One Body, Two Minds: Alternating VR Perspective During Remote Teleoperation of Supernumerary Limbs

Hongyu Zhou
School of Computer Science
The University of Sydney
Sydney, NSW, Australia
hzho4130@uni.sydney.edu.au

Xincheng Huang
Department of Computer Science
University of British Columbia
Vancouver, British Columbia, Canada
xincheng.huang@ubc.ca

Winston Wijaya
The School of Computer Science
The University of Sydney
Sydney, NSW, Australia
wwij0922@uni.sydney.edu.au

Yi Fei Cheng
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
yifeic2@andrew.cmu.edu

David Lindlbauer
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
davidlindlbauer@cmu.edu

Eduardo Velloso
School of Computer Science
The University of Sydney
Sydney, New South Wales, Australia
eduardo.velloso@sydney.edu.au

Andrea Bianchi
Industrial Design
KAIST
Daejeon, Republic of Korea
andrea.whites@gmail.com

Zhanna Sarsenbayeva
School of Computer Science
University of Sydney
Sydney, Australia
zhanna.sarsenbayeva@sydney.edu.au

Anusha Withana
School of Computer Science
The University of Sydney
Sydney, NSW, Australia
anusha.withana@sydney.edu.au

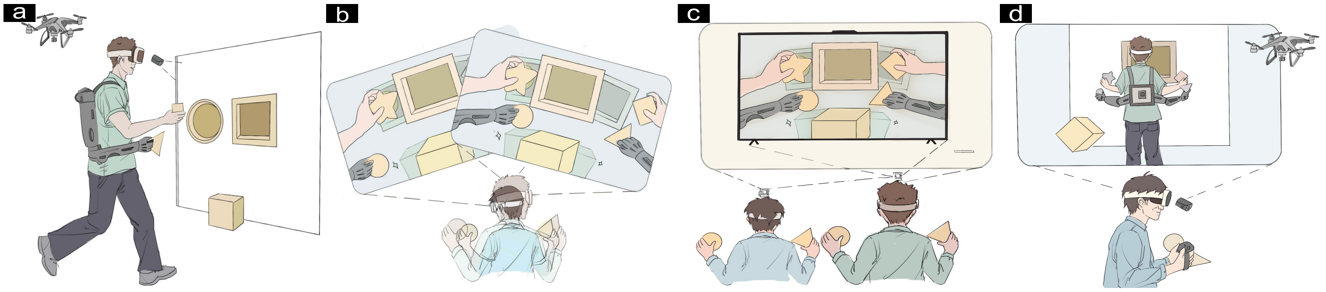

Figure 1: Illustration of three perspectives for remote collaborative teleoperation using virtual supernumerary limbs (VSLs): (a) *Host* user physically walking with VSLs attached to their back; (b) SHARED EMBODIED VIEW: the *guest* shares the *host*'s position with independent head rotation; (c) EMBEDDED ANCHORED VIEW: the *guest* views from the *host*'s head location through a stabilized virtual window; (d) OUT-OF-BODY VIEW: the *guest* independently navigates a drone-like external viewpoint using a handheld controller (the drone serves as a visual metaphor for the decoupled virtual camera).

Abstract

Remote VR teleoperation with supernumerary robotic limbs enables distant users to operate in another's local space. While a shared first-person view aids hand-eye coordination, locking the guest's camera to the host's head can degrade comfort, embodiment, and coordination. Based on a formative study (N=10) using a virtual supernumerary robotic limbs configuration to stress-test coordination, we propose guest-driven perspective switching from a shared first-person baseline (SHARED EMBODIED VIEW) to two alternatives: (a) a stabilized view with guest-controlled rotation (EMBEDDED ANCHORED VIEW), and (b) a fully decoupled third-person view (OUT-OF-BODY VIEW). We ran a user study with 24 pairs (N=48), who switched between the baseline and proposed views as task demands changed. We measured performance, embodiment, fatigue, physiological arousal, and switching behaviors. Our results reveal role-dependent trade-offs: OUT-OF-BODY VIEW improves navigation efficiency and reduces errors, while EMBEDDED ANCHORED VIEW supports embodiment. We conclude with guidelines: use EMBEDDED ANCHORED VIEW for hand-centric adjustments, OUT-OF-BODY VIEW for navigation and object placement, and ensure smooth transitions.

CCS Concepts

• **Human-centered computing → Virtual reality**; **Empirical studies in HCI**; **User studies**.

This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '26, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791433

Keywords

Virtual Reality, User Experience, Supernumerary Robotic Limbs

ACM Reference Format:

Hongyu Zhou, Xincheng Huang, Winston Wijaya, Yi Fei Cheng, David Lindlbauer, Eduardo Velloso, Andrea Bianchi, Zhanna Sarsenbayeva, and Anusha Withana. 2026. One Body, Two Minds: Alternating VR Perspective During Remote Teleoperation of Supernumerary Limbs. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3772318.3791433

1 INTRODUCTION

"When you drift with someone, you feel like there's nothing to talk about." This evocative quote from *Pacific Rim [1]* describes the deep mutual understanding achieved through the fictional mental process that allows two users to co-pilot a single machine by sharing bodily sensations, intentions, and thoughts. It captures the promise of shared bodily control and sensorimotor synchrony. Though fictional, such scenarios are becoming increasingly plausible thanks to advances in telepresence and virtual reality (VR), enabling two individuals to act through a shared body, whether physical or virtual, in tightly coordinated ways [116, 151, 155].

Emerging robotic platforms, such as the humanoid Neo [2], suggest a future where remote operators inhabit complex physical bodies. However, controlling locomotion while simultaneously performing precise manipulation can place high cognitive demands on a single operator [126]. Shared control is emerging in these scenarios, for instance, Multiple Operator, Single Robots (MOSR)[35]. A natural emergence of this is co-embodiment, splitting control between a host and a guest, for instance, between navigation and manipulation. This, in turn, raises a central design question: *viewpoint coupling*, that is, how tightly a guest user's camera is linked to the host's head and body motion, which directly affects how fluidly the pair can collaborate. In most VR systems, the remote first-person view is tied to the local user's head; small head movements therefore propagate to the shared camera, destabilizing near-body control and blurring role boundaries [22, 156]. Simultaneously, advances in AI-based 3D scene reconstruction and world models are making it feasible for teleoperators to work inside navigable virtual replicas of real environments [56, 74, 77, 107, 142]. In such digital twins, anchored or "drone-like" viewpoints can be realised virtually, even while the robot remains grounded in the physical world. Therefore, studying shared viewpoint control in VR offers a way to probe how co-embodied collaborators should manage perspectives for future telepresence systems. In this work, we focus on viewpoint management and shared visualisation, rather than simulating specific robot kinematics or control hardware.

In this paper, we developed a collaborative VR scenario in which two users, a *host* and a *guest*, share control over the same virtual body. To examine viewpoint management under demanding conditions, we augment this body with virtual supernumerary limbs (VSLs). We use this multi-limb configuration not as the sole target application, but as a challenging testbed that maximizes overlap in the near-body workspace, so that insights about viewpoint coupling can generalize to less complex shared-control scenarios. A formative study with five participant pairs identified four empirical patterns: (1) prolonged visual coupling during locomotion co-occurring with *guest* discomfort and blurred self-location; (2) limited perspective flexibility aligning with grasping errors; (3) VSL control instability during the *host*'s head position changes; and (4) role ambiguity leading to *host* pauses. Motivated by these observations, we adopt a task-contingent approach: the *host* remains in a standard HMD first-person view, while the *guest* may switch their viewpoint on demand to align coupling with navigation versus near-body phases.

Building upon these insights, we implemented three perspectives for the *guest*: (i) Shared Embodied View, our design baseline, which co-locates the *guest*'s camera with the *host*'s head position. In this configuration, the *guest*'s head position is locked to the *host*, and moving the *guest*'s head does not change the viewpoint location but only rotates the view direction independently. (ii) Embedded Anchored View shows a stereoscopic portal near the *host*, streaming a steady first-person view for the *guest*; and (iii) Out-of-body View, a world-anchored, fully decoupled 6-DoF third-person camera for spatial awareness. Operationally, the *host* remains in Shared Embodied View throughout. The *guest* can switch between Shared Embodied View and either Embedded Anchored View or Out-of-body View, and our study investigates two *guest*-driven switching strategies: Shared Embodied View ↔ Embedded Anchored View and Shared Embodied View ↔ Out-of-body View.

We then conducted a controlled, within-subjects study with 24 pairs of participants (N=48) to examine how perspective modes affect task performance, embodiment, workload/fatigue, physiology, and guest switching. Results show that Out-of-body View yielded fewer errors in the Factory task with comparable completion times, while participants switched to Embedded Anchored View more frequently for precision-demanding subtasks. Guests reported higher subjective workload and fatigue in Out-of-body View, although HRV indicated lower physiological stress; participants tended to use Embedded Anchored View for precision phases and Out-of-body View for navigation.

Hence, this research offers the following key contributions:

- **Empirical diagnosis.** A formative study diagnosed two systemic breakdowns, coordination friction and guest disorientation, arising from fixed first-person coupling in co-embodied multi-limb VR control.
- **System & evaluation.** We introduce guest-driven perspective switching and evaluate two regimes (Shared Embodied View↔Embedded Anchored View Shared Embodied View↔Out-of-body View) in a controlled within-subjects study (N=48), measuring performance, workload / fatigue, embodiment, physiology, and switching behaviour.
- **Design implications.** We derive guidelines on when to use Shared Embodied View, Embedded Anchored View, and Out-of-body View to manage viewpoint coupling and support comfort, coordination, and embodiment in co-embodied multi-limb VR.

2 RELATED WORK

In this section, we contextualize our research within existing literature on perspective control, remote embodiment, and supernumerary robotic limbs. We highlight key insights and identify the knowledge gaps our study aims to address.

2.1 Perspective Control

Viewpoint selection significantly impacts remote collaboration and coordination, influencing both task performance and user experience. Here, we distinguish coordination from collaboration. Coordination stands for the alignment of actions and timing between collaborators, whereas collaboration refers to the broader process of jointly working toward a shared goal, including planning, decision-making, and mutual understanding [46]. Prior research has explored the effects of different perspectives on remote interactions and their impact on coordination. Galvan et al. [40] demonstrated how alternating between first-person and third-person perspectives can leverage the unique advantages of each viewpoint without significantly harming body ownership of the virtual body. Similarly, Komiyama et al. [80] constructed a system that allows users to freely switch between first-person and third-person views in a remote work setting, enabling precise task execution through first-person images while maintaining situational awareness via third-person viewpoints. Saraiji et al. [116] introduced a wearable system where a teleoperator controls two robotic limbs remotely, using an HMD to observe the environment from a shoulder-mounted perspective, highlighting the importance of viewpoint selection for skill-sharing and collaboration.

However, viewpoint discrepancies between collaborators remain a significant challenge, particularly in object-focused tasks where precise spatial alignment and gesture interpretation are crucial. Nagai et al. [94] explored 360-degree wearable cameras, allowing users to observe and communicate with an entire environment from a remote location, potentially enhancing collaborative spatial awareness. Similarly, Jones et al. [71] highlighted the inherent limitations of handheld perspectives in mobile video chat, as collaborators struggled to perceive shared scenes clearly due to instability and narrow visual fields. Fussell et al. [38] found that scene-focused camera angles outperformed head-mounted perspectives in physical task coordination, as they provided better visibility of the shared workspace. Further, Tang et al. [128] showed that different perspectives optimize different task demands, with shared perspectives aiding textual reading, while asymmetric perspectives facilitate the creation of shared and personal workspaces.

Beyond egocentric and third-person views, prior work proposed techniques like Worlds-in-Miniature (WIM) [125], which offers a scaled-down, allocentric replica of the environment, later adapted for real-world control [119]. Go-Go [110] enables non-linear arm extension to reach distant objects. These methods expand the design space for remote perspective control and inform our investigation of dynamic viewpoint switching.

In summary, prior work shows that flexible viewpoint control can improve spatial awareness and collaboration in VR and telepresence. However, we still know little about how dynamic, user-driven perspective switching shapes coordination, embodiment, and fatigue when two users co-embody a single avatar and jointly control multiple limbs in a shared virtual environment. Our study introduces adaptive guest-driven switching between Shared Embodied View, Embedded Anchored View, and Out-of-body View, to probe the trade-offs in control fluency, comfort, and visual stability under collaborative embodiment.

2.2 Embodiment in Collaboration

In this paper, we distinguish between "self-embodiment", which describes a user's ownership/agency over their own avatar (commonly measured with AEQ [41, 152]), and "remote/co-embodiment" or "social presence", which describe mutual awareness of a partner. These constructs are related but distinct. In our evaluation, we quantify self-embodiment and user experience.

