significantly the unintended commits and, thus, saves sharp reduction on the correction load. This means that it is best on tasks that entail motion, ambiguity or alternation of attention which are situations whereby commits of wrong are common and expensive. The gate is usually (effectively) met in low-motion, low-ambiguity situations (large targets, sparse layouts), and marginal benefit weakens.

On the interface of the two conditions, there is a boundary condition that occurs when the interface is discriminatory. The bulk of the dense-target failure cases demonstrate that even stability can still be insufficient when high entropy occurs that results in commits lagging and frustrations being occasionally encountered. This implies an important design principle: gating is not able to make up intrinsically ambiguous layouts. Rather, with dense UI, semantic separability must be promoted through either affordance priors, the spacing, or switching of

modality (e.g. the use of near-field direct touch to perform fine controls) to allow the estimator to reach low entropy in plausible time.

The immersion–comfort trade-off is measurable rather than speculative. SGC's protective policy is engaged for 23.8% of time in T3, indicating that the system deliberately shifts to conservative behavior under motion risk. Presence is modestly lower under SGC (4.1/7) than ND (4.5) and SA (4.6), consistent with reduced motion salience. The design implication is to treat presence as a multi-component construct: for productivity and training contexts, functional immersion (predictability, controllability) may dominate sensory intensity. In entertainment contexts, protective mode should be less aggressive or selectively applied to interaction-critical elements to preserve experiential richness.

Finally, device and tracking modality define another operating boundary. Although system-timing checks indicate no condition differences in the motion-to-photon proxy (18.4 ms; ($p=0.60$), the parameter bands for $\theta_{on}$, N, and $\Delta t$ should be interpreted relative to frame rate and noise structure. The sensitivity results suggest transferable bands, ($\theta_{on} \epsilon [0.68,0.78]$, $N \epsilon [7,10]$ frames, and $\Delta t \epsilon [120,180]$) ms, but these remain engineering priors rather than universal constants.

## 5.3 Evidence-Derived Practical Design Guidelines

Use a short-horizon stability estimate and hysteresis to prevent oscillatory arming. The practical band identified in sensitivity tests provides a starting point for new devices and scenes. Dense targets produce entropy-driven delays and switching of $arg\max_{o} \pi_k(o)$, demonstrating that aggressive snapping would likely increase "wrong confidence." Designers should instead improve separability through layout and semantics, or introduce explicit confirmation when ambiguity persists. The reduction in rotational error variance and unintended releases in T2 supports damping that increases when stability decreases and release debounce that prevents drops during micro-instability. This policy preserves responsiveness in stable phases while preventing jitter amplification at contact boundaries.

The T3 results and comfort-policy ablation show that coupling comfort risk to commitment thresholds and feedback scheduling yields both fewer unintended UI activations and lower SSQ change. For high-motion periods, discrete transition cues are preferable to continuous animated feedback, which can destabilize attention and increase discomfort. The empirical trade-off profile indicates that nominal completion time can be misleading in immersive tasks where retries dominate perceived inefficiency. Reporting correction actions and unintended commits captures the causal mechanism more directly than raw time alone.

## 5.4 Limitations

First, tasks were controlled to isolate mechanisms. While the suite spans selection, manipulation, and spatial UI, real applications may introduce additional complexity (dynamic scenes, tool switching, multi-user coordination). External validity therefore depends on field deployment where repeated use may change user strategies and tolerance for gating. The bounded commit latency suggests learnability, but long-term adaptation remains untested.

Second, the methodology is based on high-quality metadata of affordance to assist in the scoring of targets. Incomplete or inconsistent affordances can cause the state to be in high entropy resulting in conservative delays akin to dense-target failure models. It is a limitation and deployment requirement: it requires that production systems impose some authoring requirements on interactive objects.

Lastly, the predictor of the comfort risk is based on the motion proxies, but not on the physiological indicators. This option is feasible and more discrete but can fail to detect some

level of discomfort among users. Sensitivity can be enhanced through addition of biosignals but this presents practical problems (sensor availability, calibration, and privacy restrictions). The current design must then be seen as a user deployable baseline which has the capacity to be stretched in cases where the richness of sensing can be tolerated.