Self-embodiment relates to a user's ownership, agency, and self-location with respect to their own avatar. Remote embodiment refers to visual proxies that convey a collaborator's bodily state and support awareness and coordination. Our study quantifies self-embodiment while using lightweight partner cues to aid reference. Research has explored methods where multiple users share control over a single avatar, using techniques like the weighted-average co-embodiment method, which distributes control percentages between participants [37, 48, 79]. Studies show that increasing control weight improves users' sense of agency, enhancing task coordination [79]. This method has been effective in contexts like VR-based rehabilitation, where even users with less control feel engaged [42, 72]. Further work in multi-operator single robot (MOSR) systems examined how multiple users control a robot, later evolving to include shared control in VR environments with a focus on first-person perspectives and multisensory feedback, affecting psychological aspects like intention alignment [60, 68, 75]. The phenomenon of "enfacement", where users perceive a merged identity, highlights the immersive potential of shared control [118].

Complementary to these control-centric approaches, effective collaboration also depends on how partners are represented and perceived, i.e., embodiment cues that support mutual awareness, reference, and non-verbal coordination. Partner embodiments span a spectrum from minimal surrogates such as the Telepointer [45], through visualisations of partner hands [57, 58, 129, 140, 151], arms [28, 122, 127, 128], and feet [4], to full-body AR/VR telepresence avatars [91, 100, 105]. These representations improve joint attention, reference establishment, and social presence. When gestures alone are insufficient for object reference, raycasting and local scene reconstruction offer stable referents [29, 98].

While remote embodiment is known to support collaboration, its interplay with dynamic perspective control remains underexplored in co-embodied VR scenarios. We examine this relationship in a co-embodied avatar, focusing on how shared-viewpoint schemes shape coordination and embodied experience.

2.3 Supernumerary Robotic Limbs

Research on supernumerary robotic limbs (SRL) has investigated various ways of augmenting human bodily capabilities [55, 111, 115, 134, 153], including additional robotic legs [104], fingers [55], and arms [134]. Our study focuses specifically on virtual supernumerary limbs (VSLs), aiming to expand user capabilities within virtual reality environments.

Initially, SRLs emerged from industrial contexts, designed to assist workers in demanding physical tasks such as construction and assembly operations [101–103]. Over time, the scope of SRL research broadened significantly, encompassing novel designs such as robotic tails to assist balance [90, 93], and extra limbs facilitating complex manipulations [30, 111, 132, 145]. For instance, Maekawa et al. [90] presented a wearable robotic tail that enhanced stability during physically challenging tasks. Control strategies for SRLs have consistently been a central research theme, with multiple intuitive mechanisms proposed to facilitate efficient limb manipulation. These include methods like remapping existing body movements (e.g., utilizing foot or shoulder gestures)[115, 117, 121], and employing brain-computer interfaces (BCI) for direct neural control[106]. While promising, many of these approaches depend heavily on fixed control mappings, which restrict adaptability to dynamic and varying task conditions. Further, comprehensive user evaluations remain scarce, as most studies are limited to concept validations involving single-user scenarios [115, 133, 134]. In collaborative contexts, sharing body control between two users introduces additional demands, such as division of roles, viewpoint negotiation, and real-time coordination, that fixed mappings do not address.

In summary, SRLs and VSLs demonstrate substantial potential, while most prior evaluations emphasize single-user prototypes with fixed mappings. We complement prior work by studying guest-driven switching from Shared Embodied View to Embedded Anchored View or Out-of-body View, improving coordination, efficiency, and comfort when two users co-control one body.

3 FORMATIVE STUDY

To surface early-stage challenges in shared-body collaboration during locomotion, we conducted a diagnostic formative study focused on the limitations of a fixed shared first-person view, which is commonly adopted in prior work [116, 154].

3.1 Method

We recruited five dyads ($N = 10$), each composed of a *host* and a *guest*, all with prior VR experience but no exposure to shared-control systems. This ensured that participants were not distracted by the novelty of VR itself or its basic controls, allowing us to better isolate usability challenges specific to shared-body interaction. Both participants were immersed in the same virtual environment via head-mounted displays. The *host* was embodied as a full-body avatar augmented with a pair of Virtual Supernumerary Limbs (VSLs) attached to their back. The *guest* shared the *host*'s first-person view via a camera rigidly anchored to the *host*'s head position. While the *guest* could not change the camera's spatial location, they retained independent control over view orientation through head rotation. This configuration, which we term the Shared Embodied View, served as the baseline design condition (Figure 2).

Task and Procedure: Participants performed a simplified collaborative task designed to surface visual, operational, and embodiment challenges specific to shared-body teleoperation with VSLs. The task involved the joint transport of small boxes across a 2–3 meter space. The *guest* used VSLs to grasp objects from a table; after a successful grasp, the *host* navigated to a placement zone. Before the main trials, each pair had a short familiarisation period to practise grasping and transporting boxes using the VSLs. They then completed two 5-minute recorded trials. The short duration was intended to foreground early-stage coordination breakdowns and discomfort, rather than capture long-term adaptation effects. We collected observational notes, head rotation logs, VSL activity traces, and think-aloud comments [50, 51]. Post-trial, participants were briefly interviewed on discomfort, control issues, and coordination experiences [25].

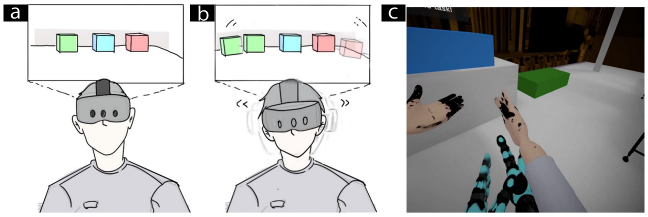

Figure 2: Formative study configuration showing (a) the stable first-person perspective from the *host*, (b) the *guest*'s co-located but independently rotatable view illustrating visual instability, and (c) the avatar embodiment with VSLs used for near-body manipulation.

Measures and Analysis. To identify breakdowns in shared-body coordination under fixed first-person coupling, we used three diagnostic indicators: (i) prolonged head rotations by the *host* (lasting >10 s), which are known to induce discomfort in the *guest* [123, 130]; (ii) grasping errors (misses or drops per attempt) to assess VSL precision [3, 99]; and (iii) self-reported loss-of-control from the *guest* and observed host pauses as markers of coordination breakdown [52]. Two authors independently coded observation notes and interview/think-aloud transcripts using these three indicators as deductive anchors. Discrepancies were resolved through discussion, and related codes were grouped into macro-level themes following reflexive thematic analysis [11, 12, 136]. We triangulated qualitative excerpts with synchronized system logs (e.g., head rotation, VSL activity, locomotion events) to support interpretation [17, 24]. Given the small sample size ($N = 10$), we report dyad-level coverage, representative excerpts, and cross-modal log co-occurrence to inform design, not to claim generalizable causality [92].

3.2 Findings

3.2.1 *Guest Control Drops Without Perspective Flexibility (5/5 dyads).*

Guests reported losing spatial orientation when the *host* was moving. We define this as disorientation during *host* movement that coincides with grasping errors. Participants remarked, "I kept losing where the arms" (*guest*, P04) and "felt like the hands lagged behind what I wanted to do" (*guest*, P08); *hosts* noticed this too ("When I started walking, he said 'wait'," *host*, P03). Grasping errors often occurred shortly after the *host* began moving, according to system logs—suggesting a close relationship, though not a causal one. These findings highlight the need for a viewpoint mode that maintains stable, body-centered alignment during fine motor tasks, allowing *guests* to invoke it as needed for precise control.

View	Camera Anchoring	Orientation	Translation	Render
SHARED EMBODIED VIEW	Host-anchored	Guest head (independent)	No	Full-screen
EMBEDDED ANCHORED VIEW	Decoupled-anchored	Guest head (linked)	No	Stabilized portal
OUT-OF-BODY VIEW	World-anchored (free-floating)	Guest head / controller	Yes (6-DoF)	Full-screen

Table 1: Differences across views along two axes: position anchoring and view control. Host always stays in SHARED EMBODIED VIEW; the guest switches between SHARED EMBODIED VIEW and one alternative per condition.

3.2.2 *Hosts Stop Moving Due to Role Confusion (3/5 dyads).*

Following prior work on detecting stillness in movement trajectories and HMD logs [9, 137], we identified *host* pauses as periods of minimal movement lasting several seconds during ongoing tasks. *Hosts* often attributed these pauses to uncertainty about whether continued movement would interfere with the *guest*'s manipulation. Participants remarked, "I kept stopping, unsure if moving would disrupt them" (*host*, P03) and "I waited here until he finished" (*host*, P05), with *guests* corroborating this dynamic ("I needed to say 'don't move"' *guest*, P06). These co-occurrences suggest hesitation rooted in unclear role boundaries, rather than deliberate coordination.

3.2.3 *Prolonged Shared View Causes Cybersickness (5/5 dyads).*

When the *guest* shared the host's first-person view, extended walking or turning by the host often coincided with verbal reports of dizziness, disorientation, or difficulty controlling the arms. Participants described feeling dizzy or unsure of their location after several seconds of walking or turning, e.g., "long turns made me dizzy" (*guest*, P02), and "I [kept] losing where the arms [were], and felt nauseous" (*guest*, P06). Discomfort did not arise from sudden turns alone, but more often from sustained motion lasting over 10 seconds, even when the motion was low in intensity. In the system logs, these episodes overlapped with sustained head rotations or locomotion events. Grasping errors were also more likely to occur shortly after such movement sequences. Although we make no causal claims, this repeated co-occurrence points to a possible mismatch between visual motion and control stability during prolonged shared viewing.

3.2.4 *Camera Shifts Disrupt Guest's VSL Control (4/5 dyads).*

When the *guest* shared the host's camera view, sudden or sharp head turns by the *host* often disrupted the *guest*'s ability to control the virtual arms. Participants described losing targeting accuracy or feeling disoriented during these transitions, such as "Hard to keep control when everything shifts" (*guest*, P06) and "every turn jolted my aim" (*guest*, P02). Triangulating with locomotion logs, these breakdowns frequently occurred within short windows after *host* heading-change events, evidence of time-adjacent co-occurrence rather than causality.

3.2.5 *Summary of Findings.*

Our formative study revealed recurring difficulties with a fixed shared first-person view. First, *guests* often lost track of hand position during precise actions, especially when *hosts* moved their head mid-task, leading to grasping errors and hesitation. At the same time, *hosts* frequently paused movement, unsure whether the guest was in control, reflecting ambiguity in coordination roles. Second, shared locomotion episodes where the *host* moved while the *guest* remained visually coupled caused discomfort, disorientation, and disengagement for the guest. These issues clustered into two under-explored challenge areas: (C1) coordination breakdowns during precision tasks, and (C2) discomfort and control loss from shared movement. System logs confirmed that most errors clustered around host motion: about 71% of logged errors (45/63) occurred within 4 s of walking onset or a sharp head turn. These findings highlight the need to restore guest perceptual agency and reduce coordination friction during task transitions.

4 DESIGN AND IMPLEMENTATION

To address the issues surfaced in our formative study, we designed two guest-driven perspective modes that directly target the observed challenges. In C1, coordination broke down during precise manipulation due to ambiguous visual control. To address this, we introduce the EMBEDDED ANCHORED VIEW, which grants the guest independent rotational control of the viewpoint. This stabilizes the visual frame during fine-grained tasks and helps clarify interaction roles. In C2, guests experienced discomfort and disorientation caused by involuntary shared locomotion. To address this, we implemented the OUT-OF-BODY VIEW, a third-person, freely positionable camera that decouples the guest's perspective from the host's body, easing spatial confusion and supporting higher-level navigation decisions. The specific distinctions between these perspective modes are summarized in Table 1.

4.1 Embedded Anchored View (C1)

We implemented EMBEDDED ANCHORED VIEW as a body-anchored yet independently rotatable viewpoint, as shown in Figure 4 (a). The system renders a stabilized egocentric feed as a stereoscopic portal, positioned at a fixed offset from the host and oriented toward the host's head-camera direction. This provides visual stability while preserving directional control. We also apply light IPD tuning and frame smoothing to support accurate manipulation.

Previous research has demonstrated improved depth perception and task accuracy with slight inter-pupillary distance (IPD) adjustments in VR [47, 97]. Thus, we slightly reduced the guest's effective IPD within the EAV window to enhance binocular depth cues, facilitating precise grasping and alignment tasks.

Moreover, as rapid *host* movements can disrupt the *guest*'s reference and degrade VSL stability [19], we employed Unity SmoothDamp stabilization (exponential smoothing) on EAV camera updates to soften *host* jitter. This preserves a steady body-anchored reference during *host* motion, improving comfort and control precision.

Finally, to reinforce role clarity, we emphasized visual embodiment cues known to reduce hesitation in collaboration [81]. While the shared avatar is visible in all modes, EMBEDDED ANCHORED VIEW shows a steady 3D window that keeps the host's hands, arms,

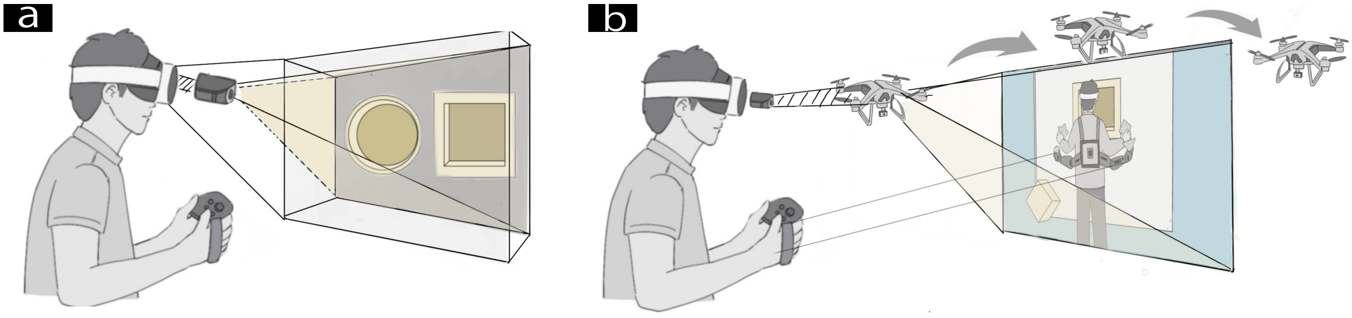

Figure 3: Illustration of the Embedded Anchored View and Out-of-body View. (a) The *guest* user's viewpoint is spatially aligned with the *host*'s viewpoint but rendered as a video-like window. (b) The *guest* user independently controls a virtual drone camera with 6 degrees-of-freedom (6-DoF), enabling flexible positional and rotational adjustments independent of the *host*'s movements (the drone serves as a visual metaphor for the decoupled virtual camera).

and VSLs visually locked to the body with correct in-front/behind relationships, making the control locus near the body unambiguous.

4.2 Out-of-body View (C2)

Out-of-body View provides a fully decoupled, world-frame third-person camera with 6-DoF motion (rotation and translation), independent of the *host*'s movement, as shown in Figure 4 (b). The *guest* controls the drone camera via joystick inputs. We apply smooth interpolation and bound speed/acceleration, with collision checks to prevent disruptive motion.

To improve the *guest*'s spatial awareness and viewpoint flexibility, we implemented a 6-degree-of-freedom (6-DoF) virtual drone camera. The *guest* controlled positional (forward, backward, lateral, vertical) and rotational (pitch, yaw, roll) movements using intuitive VR joystick inputs. Smooth camera transitions were ensured through spherical linear interpolation (Slerp) and linear interpolation (Lerp), enhancing visual comfort and spatial understanding.

Stabilizing VSL control during *host*'s movements required a stable spatial reference frame. Thus, we implemented the drone camera in a fixed world coordinate system, independent of *host* movement [78]. Additionally, real-time collision detection and spatial constraints via Unity's raycasting prevent visual disruptions and maintain stable associations with the shared environment and avatar [13].

4.3 Perspective Switching Mechanism

In our system, Shared Embodied View is the default host-centric view and serves as the baseline from which other perspectives are invoked. To overcome limitations of this fixed shared view [8, 16], we implement a *guest-driven* perspective-switching mechanism that lets users fluidly alternate between views based on task demands.

This mechanism allows the *guest* to transition from Shared Embodied View to a condition-specific alternative view using VR controller input, which activates or deactivates dedicated Unity camera components. To preserve comfort and spatial context, each switch is rendered with short cross-fades, input debounce, and on-screen indicators. The design enables the *guest* to time perspective transitions to task needs: invoking Embedded Anchored View's stabilization, body-aligned portal for fine manipulation, or Out-of-body View's fully decoupled third-person view for spatial planning. This mechanism directly supports user agency, reduces prolonged visual coupling, and maintains stable VSL control during host movement—addressing the challenges identified in our formative study.

4.4 VSLs Independent Control and Stabilization

Formative findings indicated that VSL control became fragile under *host* locomotion and heading changes. We therefore decouple VSL actuation from *host* head/body transforms and drive VSL poses in a world-referenced frame initialized when the *host* is stationary. Using Meta Quest 3 tracking, we monitor head orientation and translation, and flag short-window spikes in angular velocity or position delta as "significant motion." When flagged, a stabilization mode engages: VSLs are either held in place or eased back to a predefined neutral pose via Unity `SmoothDamp` to avoid abrupt shifts. Guest input immediately disables stabilization and enters active control to preserve precision and responsiveness. This mechanism mitigates motion-induced disruptions without altering the *host*'s baseline view and applies uniformly under Shared Embodied View and Embedded Anchored View (the difference is rendering only).

4.5 Implementation Details

We developed the system in Unity 2022.3.7f1 and deployed it on two Meta Quest 3 headsets. Both users share one anthropomorphic avatar in VR: the *host* navigates and controls the primary arms, and the *guest* operates the back-mounted VSLs. Each headset was tethered via USB to a dedicated PC on the same local network, achieving latency below 5ms [32]. Multiplayer synchronization used Photon Unity Networking with a locally hosted Photon Server to minimize network hops. The *host* and *guest* were located in separate rooms to preserve a remote collaboration context while enabling natural voice communication via audio only. Perspective control, VSL manipulation, and task logic were integrated into the Unity scene. On-screen indicators show the current mode. A logging module records head kinematics, perspective switches, VSL commands, and grasp events.

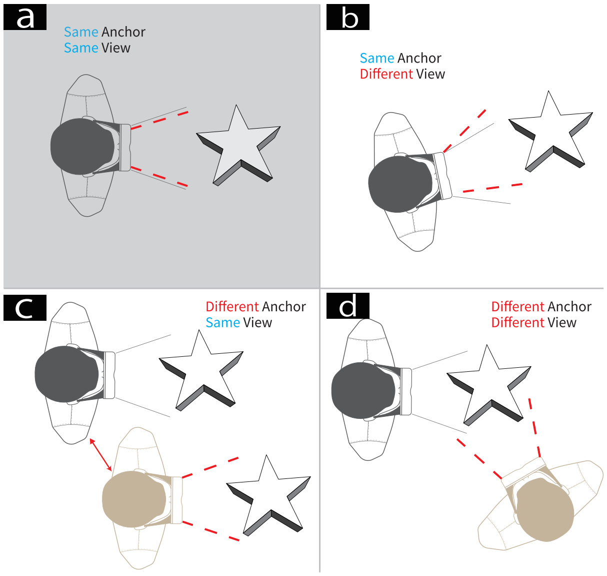

Figure 4: Guest perspectives organized as an *Anchor–View* matrix relative to the host. (a) Locked Egocentric: *same anchor, same view*; guest camera is rigidly locked to the host's head pose and view. (Shown for comparison; not studied, as it grants the guest no agency and renders them a passive observer under full host control). (b) Shared Embodied View: *same anchor, different view*; guest shares the host's position but rotates the camera independently. (c) Embedded Anchored View: *different anchor, same view*; guest remains independently positioned but views a stabilized window of the host's egocentric feed. (d) Out-of-body View: *different anchor, different view*; guest drives a free-floating 6-DoF drone camera, decoupled from the host.

5 EXPERIMENTAL METHOD

We ran a within-subjects experiment to examine how two guest-controlled view-switching conditions shape user strategies and collaborative behaviors when operating VSLs with a shared-body avatar. Both tasks required tight coordination between the *host* and *guest*, emphasizing precision, timing, and role-dependent experience in dynamic VR scenarios.

To align with our revised research scope, we address the following RQs:

- **RQ1**: How do Embedded Anchored View and Out-of-body View affect subjective workload, fatigue, and physiological responses during collaboration?
- **RQ2**: How do Embedded Anchored View and Out-of-body View influence self-embodiment (AEQ) for hosts and guests?
- **RQ3**: How do Embedded Anchored View and Out-of-body View affect task performance (completion time, errors) in dynamic shared-body collaboration when switching is initiated from the Shared Embodied View?
- **RQ4**: When and why do *guests* switch between Shared Embodied View ↔ Embedded Anchored View or Shared Embodied View ↔ Out-of-body View as task demands evolve, and how are these switches associated with performance and experience?

5.1 Experimental Conditions and Procedure

5.1.1 Experimental Conditions. We tested two conditions: Shared Embodied View↔Embedded Anchored View and Shared Embodied View↔Out-of-body View. In both conditions, tasks began in Shared Embodied View; the *guest* could switch to the condition-specific alternative, while the *host* remained in Shared Embodied View throughout. Each session comprised two task runs. Task type and condition order were fully counterbalanced across participant pairs to mitigate learning effects and fatigue, ensuring each pair experienced one run per condition.

5.1.2 Participants. We recruited 24 participant pairs (48 participants; 24 female, 24 male; M=22.8 yrs, SD=3.2 yrs; no color vision deficiencies). This sample size aligns with common practice in HCI research involving collaborative VR systems [14]. All participants had prior experience with VR [21]. 16 participants reported regular VR use, while 32 primarily engaged with VR for entertainment. Each participant pair consisted of a *host* and a *guest*.

5.2 Experimental Tasks

5.2.1 Transportation Task. The first task simulated a mobile manipulation scenario commonly encountered in collaborative work and human-robot collaboration environments [23, 112, 139, 141], as shown in Figure 5 (a,b). In each trial, the participant pair (host and guest) started from a designated starting location. Their goal was to collaboratively transport colour-coded boxes placed centrally on a table into corresponding target bins distributed across a virtual room approximately 6×8 meters in size, populated with static obstacles such as tables, walls, and barriers. The *host* navigated the virtual environment freely by physically walking in the tracking space, ensuring balance, obstacle avoidance, and efficient positioning, while the *guest* remotely controlled the VSLs using VR controllers, pressing specific buttons to grasp and release boxes accurately, a role distribution consistent with prior collaborative robotics studies [39, 146]. To enforce joint action, the boxes were designed to be large enough to require two hands to lift and carry, mimicking the size and weight of bulky real-world objects. While two hands stabilized the load, the remaining two were required to interact with the environment by pushing aside obstacles or lifting spring-loaded bin covers, creating situations in which all four hands had to operate simultaneously.

This task required continuous collaboration: the *host* needed to provide optimal positioning for the *guest*'s manipulation tasks, whereas the *guest* had to adapt to continuous perspective shifts caused by the *host*'s movement. The design intentionally introduced continuous demands for spatial navigation and fine-grained manipulations, enabling systematic investigation of the impacts of dynamic viewpoint control, cognitive load, and collaborative efficiency in shared-body remote teleoperation scenarios.

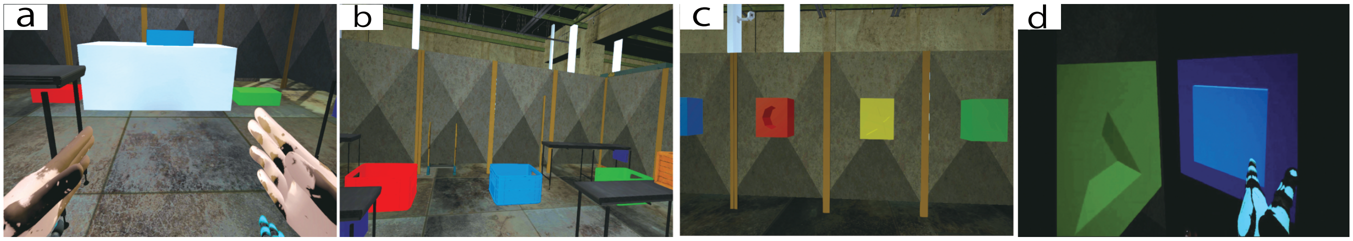

Figure 5: Task overview: Transportation and Factory. Panels (a–b) depict the *Transportation* task: (a) colour-coded boxes placed on a central table; (b) matching colour-coded bins distributed throughout the VR environment. Panels (c–d) depict the *Factory* task: (c) geometrically distinct holes on a wall panel as precise insertion targets; (d) collaborative insertion of objects (e.g., cubes) into the corresponding holes.

5.2.2 Factory Task. The second task required participants to perform precision object manipulation and insertion tasks collaboratively, emphasizing precise spatial coordination and multi-limb synchronization [144], as shown in Figure 5 (c,d). In each trial, participants simultaneously controlled the avatar's four hands (two controlled by the *host* and two by the *guest*) to insert four geometrically distinct objects (e.g., cubes, cones, spheres, and cylinders) into corresponding matching holes on a virtual wall panel. Each insertion required precise alignment and steady positioning to succeed; insertion errors included misalignment, improper orientation, or unsuccessful attempts. The *host* controlled the avatar's locomotion and two primary arms, while the *guest* controlled the two extra arms. Clear visual feedback indicated successful insertion. To encourage multi-limb synchronization, we designed a timed batch-insertion task. All four target holes were always available, and participants were asked to complete all insertions before the trial time limit expired. The pick-up area and insertion panel were spatially separated, encouraging participants to carry multiple objects per trip when possible. Using all four hands in parallel helped reduce walking, minimize idle time between insertions, and improve overall efficiency. While not strictly required, coordinating all four hands became the most effective strategy for completing the task on time. The task specifically aimed to evaluate how dynamic perspective-switching influenced precision tasks requiring fine motor coordination, collaborative synchronization, and role-specific embodiment under varying viewpoints.