# 6 Conclusion

The paper explored an implementable immersive VR interaction design in a way that interaction is viewed as a closed-loop system and not a set of unrelated techniques. Inspired by the cascading errors and amplification of discomfort encountered in realistic VR settings, including tracking noise, differences at the origin of user motion, and bounded strictness of latency, we integrated an architecture that combined probabilistic intent estimation, stability-conditioned commitment, constraint-based execution, and comfort-sensitive multimodal feedback scheduling. The fundamental design concept, which is late commitment, is the distinction between continuous inference and irreversible action triggers, enabling the system to offer low-cost previews, but committing only when confidence and micro-stability in combination meet a hysteresis gate. The resulting contract of interaction is predictable minimizing carry-over triggers and unintended activations without any apparent lag.

The evaluation framework is built with emphasis on the not only completion time but also the form of errors and corrections, system dynamics (stability, uncertainty, commit latency), and subjective workload/comfort, all as an evaluation. This view is fundamental to immersive HCI, in which perceived efficiency can be overwhelmed by discrete reselect/regras cycles and attentional interference, instead of sheer movement time. It was also explained in the discussion that the benefit of stability gating is most in situations of motion, ambiguity and switching of attention but the gating cannot hope to counter fundamental indiscriminable layouts where posterior uncertainty persists at very high levels. In these situations, designers need to enhance semantic separability, temporalise spatial layout density or come up with explicit confirmation.

The article has evidence-based design implications to sturdy immersive interaction: (i) use unintended commitments and cost of correction as key optimization objectives; (ii) embed uncertainty indicators (e.g., entropy) in commitment logic and not just stronger snapping; (iii) adjust solver gains and release behavior under instabilities to avoid amplifying jitter on manipulation; and (iv) bind comfort control to interaction policy (thresholds and feedback timing), and not just rendering. Long-term adaptation in the event of extended use remains yet to be empirically tested in future work, along with extension to multi-user and dynamic scenes and more rich sensing (e.g., eye tracking, physiological signals) where privacy and feasibility allow.

# References

[1] Maslych, M., Yu, D., Ghasemaghaei, A., Hmaiti, Y., Martinez, E. S., Simon, D., ... & LaViola, J. J. (2024, October). From research to practice: Survey and taxonomy of object selection in consumer VR applications. In 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 990-999). IEEE.

[2] Frata Furlan Peres, F., Nunes, F., Teixeira, J. M., Maurício, C. R., Conceição, K. P., & Yoshida, L. (2024, September). Methods for Evaluating Immersive 3D Virtual Environments: a Systematic Literature Review. In Proceedings of the 26th Symposium on Virtual and Augmented Reality (pp. 140-151).

[3] Friedl-Knirsch, J., Pointecker, F., Pfistermüller, S., Stach, C., Anthes, C., & Roth, D. (2024, June). A Systematic Literature Review of User Evaluation in Immersive Analytics. In Computer Graphics Forum (Vol. 43, No. 3, p. e15111).

[4] Weller, R., Wegele, W., Schröder, C., & Zachmann, G. (2021). Lenselect: Object selection in virtual environments by dynamic object scaling. Frontiers in Virtual Reality, 2, 684677.

[5] Shi, R., Wei, Y., Hu, X., Liu, Y., Yue, Y., Yu, L., & Liang, H. N. (2024). Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays. Proceedings of the ACM on Human-Computer Interaction, 8(ISS), 93-111.

[6] Maslych, M., Hmaiti, Y., Ghamandi, R., Leber, P., Kattoju, R. K., Belga, J., & LaViola, J. J. (2023, March). Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map. In 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 460-470). IEEE.

[7] Chen, D. L., Balakrishnan, R., & Grossman, T. (2020, March). Disambiguation techniques for freehand object manipulations in virtual reality. In 2020 IEEE conference on virtual reality and 3D user interfaces (VR) (pp. 285-292). IEEE.

[8] Steed, A., & Lai, J. (2025). Comparison of hand tracking-based and controller-based interaction in a consumer virtual reality game. Virtual Reality, 29(3), 120.

[9] Khundam, C., Vorachart, V., Preeyawongsakul, P., Hosap, W., & Noël, F. (2021, September). A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In Informatics (Vol. 8, No. 3, p. 60). MDPI.

[10] Zhu, F., Sousa, M., Sidenmark, L., & Grossman, T. (2024, May). PhoneInVR: an evaluation of spatial anchoring and interaction techniques for smartphone usage in virtual reality. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

[11] Bashar, M. R., & Batmaz, A. U. (2024, June). Virtual task environments factors explored in 3d selection studies. In Proceedings of the 50th Graphics Interface Conference (pp. 1-16).

[12] Zhang, C., Chen, T., Shaffer, E., & Soltanaghai, E. (2024, May). FocusFlow: 3D gaze-depth interaction in virtual reality leveraging active visual depth manipulation. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-18).