5.3 Experimental Data Collection

To evaluate task performance and switching strategies, we collected quantitative data covering subjective embodiment, workload and fatigue, performance metrics, physiological responses, and qualitative data through interviews and think-aloud protocols.

5.3.1 Qualitative Data. We recorded audio and video during the think-aloud session and post-task semi-structured interviews as qualitative data.

5.3.2 Performance Measures. Task performance metrics included completion time, error rate, and perspective-switching behaviors. In the Transportation Task, we recorded completion time (from grasp initiation to successful bin placement) and errors (failed grasps, drops, or wrong-bin placements). In the Factory Task, we measured insertion time per object and recorded insertion errors, which included both translational misalignments (incorrect positional offsets relative to the target hole) and rotational misalignments (incorrect orientation of objects relative to the required target orientation). Participants received no performance feedback during trials.

5.3.3 Subjective Feedback. We administered three standardized questionnaires:

The Avatar Embodiment Questionnaire (AEQ) measured self-embodiment, consistent with prior work on VR embodiment [41]. Following prior SRL-in-VR work that administers AEQ in a pre/post-adaptation schedule [6], we followed the same timing logic and extended it to our two perspective-switching strategies by collecting one pre-adaptation AEQ under SHARED EMBODIED VIEW and two post-adaptation AEQ: one after SHARED EMBODIED VIEW→ EMBEDDED ANCHORED VIEW and one after SHARED EMBODIED VIEW → OUT-OF-BODY VIEW.

NASA Task Load Index (NASA-TLX) assessed perceived workload (mental, physical, temporal, effort, performance, frustration) [49]. Visual Analog Scale for Fatigue (VAS-F) measured subjective fatigue across 18 dimensions [120]. While established measures of cybersickness such as the Simulator Sickness Questionnaire (SSQ) [75] are available, in this first study we focused on workload (NASA–TLX) and fatigue (VAS–F) as our primary subjective indicators of discomfort in shared viewpoints, to limit questionnaire burden and align the scales with our task demands. We acknowledge the absence of a dedicated cybersickness instrument as a limitation and see it as an important addition for future work.

5.3.4 Physiological Data - Heart Rate Monitoring. We continuously recorded heart rate data using a Polar H10 chest strap [135], known for accuracy and minimal interference with VR [83, 114]. Specifically, we analyzed heart rate variability (HRV) and peak heart rate during intensive task phases and critical perspective-switching events to quantify physiological correlates of subjective experience [5, 59]. RR-interval series were preprocessed following psychophysiology recommendations [85]. We removed artefact-contaminated/ectopic intervals, interpolated only short gaps, and computed RMSSD in sliding windows during active task periods (full parameters in Appendix A). Heart rate variability (HRV) was computed using the time-domain RMSSD index (root mean square of successive differences between adjacent NN intervals). Higher values indicate lower physiological stress.

5.3.5 Procedure. We adopted a fully within-subjects, repeated-measures design over two consecutive days. Each dyad experienced both collaborative tasks (Transportation and Factory) and both switching regimes (Shared Embodied View→ Embedded Anchored View and Shared Embodied View→ Out-of-body View). Tasks were split across days such that each day focused on one task, and task order was balanced across dyads (half started with Transportation on Day 1 and Factory on Day 2, and the remainder received the opposite order). Within each task, the order of the two switching regimes was also balanced across dyads so that each regime appeared equally often in the first and second position. For each task, the number of targets, spatial layout, and time limits were held constant across perspective conditions; only the guest's viewpoint control differed, making task difficulty comparable between Embedded Anchored View and Out-of-body View.

Participants first completed a demographics form, received a study briefing, and provided informed consent. We then introduced the switching modes, control mappings, and role division: the *host* managed avatar locomotion and native arms and remained in Shared Embodied View; the *guest* controlled the virtual supernumerary limbs (VSLs) and initiated perspective switches, which occurred only between Shared Embodied View and Embedded Anchored View or Out-of-body View.

Participants completed a 5-minute familiarization session in Shared Embodied View to ensure fluency before entering the main conditions. We then administered the Avatar Embodiment Questionnaire (AEQ) to establish pre-condition embodiment. Following prior SRL-in-VR work [6], we administered AEQ after each condition to obtain post-adaptation scores. This enabled both pre–post contrasts and condition-level comparisons of self-embodiment.

Each daily session began with five minutes of calming music ("Zen mode soundtrack" [10]) to minimize participant fatigue, following a similar approach used in prior research examining sensory conflicts and cognitive fatigue in VR [89]. Short breaks were offered between blocks as needed, and although we did not model learning or fatigue effects explicitly, we did not observe obvious performance degradation over time. Within each daily session, participants performed the assigned task twice, once under each of the two perspective conditions (Embedded Anchored View and Out-of-body View), with the presentation order of these conditions also counterbalanced. During task execution, the host controlled the avatar's locomotion and physical arms, while the guest operated the VSLs. Verbal communication was encouraged to allow participants to articulate strategies and decisions.

At the end of both perspective conditions, participants removed their VR headsets and completed the NASA Task Load Index, the Visual Analog Scale for Fatigue, and the AEQ. The study received ethics clearance from the Human Research Ethics Committee (HREC) of the University of Sydney (2019/553).

5.4 Data Analysis

We analysed questionnaire, performance, and physiological measures using linear and generalised linear mixed-effects models. For approximately normal continuous outcomes (e.g., AEQ factors, NASA–TLX, VAS–F, completion times) we used Gaussian models with identity link; for counts and binary outcomes (e.g., errors, trial success) we used Poisson or binomial models with log/logit links. Unless noted, models included fixed effects of Perspective and, where applicable, User Role and Task, and random intercepts for participant and dyad plus random slopes for within-subject factors when supported by the data. Model diagnostics (residuals, Q–Q plots, overdispersion) and full formulas are reported in Supplementary Materials. Results in the main text are summarised as estimated marginal means (EMMs) with 95% confidence intervals, standardised effect sizes, and p-values for planned contrasts.

To characterise the sensitivity of our design, we conducted a G*Power [33] analysis. For the planned within-subject contrasts between Embedded Anchored View and Out-of-body View ($\alpha = .05$, power $= .80$), our sample of $N = 48$ provides sensitivity to paired-samples effects of approximately $d_z \approx 0.41$. For the AEQ analysis with three Perspective levels (SEV, EAV, OOB), a repeated-measures ANOVA approximation (correlation among repeated measures $r = .50$, $\varepsilon = 1.0$) yields a detectable within-subject main effect of $f \approx 0.20$. For between-role comparisons (Host vs. Guest; $n = 24$ per role), an independent-samples approximation indicates 80% power to detect effects of roughly $d \approx 0.80$ at $\alpha = .05$.

6 RESULTS

We report quantitative and qualitative findings addressing our four research questions (RQs). RQ1 investigates how perspectives impact subjective workload, fatigue, and physiological responses. RQ2 explores the effects of visual perspectives on embodiment. RQ3 examines how different perspective strategies affect task performance, including completion times and error rates. RQ4 analyzes strategic switching behaviours in response to evolving task demands.

Because of the within-subjects design with repeated measures, we used Generalized Linear Mixed Models (GLMMs) to account for random effects. We refer to *EAV* and *OOB* as the two dynamic switching conditions (Shared Embodied View→Embedded Anchored View, Shared Embodied View→Out-of-body View). For all RQs except RQ2, we focus inferential statistics on EAV vs. OOB; SEV is plotted as a baseline but excluded from planned pairwise comparisons unless noted. Results are presented as estimated marginal means (EMMs) with 95% confidence intervals (CIs). Full model details are available in the Supplementary Materials.

6.1 Physiological Responses Measures (RQ1)

Complementing our subjective findings (RQ1), we report physiological data (HR, HRV) summarized in Figure 6. GLMM analyses revealed significant effects of Perspective, Task, and Role on both peak heart rate and HRV (RMSSD), including Perspective × Task and Perspective × Role interactions for peak heart rate (e.g., Perspective main effect for *hosts* in the Transportation baseline: $\beta = -23.0$ bpm, 95% CI [−25.4, −20.6]; Perspective×Task: $\beta = 16.2$ bpm, 95% CI [13.3, 19.0]; Perspective×Role: $\beta = 22.1$ bpm, 95% CI [19.2, 24.9]).

Peak Heart Rate. Participants exhibited significantly elevated peak heart rates in Out-of-body View compared to Embedded Anchored View during the precision-oriented Factory task. This effect was more pronounced for *guests* ($M_{OOB} = 102.6$ bpm, 95% CI [100.2, 104.9]; $M_{EA} = 87.3$ bpm, 95% CI [85.0, 89.7]; $\Delta M = 15.3$ bpm, $p < .05$). For *hosts*, Factory showed a small but reliable elevation under Embedded Anchored View relative to Out-of-body View ($M_{EA} = 115.4$ bpm, 95% CI [113.0, 117.8]; $M_{OOB} = 108.6$ bpm, 95%

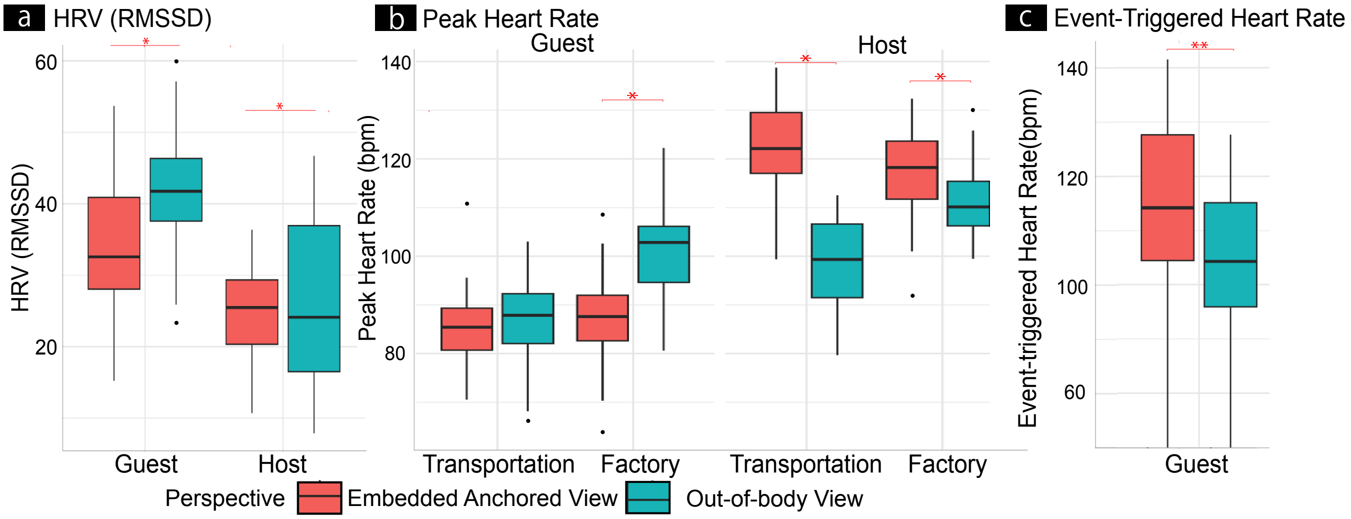

Figure 6: Physiological responses for *hosts* and *guests* under Embedded Anchored View and Out-of-body View across two different tasks. (a) HRV (RMSSD), a time–domain index of heart-rate variability. (b) Peak heart rate (bpm) by role and task. (c) Event-triggered heart rate aligned to perspective switches from the Shared Embodied View to the target view (Shared Embodied View → Embedded Anchored View or Shared Embodied View → Out-of-body View). * indicates p < .05 and ** indicates p < .01.

CI [106.2, 110.9]; ΔM = 6.8 bpm, $p < .05$). In the navigation-focused Transportation task, *hosts* showed significantly higher peak heart rates in Embedded Anchored View (M_{EA} = 118.4 bpm, 95% CI [116.1, 120.8]) than Out-of-body View (M_{OOB} = 95.4 bpm, 95% CI [93.1, 97.8]; ΔM = 23.0 bpm, $p < .05$), while *guests* showed no reliable difference (M_{EA} = 90.4 bpm, 95% CI [88.0, 92.7] vs. M_{OOB} = 89.4 bpm, 95% CI [87.1, 91.8]; ΔM = 1.0 bpm, $p > .05$).

Heart Rate Variability (RMSSD). We observed a significant Perspective × Role interaction (β = 5.0 ms, 95% CI [0.4, 9.6]). For *guests*, RMSSD was higher in Out-of-body View (M = 42.5 ms, 95% CI [40.2, 44.8]) than in Embedded Anchored View (M = 35.7 ms, 95% CI [33.4, 38.0]; ΔM = 6.8 ms, $p < .05$), indicating lower physiological stress in the observer perspective. For *hosts*, RMSSD was also higher in Out-of-body View than in Embedded Anchored View (M_{OOB} = 27.3 ms, 95% CI [25.0, 29.6]; M_{EA} = 25.4 ms, 95% CI [23.2, 27.7]), although the magnitude of the difference was modest.

Event-triggered heart-rate analyses confirmed a significant immediate increase when switching from Embedded Anchored View to Out-of-body View (+11.1 bpm, 95% CI [9.0, 13.2], $p < .01$), highlighting the anticipatory cognitive demand associated with viewpoint adjustments.

These physiological results corroborate our subjective findings, showing that Out-of-body View reduced physiological stress and arousal compared to Embedded Anchored View, particularly for guests. This supports RQ1 by highlighting how observer perspectives can mitigate workload during collaboration.

6.2 Workload and Fatigue Analysis

This analysis addresses RQ1 by comparing subjective workload and fatigue between Embedded Anchored View and Out-of-body View across task and role conditions, as shown in Figure 7. GLMMs revealed significant Perspective × Task and Perspective × Role interactions for workload (NASA-TLX). During the Factory task, both *hosts* and *guests* reported higher workload in Out-of-body View compared to Embedded Anchored View (p<.05 for *hosts*, p<.001 for *guests*; Δ=12.03, 95% CI [8.02, 16.04] for *hosts*; Δ=17.35, 95% CI [13.27, 21.43] for *guests*; *hosts*: Out-of-body View = 57.60 ± 5.87, Embedded Anchored View = 45.57 ± 4.29; *guests*: Out-of-body View = 54.58 ± 5.06, Embedded Anchored View = 37.23 ± 5.30). For the Transportation task, no reliable workload difference was observed for *guests* (p>0.10; Δ=4.77, 95% CI [0.95, 8.58]; Out-of-body View = 48.95 ± 5.66, Embedded Anchored View = 44.18 ± 4.52), whereas *hosts* reported significantly higher workload in Out-of-body View than Embedded Anchored View (p<.05; Δ=4.82, 95% CI [4.82, 4.82]; Out-of-body View = 53.17 ± 5.16, Embedded Anchored View = 48.35 ± 4.95).

For subjective fatigue (VAS-F), *guests* reported significantly higher fatigue in Out-of-body View than Embedded Anchored View during both Factory (p<.05; Δ=3.25, 95% CI [0.05, 6.45]; Out-of-body View = 22.65 ± 4.99, Embedded Anchored View = 19.40 ± 2.67) and Transportation (p<.001; Δ=2.67, 95% CI [0.25, 5.09]; Out-of-body View = 27.22 ± 4.97, Embedded Anchored View = 24.55 ± 5.11). For *hosts*, we observed significantly higher fatigue in Embedded Anchored View than Out-of-body View during Transportation (p<.05; Δ= − 16.46, 95% CI [-19.84, -13.08]; Embedded Anchored View = 42.67 ± 4.92, Out-of-body View = 26.21 ± 3.71), and a marginally higher fatigue during Factory, although the effect did not reach significance (p>0.05; Δ= − 12.96, 95% CI [-15.29, -10.63]; Embedded Anchored View = 37.77 ± 3.35, Out-of-body View = 24.81 ± 4.99).

Together, these results support RQ1 by showing that Out-of-body View can increase perceived workload and fatigue for guests,

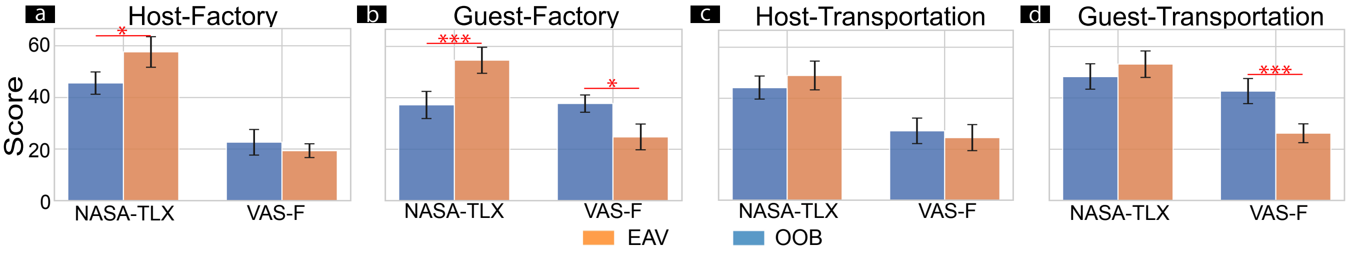

Figure 7: Subjective measures (VAS-F and NASA-TLX) for *hosts* and *guests* across tasks: (a) *Host* scores for Factory Task, (b) *Guest* scores for Factory Task, (c) *Host* scores for Transportation Task, and (d) *Guest* scores for Transportation Task. Scores are compared between Embedded Anchored View and Out-of-body View. * indicates $p < .05$, ** indicates $p < .01$ and *** indicates $p < .001$.

whereas hosts experience more fatigue under Embedded Anchored View, particularly in navigation-intensive tasks. This highlights a role-dependent divergence in how perspective affects burden.

Fatigue (VAS–F). For *guests*, fatigue was higher in Out-of-body View than in Embedded Anchored View in the *Factory* task ($p<.05$; $\Delta=3.25$, 95% CI [0.05, 6.45]), and showed the same direction without reaching significance in *Transportation* ($p>0.10$; $\Delta=2.67$, 95% CI [0.25, 5.09]), consistent with Fig. 7(b)(c). For *hosts*, fatigue was higher in Embedded Anchored View than in Out-of-body View during *Transportation* ($p<.05$; $\Delta = -16.46$, 95% CI [-19.84, -13.08]), and showed no reliable difference in *Factory* ($p>0.10$; $\Delta = -12.96$, 95% CI [-15.29, -10.63]), matching Fig. 7(a)(d).

6.3 Embodiment Questionnaire Analysis (RQ2)

RQ2 examines how perspective-switching strategies shape subjective embodiment. AEQ ratings were collected immediately before and after each perspective-switch condition, following standard practice in VR embodiment research [6]. Factor scores were computed as the mean of their constituent items:

$$\text{Factor Score} = \frac{\sum \text{Item Scores}}{\text{Number of Items per Factor}}.$$

We analysed factor scores with linear mixed-effects models (lme4 [7]; lmerTest [84]), including fixed effects of Perspective, User Role (host vs. guest), and Embodiment Factor, and by-participant random intercepts with random slopes for within-subject factors. Degrees of freedom for fixed effects were estimated via the Kenward–Roger method [76]; omnibus tests are reported as Type-III F statistics, and post-hoc pairwise contrasts were computed with *emmeans* [86] using Tukey adjustment [54].

We observed a main effect of Perspective ($p < .001$) and Embodiment Factor ($p < .01$), and a Perspective×User Role interaction ($p < .001$). Post-hoc comparisons aligned with Fig. 8: across factors, Shared Embodied View yielded the highest embodiment, Embedded Anchored View was intermediate, and Out-of-body View was lowest. For *guests*, body ownership and presence were higher in Shared Embodied View than in Out-of-body View ($M_{\text{SEV}} = 4.47$, $SD = 0.97$ vs. $M_{\text{OOB}} = 3.52$, $SD = 1.04$; both Tukey $p < .001$). corresponding to a large standardised mean difference of approximately $d = 0.94$, 95% CI $[0.34, 1.53]$. For *hosts*, agency was higher in Shared Embodied View than in Embedded Anchored View ($M_{\text{SEV}} = 4.26$, $SD = 0.94$ vs. $M_{\text{EAV}} = 3.72$, $SD = 1.01$; Tukey $p < .001$); corresponding to a moderate standardised mean difference of approximately $d = 0.60$, 95% CI $[0.37, 0.82]$.

These findings answer RQ2 by showing that self-embodiment is perspective- and role-dependent: Out-of-body View reduces both presence and ownership, particularly for guests, while Embedded Anchored View improves embodiment relative to Out-of-body View, but remains lower than the Shared Embodied View.

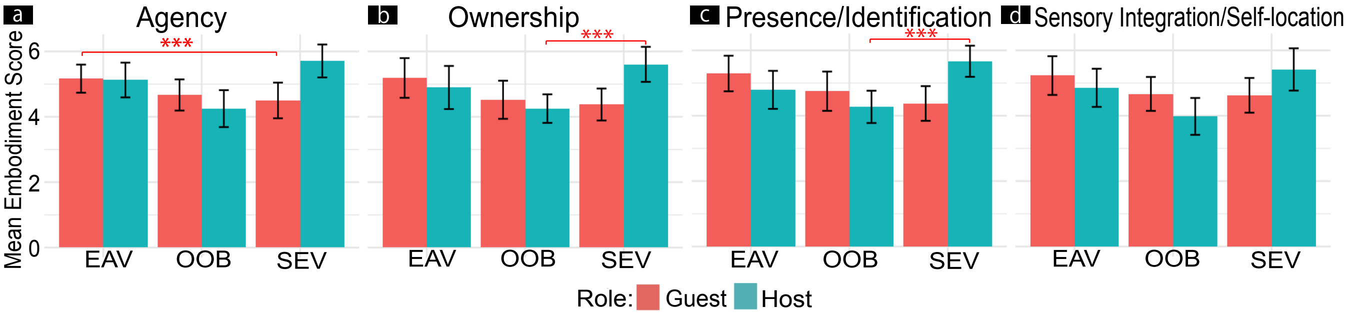

Figure 8: Mean embodiment scores (±SE) across three perspective conditions (Embedded Anchored View, Out-of-body View, and Shared Embodied View), separated by user role (Host and Guest). Subplots represent AEQ factors: (a) Agency, (b) Body Ownership, (c) Presence/Identification, and (d) Sensory Integration/Self-location. * indicates $p < .05$, ** indicates $p < .01$ and *** indicates $p < .001$.

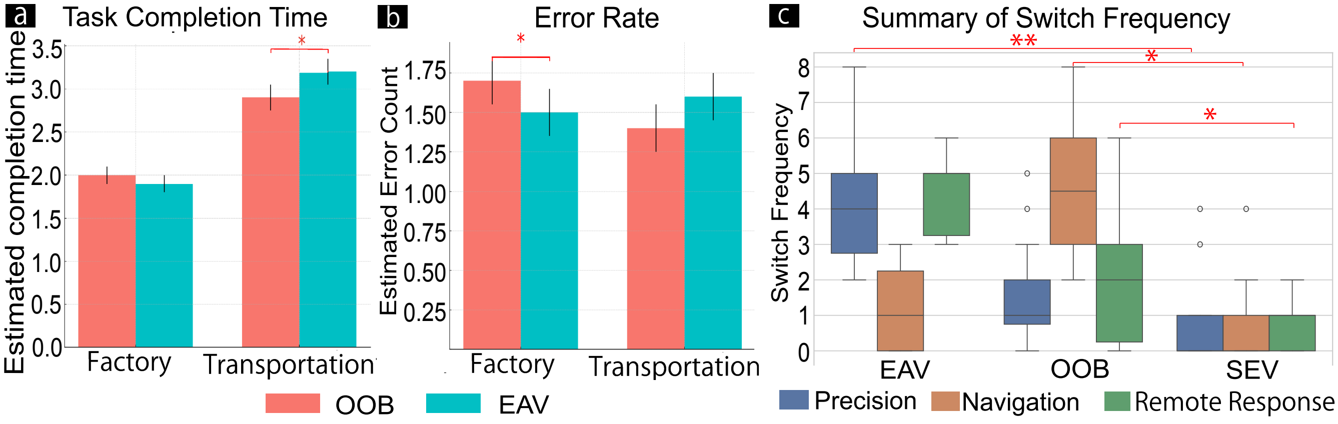

Figure 9: Effects of perspective mode and task phase. (a) Completion time: Out-of-body View led to significantly faster task completion in the Transportation task. (b) Error count: Out-of-body View resulted in significantly fewer errors in the Factory task. (c) Switch frequency by task phase. Switching was infrequent in the baseline Shared Embodied View condition (median ≤ 1 across phases). In comparison, participants switched more frequently to Embedded Anchored View during precision subtasks, and to Out-of-body View during navigation and remote-response phases. * indicates $p < .05$, ** $p < .01$, and *** $p < .001$.

6.4 Task Performance Analysis (RQ3)

Here, we answer RQ3 by examining task performance metrics across different perspective modes. Task completion times and error rates were analyzed using GLMMs with gamma and Poisson distributions respectively (random intercept for participant ID). Results are summarized in Figure 9 (a,b). Participants completed the Transportation task significantly faster in Out-of-body View ($M = 15.5$, s, $SE = 1.2$) than in Embedded Anchored View ($M = 22.6$, s, $SE = 1.7$; $z = 6.64$, $p < .001$). with the GLMM indicating an estimated time ratio of ≈ 0.69, 95% CI $[0.55, 0.84]$ (i.e., about 31% faster in Out-of-body View). In contrast, completion times for the Factory task did not differ reliably between perspectives ($p = .07$). In the Factory task, significantly fewer errors were observed in Out-of-body View ($M = 3.4$, $SE = 0.5$) than in Embedded Anchored View ($M = 4.4$, $SE = 0.6$; $z = -2.49$, $p < .05$). The Transportation task yielded no statistically reliable difference in error rate.

In answering RQ3, our results suggest that the observer perspective (Out-of-body View) offers performance advantages under different demands: faster navigation in dynamic tasks (Transportation) and fewer errors in precision tasks (Factory).