[13] Kurzweg, M., Weiss, Y., Ernst, M. O., Schmidt, A., & Wolf, K. (2024). Survey on haptic feedback through sensory illusions in interactive systems. ACM Computing Surveys, 56(8), 1-39.

[14] El Saddik, A., Ahmad, J., Khan, M., Abouzahir, S., & Gueaieb, W. (2025). Unleashing creativity in the metaverse: Generative ai and multimodal content. ACM Transactions on Multimedia Computing, Communications and Applications, 21(7), 1-43.

[15] Liu, H., & Gu, Z. (2025). Enhancing Virtual Weight Perception with Material-Informed

Audio-Vibrational Feedback. International Journal of Human–Computer Interaction, 1-16.

[16] Oh, H., & Son, W. (2022). Cybersickness and its severity arising from virtual reality content: A comprehensive study. Sensors, 22(4), 1314.

[17] Mareta, S., Thenara, J. M., Rivero, R., & Tan-Mullins, M. (2022). A study of the virtual reality cybersickness impacts and improvement strategy towards the overall undergraduate students' virtual learning experience. Interactive Technology and Smart Education, 19(4), 460-481.

[18] Cossio, S., Chiappinotto, S., Dentice, S., Moreal, C., Magro, G., Dussi, G., ... & Galazzi, A. (2025). Cybersickness and discomfort from head-mounted displays delivering fully immersive virtual reality: A systematic review. Nurse Education in Practice, 104376.

[19] Groth, C., Magnor, M., Grogorick, S., Eisemann, M., & Didyk, P. (2024). Cybersickness reduction via gaze-contingent image deformation. ACM Transactions on Graphics (TOG), 43(4), 1-14.

[20] Terzioğlu, B., Celikcan, U., & Capin, T. K. (2024). Gaze-contingent adaptation of VR stereo parameters for cybersickness prevention. The Visual Computer, 40(7), 5017-5028.

[21] Monteiro, P., Peixoto, B., Gonçalves, G., Coelho, H., Barbosa, L., Melo, M., & Bessa, M. (2025). Beyond the hands: Evaluating the usability of hands-free methods and controllers for menu selection during an immersive VR experience. International Journal of Human–Computer Interaction, 1-38.

[22] Buckingham, G. (2021). Hand tracking for immersive virtual reality: opportunities and challenges. Frontiers in Virtual Reality, 2, 728461.

[23] Akhoroz, M., & Yildirim, C. (2024, December). Heavy is the Hand: Effects of Hand-Tracking Input and Gestures on Locomotion Performance and User Experience in Virtual Reality. In Proceedings of the International Conference on Mobile and Ubiquitous Multimedia (pp. 61-71).

[24] Yu, D., Dingler, T., Velloso, E., & Goncalves, J. (2024). Object Selection and Manipulation in VR Headsets: Research Challenges, Solutions, and Success Measurements. ACM Computing Surveys, 57(4), 1-34.

[25] Papaefthymiou, S., Giannakopoulos, A., Roussos, P., & Kourtesis, P. (2024, November). Mitigating cybersickness in virtual reality: impact of eye–hand coordination tasks, immersion, and gaming skills. In Virtual Worlds (Vol. 3, No. 4, pp. 506-535). MDPI.

[26] Tzimos, N., Parafestas, E., Voutsakelis, G., Kontogiannis, S., & Kokkonis, G. (2025). Multimodal Interaction with Haptic Interfaces on 3D Objects in Virtual Reality. Electronics, 14(20), 4035.

[27] Palmisano, S., Mcfadyen, M., Miellet, S., Allison, R. S., & Kim, J. (2025). "Differences in Virtual and Physical Head Pose" Predict Cybersickness When Naturalistic Head-Movements are Made in VR. International Journal of Human–Computer Interaction, 1-14.

[28] Ozkan, A., & Celikcan, U. (2023). The relationship between cybersickness and eye-activity in response to varying speed, scene complexity and stereoscopic VR parameters. International Journal of Human-Computer Studies, 176, 103039.

[29] Zhai, X., Sun, Y., Wang, M., Asmi, F., Cai, W., & Chu, X. (2022, May). Exploring the effect of virtual reality with haptics on educational research: A meta-analysis from 2010 to 2020. In 2022 8th International Conference of the Immersive Learning Research Network (iLRN) (pp. 1-7). IEEE.

[30] Zhu, F., Sousa, M., Sidenmark, L., & Grossman, T. (2024, May). PhoneInVR: an evaluation of spatial anchoring and interaction techniques for smartphone usage in virtual reality. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1-16).