6.5 Perspective Switching Behaviors (RQ4)

To address RQ4, we examined how participants strategically switched perspectives during tasks. As shown in Figure 9(c), only *guests* could initiate perspective switches. Following prior work in collaborative HRI/VR [15, 96, 124], each trial was segmented (from logs) into three phases: *precision* (near-body manipulation), *navigation* (locomotion/wayfinding), and *remote response* (reacting to host movement or environment events).

- **Navigation phase**: *host* locomotion (>1.5s continuous motion), no active VSL control from *guest*. This phase aligns with “approach” in handover models [15].
- **Precision-focused subtasks**: stationary *host*, active VSL grasp/place attempts by *guest*. These represent the “manipulation” phase typical in object transfer tasks [96].
- **Remote-response events**: short *guest*-initiated viewpoint switches in response to unexpected *host* movement (e.g., rotation, misalignment). These reflect reactive perspective adjustments during miscoordination or disorientation [82].

We modeled switch frequency with a GLMM (Poisson; random intercepts for participant), revealing strong modulation by task phase. Guests switched predominantly to Embedded Anchored View during *precision* subtasks ($M = 4.20$, $SE = 0.30$, 95% CI $[3.61, 4.79]$, $p<.01$), and to Out-of-body View during *navigation* ($M = 4.89$, $SE = 0.31$, 95% CI $[4.28, 5.50]$, $p<.05$). For *remote-response* moments, guests again preferred Out-of-body View ($M = 3.17$, $SE = 0.32$, 95% CI $[2.54, 3.80]$; $p<.05$ versus other views). Switching while in Shared Embodied View was rare (median ≈ 1 across phases), indicating little perceived need to change away from the baseline shared first-person view except when task demands shifted.

These findings answer RQ4 by showing that guests adaptively switch perspectives based on phase-specific demands, preferring Embedded Anchored View for close manipulation and Out-of-body View for navigation and rapid corrections. This indicates that switching behaviors are tightly coupled to functional needs and task context.

6.6 Qualitative Findings

We conducted an inductive thematic analysis [11] to explore participants' experiences with VSLs operation and perspective-switching during collaborative tasks. This analysis aimed to identify interaction strategies, user preferences, and challenges in managing viewpoint shifts and VSLs control. Two independent coders analyzed interview transcripts, achieving good inter-rater reliability (Cohen's Kappa, $\kappa = 0.78$).

Following Braun and Clarke's framework [12], we initially familiarized ourselves with the data by reviewing interview transcripts and noting key insights. During initial coding, we labeled relevant segments related to strategies, cognitive load, viewpoint preferences, spatial awareness, and collaborative challenges. Independent coding was followed by joint discussions to consolidate these codes into initial themes.

Iterative refinement grouped codes into four broader themes: "Visual Decoupling and Embodiment", "Prior Experience and Perspective Adaptation", "Strategies for Managing Perspective Switching", and "Emergent Communication through Gestures". These themes structured our understanding of participants' adaptive behaviors, cognitive demands, and collaboration strategies during dynamic perspective switching in VSLs supported VR teleoperation.

The following sections detail these themes, illustrating participants' approaches to viewpoint control, collaboration strategies, and adaptations to dynamic teleoperation demands.

6.6.1 Role-asymmetric Embodiment Trade-offs. As previously introduced, *Visual Decoupling* refers to the *guest*'s adoption of an independent external perspective (OUT-OF-BODY VIEW), decoupled from the *host*'s visual alignment. Initially, based on prior research [41], we hypothesized that visual decoupling would negatively impact users' sense of embodiment due to the visual and proprioceptive mismatch. However, *hosts* overwhelmingly reported an enhanced sense of embodiment when their remote partners adopted the OUT-OF-BODY VIEW, emphasizing a feeling of freedom and control over their movements.

Several *hosts* explicitly mentioned that they felt less inhibited when their partners were not visually coupled to their head movements. For example, one participant stated: "I felt comfortable moving naturally without constantly worrying" (*host*, P17). Specifically, *hosts* were concerned that their own head movements or sudden turns could cause visual instability or discomfort for their remote partners, which often made them overly cautious and hesitant during collaboration. When freed from these concerns, they reported increased confidence and spontaneity in their movements, contributing to a stronger sense of ownership over the shared body.

In contrast, *guests* generally reported reduced self-embodiment in the OUT-OF-BODY VIEW, often describing a sense of detachment: "felt like watching myself from outside" (*guest*, P19). This was attributed to the lack of visual alignment with their avatar's first-person perspective, which made it harder to associate with the embodied body. One *guest* described: "When I was not seeing through the avatar's eyes, it didn't feel like it was me moving." We interpret this divergence as a role-dependent trade-off: for *hosts*, the removal of visual coupling alleviated their self-imposed constraints and enabled greater natural movement and self-ownership; for *guests*, the visual decoupling created psychological distance and disrupted the integration of the avatar into their body schema. These asymmetrical embodiment effects reveal that OUT-OF-BODY VIEW can simultaneously enhance host autonomy while disrupting guest self-identification with the avatar. Consistent with this interpretation, RMSSD-based HRV did not show large role-specific divergences across perspectives, suggesting that these embodiment trade-offs are primarily driven by visual coupling and perceived responsibility rather than gross differences in physiological arousal.

6.6.2 Importance of Gaming Experience over VR Experience. Participants with extensive gaming experience adapted more rapidly to OUT-OF-BODY VIEW than those with mainly non-gaming VR backgrounds. Several participants linked this to prior fluency in third-person or drone-style games. One noted, "This view feels just like controlling a game character in third-person games" (*guest*, P8), while another commented, "I like playing games. The drone one felt like a game" (*guest*, P10). These users consistently described OUT-OF-BODY VIEW as intuitive and familiar, citing spatial awareness and mental models shaped by gaming.

In contrast, participants with strong VR experience but limited gaming backgrounds struggled more with the externally detached spatial perspective. One such participant shared, "I kept getting disoriented" (*guest*, P15), attributing their discomfort to the lack of visual alignment and direct embodiment. These findings suggest that gaming, rather than general VR use, better equips users to operate within non-egocentric, third-person viewpoints, as common in the OUT-OF-BODY VIEW condition.

These contrasting patterns highlight that prior gaming experience, particularly with third-person perspectives, plays a more critical role than general VR familiarity in shaping users' ability to navigate detached viewpoints like OUT-OF-BODY VIEW.

6.6.3 Strategic Minimization of Perspective Switching. Though we initially hypothesized that frequent perspective switching would improve task performance, several participants (N=6) reported deliberately minimizing their switches to maintain focus and reduce mental fatigue. As one *guest* explained, "I found fewer switches kept me focused" (P3). These participants described a strategy of limiting perspective changes to key task transitions, citing benefits such as reduced cognitive load and sustained concentration.

In contrast, a smaller group adopted a high-frequency switching strategy, believing it improved their situational awareness. One such participant remarked, "I needed different perspectives to manage different tasks better" (*guest*, P11). Although these users acknowledged the added cognitive effort and occasional disorientation, they perceived the benefits of enhanced task visibility as outweighing the costs.

Together, these divergent strategies highlight individual differences in managing cognitive resources and suggest that perspective-switching preferences may depend on thresholds for mental load and spatial tracking. Consistent with this, our HRV analysis (Sec. 6.1) shows perspective-dependent differences between trials with more versus fewer switches, suggesting that these strategies are associated with differences in physiological effort.

6.6.4 Emergence of Natural Gesture-Based Communication Strategies. A spontaneous, non-verbal communication strategy emerged in some participant pairs (N=18), particularly those who relied heavily on OUT-OF-BODY VIEW. "I use robotic arm gestures to explain" (*guest*, P30). These pairs developed gesture-based conventions to complement or replace verbal instructions, often describing this as more efficient and natural than speech alone.

However, not all dyads adopted this strategy. "We kept misunderstanding each other, so we stuck with verbal instructions" (*guest*, P14). Some participants found gestures ambiguous or lacked confidence in their clarity. Thus, while gesture-based interaction emerged organically under visual decoupling, it was not universal.

Taken together, these findings suggest that gesture-based communication strategies may emerge organically, shaped by spatial alignment, mutual visibility, and the degree of interpersonal trust.

7 DISCUSSION

Our study examined how guest-driven perspective switching (SHARED EMBODIED VIEW↔ EMBEDDED ANCHORED VIEW, SHARED EMBODIED VIEW↔ OUT-OF-BODY VIEW) strategies influence performance, embodiment, workload, and adaptive user behaviors in collaborative VR teleoperation. Results revealed clear task-dependent performance advantages (RQ1), distinct role-specific embodiment preferences (RQ2), and notable effects on workload and physiological responses (RQ3). Additionally, we identified strategic viewpoint-switching patterns aligned with evolving task demands (RQ4).

7.0.1 Reducing Workload and Physiological Strain through Perspective Design (RQ1). Our findings clearly demonstrate that perspective choice significantly influences subjective workload and physiological strain. The OUT-OF-BODY VIEW effectively reduces users' perceived fatigue (VAS-F) and workload (NASA-TLX) during dynamic tasks, confirmed by elevated RMSSD and reduced heart rate. Conversely, despite enhancing embodiment, the EMBEDDED ANCHORED VIEW unexpectedly increased cognitive and physiological burden, likely due to continuous visual stabilization demands. These results extend previous findings [73] by providing integrated subjective and physiological evidence. Our study highlights a key design trade-off between spatial embodiment and physiological comfort, suggesting designers dynamically balance perspectives, prioritizing the OUT-OF-BODY VIEW during intensive tasks and carefully managing usage duration of the EMBEDDED ANCHORED VIEW. Additionally, as several guests reported avoiding perspective switches to stay focused (Section 6.6.3), we propose intelligent guidance mechanisms to help minimize unnecessary switching. Future work should evaluate such mechanisms in realistic settings.

7.0.2 Balancing Embodiment and Spatial Awareness: Role-specific Considerations (RQ2). Our findings reveal significant role-specific differences regarding embodiment and spatial awareness across perspectives. *Hosts* experienced heightened agency and ownership primarily within the SHARED EMBODIED VIEW, benefiting from a coherent first-person grounding, whereas EMBEDDED ANCHORED VIEW aided fine control near the body. *Guests* reported stronger embodiment in the SHARED EMBODIED VIEW, but preferred the OUT-OF-BODY VIEW for better spatial awareness and stable control during navigation. This extends prior work [36], explicitly identifying distinct user needs based on collaborative roles. Critically, we found a clear design tension: the OUT-OF-BODY VIEW enhanced spatial cognition but diminished *guest*'s sense of embodiment and increased perceived effort in some tasks, consistent with prior studies highlighting the impact of visual–body mismatches [20]. To mitigate this trade-off, we recommend adaptive, role-specific viewpoint strategies to dynamically balance embodiment and spatial awareness. Future work should validate role-dependent effects in more complex, real-world scenarios.

7.0.3 Optimizing Collaborative Performance through Adaptive Perspective Switching (RQ3, RQ4). Our results demonstrated clear task-dependent advantages for each perspective mode. Specifically, the OUT-OF-BODY VIEW significantly improved performance in navigation tasks compared to the EMBEDDED ANCHORED VIEW (RQ3), confirming prior insights regarding third-person benefits for spatial awareness [43, 61]. However, unlike previous studies, we explicitly quantified these perspective effects across diverse collaborative scenarios. Role-specific embodiment differences emerged, with *host* users reporting greater agency and ownership in the SHARED EMBODIED VIEW, and *guest* experiencing stronger embodiment in the SHARED EMBODIED VIEW but better spatial orientation in the OUT-OF-BODY VIEW (RQ2), extending previous research [37]. Moreover, our subjective and physiological data indicate a trade-off: OUT-OF-BODY VIEW was associated with *lower physiological stress for guests* (higher RMSSD) but *higher subjective fatigue/workload in certain contexts* (e.g., higher NASA–TLX in Factory, and higher host workload during Transportation), while EMBEDDED ANCHORED VIEW supported near-body stability without the same spatial overview. Qualitative analysis also identified strategic user-driven viewpoint switching aligned with dynamic task demands (RQ4).

7.1 Implications for the Design of Collaborative Teleoperation Systems

Our findings provide clear practical implications for designing collaborative VR teleoperation systems, emphasizing adaptive perspective selection to balance performance, user comfort, and cognitive load. Specifically, the OUT-OF-BODY VIEW improves spatial navigation and reduces cognitive strain, while the EMBEDDED ANCHORED VIEW enhances precision and embodiment. We specifically highlight previously overlooked cognitive and physiological costs associated with frequent viewpoint switching, recommending adaptive guidance to address these trade-offs. Our practical guidelines are outlined in the Table 2.

Consider future high-stakes settings where space is constrained but expert intervention is critical, such as tele-supported surgery [95, 113] or in-situ aircraft maintenance [53, 69]. A local generalist could maintain the mobility and support functions while a remote specialist "inhabits" the same body for specialized tasks manipulation, relying on an EMBEDDED ANCHORED VIEW-like stabilized view that supports tasks that needs precision. In such cases, periodically switching to the OUT-OF-BODY VIEW may help maintain spatial awareness and reduce strain, making interfaces proactively recommend viewpoint transitions based on task demands. In sports coaching, such as dynamic rugby drills, the OUT-OF-BODY VIEW can effectively reduce discomfort during rapid movements, while detailed skill training benefits from stabilized visual feedback provided by the EMBEDDED ANCHORED VIEW. Designing seamless, context-aware switching interfaces can minimize cognitive load, improving coaching effectiveness. Similarly, in industrial safety contexts, adaptive perspective selection can significantly enhance hazard detection and precise manipulation tasks. Rapid, intuitive perspective switching combined with context-sensitive hazard alerts can optimize both safety and efficiency.

Finally, although our study operates in a purely virtual environment, the perspective trade-offs we identify are directly relevant to emerging teleoperation systems built on AI-based 3D world models and digital twins [77, 142]. As AI-based 3D world models and digital twins mature, remote operators will work inside persistent virtual

Table 2: Design Suggestions for Perspective Switching and Embodiment Cues. Rows marked † are directly grounded in our study findings, whereas rows marked ‡ are forward-looking design opportunities extrapolated from these findings.

Adaptive Viewpoint Recommendations		
Design Focus	**Main Finding**	**Design Suggestions**
Task-based Adaptive Switching† (Section 7.0.3)	Out-of-body View aided navigation; in precision phases guests preferentially chose Embedded Anchored View; frequent switching increased cognitive load for some.	Automatically recommend optimal viewpoints based on real-time task context. Provide minimal yet clear alerts for viewpoint transitions.
Intelligent Switching Cost Alerts‡ (Section 7.0.1)	Users were unaware of cognitive costs incurred by frequent viewpoint switching, leading to unnecessary cognitive strain.	Introduce subtle, predictive visual or auditory indicators informing users about switching costs; ensure alerts are concise and unobtrusive.
Role-Specific Embodiment Optimization		
Embodiment Interfaces† (Section 7.0.2)	Hosts benefited when guests used anchored views; guests needed flexible views (Shared Embodied View or Out-of-body View)	Provide differentiated interfaces and distinct embodiment cues tailored to *host* and *guest*. Clearly differentiate avatar alignment visuals.
Embodiment Reinforcement Cues† (Section 7.0.2)	Users experienced reduced embodiment due to inconsistent avatar-body alignment cues, especially during transitions.	Implement dynamic real-time IK-based visualizations of limb alignment. Use selective limb highlighting to reinforce embodiment and user agency.
Visual Stability and Emergency Response		
Visual-Stability Enhancement† (Section 7.0.1)	Rapid or frequent viewpoint transitions caused visual instability and discomfort, negatively impacting user comfort and performance.	Utilize adaptive smoothing techniques during perspective transitions; dynamically adjust smoothing intensity according to user physiology (e.g., HRV).
Rapid Emergency Switch‡ (Section 7.1)	Situational hazards required rapid spatial reassessment; delayed perspective switches negatively impacted safety and performance.	Design intuitive rapid perspective-switching gestures or single-click mechanisms for emergency use; ensure clear visual affordances minimize false activations.

replicas of physical workspaces, moving through the scene independently of the robot's physical camera pose [74, 77, 107, 142]. In such systems, Out-of-body View-like views can be implemented as world-locked virtual cameras, while Shared Embodied View and Embedded Anchored View correspond to avatar-locked and portal-like views. Our results suggest that these interfaces should expose multiple coupled and decoupled viewpoints aligned with task phases, and manage switching costs through predictive cues or semi-automatic camera positioning, even when specific robot kinematics are abstracted away.

7.2 Limitations and Future Work

Our study was conducted within a controlled laboratory environment, which may limit the generalizability of findings to real-world scenarios characterized by more unpredictable factors and complex interactions. Additionally, our participant pool predominantly consisted of university students, who might possess relatively uniform demographic characteristics, familiarity with VR systems, and cognitive strategies. Future research should include a broader and more diverse participant population, along with testing under more ecologically valid conditions, to better assess the robustness and applicability of our perspective-switching approach.

We also did not administer the Simulator Sickness Questionnaire (SSQ), relying instead on workload (NASA–TLX), fatigue (VAS–F), and HRV as proxies for discomfort in shared viewpoints. We interpret these discomfort-related findings as subjective and physiological strain rather than a standardized measure of simulator sickness, because SSQ was not collected. Future work should incorporate standardized cybersickness measures to facilitate finer-grained comparison across VR studies. We did not run an a priori power analysis for our mixed-effects design; a conservative dyad-level sensitivity check (paired comparison, two-sided α=.05, N=24) suggests the study is powered (~80%) to detect effects around $d \approx 0.60$, so smaller effects may be underpowered. A further limitation concerns the formative study: its short interaction duration (two five-minute trials) primarily surfaces early coordination breakdowns. Longer usage periods may reduce role confusion and improve synchronisation as partners adapt to one another. In extended pilot testing, we observed similar adaptation effects, with guest users becoming more anticipatory of the host's movement patterns over time.

While VR provided an accessible [87, 88, 138] and controlled environment for investigating perspective switching interactions in shared-body teleoperation scenarios involving VSLs, we acknowledge that our findings remain speculative regarding applicability to physical robotic implementations. Although previous research suggests transferability from virtual to real-world systems [6, 70], the specific dynamics of perspective-switching behaviors and collaborative interactions identified here may differ significantly when physical robotic limbs and real-world constraints are introduced. Therefore, our results should be validated through future studies using physical prototypes before generalizing these insights to practical teleoperation applications.

From an implementation perspective, recent toolkits for rapid wearable sensing and on-body feedback prototyping [31, 143, 150], together with fast fabrication pipelines for instrumented textiles and 3D-printed sensors and actuators [18, 26, 27, 44, 108, 109, 131, 147–149], make it increasingly practical to integrate proximity-aware sensing and lightweight pre-cue or hazard signaling into future SRL mounts or garments, complementing rapid perspective switching with anticipatory safety cues rather than relying on vision alone. More broadly, prior work on interactive AI and wearable/on-body systems suggests that trust is calibrated through legible system state and low-cost opportunities to reclaim control rather than maximal autonomy [32, 34, 62–67]; in our setting, guest-driven perspective switching serves as a similarly lightweight, reversible mechanism for negotiating control boundaries as task demands shift.

8 CONCLUSION

We explored dynamic, guest-driven perspective switching in shared-body VR teleoperation with virtual supernumerary limbs (VSLs). Building on a formative study that identified locomotion- and coordination-related challenges under fixed first-person viewing, we implemented two switching conditions: Shared Embodied View↔Embedded Anchored View and Shared Embodied View ↔Out-of-body View. A within-subjects study (N=48) showed that Out-of-body View supports navigation and reduces errors, while Embedded Anchored View is preferred for near-body manipulation phases, with *hosts* reporting stronger agency. Completion times showed no reliable difference across views, but Out-of-body View yielded fewer errors in the Factory task. Together, these findings yield practical, role- and phase-aware design guidelines for optimizing collaborative performance, comfort, and embodiment in mobile shared-avatar VSL systems.

Acknowledgments

This project was supported by the Australian Research Council Discovery Early Career Award (DECRA) - DE200100479. Dr. Anusha Withana is the recipient of a DECRA fellowship funded by the Australian Government. We are grateful for the support provided by the Neurodisability Assist Trust and Cerebral Palsy Alliance, Australia - PRG04219. Dr. Andrea Bianchi was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00337803). Additionally, we appreciate AID-LAB members for assisting us in various ways.

References

[1] 2013. Pacific Rim. Film. Directed by Guillermo del Toro. Warner Bros. Pictures and Legendary Pictures.

[2] 1X Technologies. 2025. *NEO Home Robot.* https://www.1x.tech/neo Accessed: 2025-11-17.

[3] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Sophie Jörg, and Yuting Ye. 2021. Evaluating Grasping Visualizations and Control Modes in a VR Game. *ACM Transactions on Applied Perception* 18, 4 (2021), 1–14. doi:10.1145/3486582

[4] Hesam Alizadeh, Anna Witcraft, Anthony Tang, and Ehud Sharlin. 2016. HappyFeet: Embodiments for Joint Remote Dancing.. In *Graphics Interface.* 117–124.

[5] Evgeniia I Alshanskaia, Natalia A Zhozhikashvili, Irina S Polikanova, and Olga V Martynova. 2024. *Frontiers in Psychiatry* 15 (2024), 1355846.

[6] Ken Arai, Hiroto Saito, Masaaki Fukuoka, Sachiyo Ueda, Maki Sugimoto, Michiteru Kitazaki, and Masahiko Inami. 2022. Embodiment of supernumerary robotic limbs in virtual reality. *Scientific reports* 12, 1 (2022), 9769.

[7] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. doi:10.18637/jss.v067.i01

[8] Naval Bhandari and Eamonn O'Neill. 2020. Influence of perspective on dynamic tasks in virtual reality. In *2020 IEEE conference on virtual reality and 3d user interfaces (VR).* IEEE, 939–948.

[9] Agnese Bonavita, Riccardo Guidotti, and Mirco Nanni. 2022. Individual and Collective Stop-Based Adaptive Trajectory Segmentation. *GeoInformatica* 26 (2022), 451–477. doi:10.1007/s10707-021-00449-8

[10] Torin Borrowdale. 2016. Zen Mode Soundtrack. *Alto's Adventure.* https://www.youtube.com/watch?v=R4Cv-yAbTKM. Accessed: 2025-04-04.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[12] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology* 18, 3 (2021), 328–352.

[13] Ludovic Burg. 2022. *Real-time virtual cinematography for target tracking.* Ph. D. Dissertation. Université Rennes 1.

[14] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems.* 981–992.

[15] Maya Cakmak and Siddhartha S. Srinivasa. 2011. Using spatial and temporal contrast for fluent robot-human hand-overs. In *Proceedings of the 6th international conference on Human-robot interaction.* ACM, 489–496.

[16] Zekun Cao, Jason Jerald, and Regis Kopper. 2018. Visually-induced motion sickness reduction via static and dynamic rest frames. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR).* IEEE, 105–112.

[17] Nancy Carter, Denise Bryant-Lukosius, Alba DiCenso, Judy Blythe, and Alan J. Neville. 2014. The Use of Triangulation in Qualitative Research. *Oncology Nursing Forum* 41, 5 (2014), 545–547.

[18] Edwin Chau, Jiakun Yu, Cagatay Goncu, and Anusha Withana. 2021. Composite line designs and accuracy measurements for tactile line tracing on touch surfaces. *Proceedings of the ACM on Human-Computer Interaction* 5, ISS (2021), 1–17.

[19] Zikuan Chen and Vince Calhoun. 2018. Effect of spatial smoothing on task fMRI ICA and functional connectivity. *Frontiers in neuroscience* 12 (2018), 15.

[20] Dixuan Cui and Christos Mousas. 2022. Evaluating the sense of embodiment through out-of-body experience and tactile feedback. In *Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry.* 1–7.

[21] James J Cummings and Jeremy N Bailenson. 2016. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media psychology* 19, 2 (2016), 272–309.

[22] Shivin Dass, Wensi Ai, Yuqian Jiang, Samik Singh, Jiaheng Hu, Ruohan Zhang, Peter Stone, Ben Abbatematteo, and Roberto Martín-Martín. 2024. Telemoma: A modular and versatile teleoperation system for mobile manipulation. *arXiv preprint arXiv:2403.07869* (2024).

[23] David De Schepper, Bart Moyaers, Gert Schouterden, Karel Kellens, and Eric Demeester. 2021. Towards robust human-robot mobile co-manipulation for tasks involving the handling of non-rigid materials using sensor-fused force-torque, and skeleton tracking data. *Procedia CIRP* 97 (2021), 325–330.

[24] Norman K. Denzin. 1978. *The Research Act: A Theoretical Introduction to Sociological Methods.* McGraw-Hill.

[25] Siyan Dong, Kai Xu, Qiang Zhou, Andrea Tagliasacchi, Shiqing Xin, Matthias Nießner, and Baoquan Chen. 2019. Multi-robot collaborative dense scene reconstruction. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–16.

[26] Yihao Dong, Praneeth Bimsara Perera, Chin-Teng Lin, Craig T. Jin, and Anusha Withana. 2026. TactDeform: Finger Pad Deformation Inspired Spatial Tactile Feedback for Virtual Geometry Exploration. In *Proceedings of the 2026 CHI*

Conference on Human Factors in Computing Systems (CHI '26) (Barcelona, Spain). Association for Computing Machinery, New York, NY, USA. doi:10.1145/3772318.3791699

[27] Yihao Dong, Pamuditha Somarathne, Craig T Jin, Juno Kim, Andrea Bianchi, and Anusha Withana. 2025. Just Before Touch: Manipulating Perceived Haptic Sensations through Proactive Vibrotactile Cues in Virtual Reality. In *Proceedings of the Augmented Humans International Conference 2025.* 79–91.

[28] Andre Doucette, Carl Gutwin, Regan L Mandryk, Miguel Nacenta, and Sunny Sharma. 2013. Sometimes when we touch: how arm embodiments change reaching and collaboration on digital tables. In *Proceedings of the 2013 conference on Computer supported cooperative work.* 193–202.

[29] Thierry Duval, Thi Thuong Huyen Nguyen, Cédric Fleury, Alain Chauffaut, Georges Dumont, and Valérie Gouranton. 2014. Improving awareness for 3D virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. *Journal on Multimodal User Interfaces* 8 (2014), 187–197.

[30] Jonathan Eden, Mario Bräcklein, Jaime Ibáñez, Deren Yusuf Barsakcioglu, Giovanni Di Pino, Dario Farina, Etienne Burdet, and Carsten Mehring. 2022. Principles of human movement augmentation and the challenges in making it a reality. *Nature Communications* 13, 1 (2022), 1345.

[31] Chia-An Fan, En-Huei Wu, Chia-Yu Cheng, Yu-Cheng Chang, Alvaro Lopez, Yu Chen, Chia-Chen Chi, Yi-Sheng Chan, Ching-Yi Tsai, and Mike Y Chen. 2024. SpinShot: Optimizing both physical and perceived force feedback of flywheel-based, directional impact handheld devices. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology.* 1–15.

[32] Hongxiang Fan, Martin Ferianc, Miguel Rodrigues, Hongyu Zhou, Xinyu Niu, and Wayne Luk. 2021. High-performance FPGA-based accelerator for Bayesian neural networks. In *2021 58th ACM/IEEE Design Automation Conference (DAC).* IEEE, 1063–1068.

[33] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160. doi:10.3758/BRM.41.4.1149

[34] Martin Ferianc, Hongxiang Fan, Divyansh Manocha, Hongyu Zhou, Shuanglong Liu, Xinyu Niu, and Wayne Luk. 2021. Improving performance estimation for design space exploration for convolutional neural network accelerators. *Electronics* 10, 4 (2021), 520.

[35] Daniela Feth, Binh An Tran, Raphaela Groten, Angelika Peer, and Martin Buss. 2009. *Shared-Control Paradigms in Multi-Operator-Single-Robot Teleoperation.* Springer Berlin Heidelberg, Berlin, Heidelberg, 53–62. doi:10.1007/978-3-642-10403-9_6

[36] Rebecca Fribourg, Ferran Argelaguet, Ludovic Hoyet, and Anatole Lécuyer. 2018. Studying the sense of embodiment in VR shared experiences. In *2018 IEEE conference on virtual reality and 3D user interfaces (VR).* IEEE, 273–280.

[37] Rebecca Fribourg, Nami Ogawa, Ludovic Hoyet, Ferran Argelaguet, Takuji Narumi, Michitaka Hirose, and Anatole Lécuyer. 2020. Virtual co-embodiment: evaluation of the sense of agency while sharing the control of a virtual body among two individuals. *IEEE Transactions on Visualization and Computer graphics* 27, 10 (2020), 4023–4038.

[38] Susan R Fussell, Leslie D Setlock, and Robert E Kraut. 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 513–520.

[39] Marco Gallipoli, Sara Buonocore, Mario Selvaggio, Giuseppe Andrea Fontanelli, Stanislao Grazioso, and Giuseppe Di Gironimo. 2024. A virtual reality-based dual-mode robot teleoperation architecture. *Robotica* 42, 6 (2024), 1935–1958.

[40] Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. 2017. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one* 12, 12 (2017), e0190109.

[41] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74.

[42] Mar Gonzalez-Franco, Rodrigo Pizarro, Julio Cermeron, Katie Li, Jacob Thorn, Windo Hutabarat, Ashutosh Tiwari, and P Bermell. 2017. Immersive mixed reality training for complex manufacturing. (2017).

[43] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. 2017. First-and third-person perspectives in immersive virtual environments: presence and performance analysis of embodied users. *Frontiers in Robotics and AI* 4 (2017), 33.

[44] Phillip Gough, Praneeth Bimsara Perera, Michael A Kertesz, and Anusha Withana. 2023. Design, mould, grow!: A fabrication pipeline for growing 3d designs using myco-materials. In *Proceedings of the 2023 CHI conference on human factors in computing systems.* 1–15.

[45] Saul Greenberg, Carl Gutwin, and Mark Roseman. 1996. Semantic telepointers for groupware. In *Proceedings sixth Australian conference on computer-human interaction.* IEEE, 54–61.

[46] Jens Emil Sloth Grønbæk, Juan Sánchez Esquivel, Germán Leiva, Eduardo Velloso, Hans Gellersen, and Ken Pfeuffer. 2024. Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* 1–16.

[47] FE Guedry, AR Rupert, and MF Reschke. 1998. Motion sickness and development of synergy within the spatial orientation system. A hypothetical unifying concept. *Brain research bulletin* 47, 5 (1998), 475–480.

[48] Takayoshi Hagiwara, Maki Sugimoto, Masahiko Inami, and Michiteru Kitazaki. 2019. Shared body by action integration of two persons: Body ownership, sense of agency and task performance. In *2019 ieee conference on virtual reality and 3d user interfaces (vr).* IEEE, 954–955.

[49] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[50] Morten Hertzum. 2016. A usability test is not an interview. *interactions* 23, 2 (2016), 82–84.

[51] Morten Hertzum, Pia Borlund, and Kristina B Kristoffersen. 2015. What do thinking-aloud participants say? A comparison of moderated and unmoderated usability sessions. *International Journal of Human-Computer Interaction* 31, 9 (2015), 557–570.

[52] Morten Hertzum and Kristin Due Holmegaard. 2015. Thinking Aloud Influences Perceived Time. *Human Factors* 57, 1 (2015), 101–109. doi:10.1177/0018720814540208

[53] Sun Hongli, Wang Qingmiao, Yu Weixuan, Liu Yuan, Cui Yihui, and Wang Hongchao. 2021. Application of AR technology in aircraft maintenance manual. In *Journal of Physics: Conference Series*, Vol. 1738. IOP Publishing, 012133.

[54] Torsten Hothorn, Frank Bretz, and Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50, 3 (2008), 346–363. doi:10.1002/bimj.200810425

[55] Yuhan Hu, Sang-won Leigh, and Pattie Maes. 2017. Hand development kit: Soft robotic fingers as prosthetic augmentation of the hand. In *Adjunct Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology.* 27–29.

[56] Xincheng Huang, Dieter Frehlich, Ziyi Xia, Peyman Gholami, and Robert Xiao. 2025. GaussianNexus: Room-Scale Real-Time AR/VR Telepresence with Gaussian Splatting. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25).* Association for Computing Machinery, New York, NY, USA, Article 189, 18 pages. doi:10.1145/3746059.3747693

[57] Xincheng Huang and Robert Xiao. 2024. SurfShare: Lightweight Spatially Consistent Physical Surface and Virtual Replica Sharing with Head-mounted Mixed-Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 162 (Jan. 2024), 24 pages. doi:10.1145/3631418

[58] Xincheng Huang, Michael Yin, Ziyi Xia, and Robert Xiao. 2024. VirtualNexus: Enhancing 360-Degree Video AR/VR Collaboration with Environment Cutouts and Virtual Replicas. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24).* Association for Computing Machinery, New York, NY, USA, Article 55, 12 pages. doi:10.1145/3654777.3676377

[59] Sarah Immanuel, Meseret N Teferra, Mathias Baumert, and Niranjan Bidargaddi. 2023. Heart rate variability for evaluating psychological stress changes in healthy adults: a scoping review. *Neuropsychobiology* 82, 4 (2023), 187–202.

[60] Masahiko Inami, Daisuke Uriu, Zendai Kashino, Shigeo Yoshida, Hiroto Saito, Azumi Maekawa, and Michiteru Kitazaki. 2022. Cyborgs, human augmentation, cybernetics, and JIZAI body. In *Proceedings of the Augmented Humans International Conference 2022.* 230–242.

[61] Heather Iriye and Peggy L St. Jacques. 2021. Memories for third-person experiences in immersive virtual reality. *Scientific reports* 11, 1 (2021), 4667.

[62] Shakyani Jayasiriwardene and Dulani Meedeniya. 2021. Architectural framework for an interactive learning toolkit. In *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Vol. 4. IEEE, 14–21.

[63] Shakyani Jayasiriwardene and Dulani Meedeniya. 2021. Interactive and adaptive learning content authoring framework for an m-learning toolkit. In *2021 1st Conference on Online Teaching for Mobile Education (OT4ME).* IEEE, 153–160.

[64] Shakyani Jayasiriwardene and Dulani Meedeniya. 2022. A knowledge-based adaptive algorithm to recommend interactive learning assessments. In *2022 2nd International Conference on Advanced Research in Computing (ICARC).* IEEE, 379–384.

[65] Shakyani Jayasiriwardene and Dulani Meedeniya. 2023. An adaptive and interactive learning toolkit (iLearn). *Software Impacts* 15 (2023), 100471.

[66] Shakyani Jayasiriwardene, Benjamin Tag, Anusha Withana, and Zhanna Sarsenbayeva. 2025. More Than Words: The Impact of Voice Assistant Personality Traits on Failure Mitigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3 (2025), 1–33.

[67] Shakyani Jayasiriwardene, Hongyu Zhou, Weiwei Jiang, Benjamin Tag, Emmanuel Stamatakis, Anusha Withana, and Zhanna Sarsenbayeva. 2026. From Fixed to Flexible: Shaping AI Personality in Context-Sensitive Interaction. *arXiv preprint arXiv:2601.08194* (2026).

[68] Camille Jeunet, Louis Albert, Ferran Argelaguet, and Anatole Lécuyer. 2018. "Do you feel in control?": towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE transactions on*

visualization and computer graphics 24, 4 (2018), 1486–1495.

[69] Yirui Jiang, Trung Hieu Tran, and Leon Williams. 2023. Machine learning and mixed reality for smart aviation: Applications and challenges. *Journal of Air Transport Management* 111 (2023), 102437.

[70] Ziyi Jiang, Yanpei Huang, Jonathan Eden, Ekaterina Ivanova, Xiaoxiao Cheng, and Etienne Burdet. 2023. A virtual reality platform to evaluate the effects of supernumerary limbs' appearance. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 1–5.

[71] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of camera work in mobile video collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 957–966.

[72] M-Carmen Juan, Julen Elexpuru, Paulo Dias, Beatriz Sousa Santos, and Paula Amorim. 2023. Immersive virtual reality for upper limb rehabilitation: comparing hand and controller interaction. *Virtual reality* 27, 2 (2023), 1157–1171.

[73] Julia M Juliano, Nicolas Schweighofer, and Sook-Lei Liew. 2022. Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer. *Journal of NeuroEngineering and Rehabilitation* 19, 1 (2022), 106.

[74] Zhizhong Kang, Juntao Yang, Zhou Yang, and Sai Cheng. 2020. A review of techniques for 3d reconstruction of indoor environments. *ISPRS International Journal of Geo-Information* 9, 5 (2020), 330.

[75] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[76] Michael G. Kenward and James H. Roger. 1997. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics* 53, 3 (1997), 983–997. doi:10.2307/2533558

[77] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–14.

[78] DoHyung Kim, Halim Yeo, and Kyoungju Park. 2025. Effects of an Avatar Control on VR Embodiment. *Bioengineering* 12, 1 (2025), 32.

[79] Daiki Kodama, Takato Mizuho, Yuji Hatada, Takuji Narumi, and Michitaka Hirose. 2022. Enhancing the sense of agency by transitional weight control in virtual co-embodiment. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 278–286.

[80] Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*. 1–9.

[81] Rachel Kornfield, Irene Rae, and Bilge Mutlu. 2021. So close and yet so far: how embodiment shapes the effects of distance in remote collaboration. *Communication Studies* 72, 6 (2021), 967–993.

[82] Geert-Jan M Kruijff, Miroslav Janíček, Tomás López, Niek Smets, Thijs Mioch, Joost van Diggelen, Peter Groenewegen, Arnoud Visser, Rosemarijn Looije, Tibor Bosse, et al. 2010. Experience in system design for human-robot teaming in urban search and rescue. In *Field and Service Robotics*. Springer, 111–125.

[83] Susanne Kumpulainen, Samad Esmaeilzadeh, and Arto J Pesola. 2024. Assessing the well-being benefits of VR nature experiences on group: Heart rate variability insights from a cross-over study. *Journal of Environmental Psychology* 97 (2024), 102366.

[84] Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. doi:10.18637/jss.v082.i13

[85] Sylvain Laborde, Emma Mosley, and Julian F Thayer. 2017. Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Frontiers in psychology* 8 (2017), 213.

[86] Russell V. Lenth. 2018. Estimated Marginal Means, aka Least-Squares Means. *Journal of Statistical Software* 69, 1 (2018), 1–33. doi:10.18637/jss.v069.i01

[87] Ruilin Liu, Tinghui Li, and Zhanna Sarsenbayeva. 2025. Effects of Ambient Illumination and Screen Luminance on Mixed Reality Interaction. *IEEE Access* 13 (2025), 192837–192855.

[88] Ruilin Liu, Tinghui Li, and Zhanna Sarsenbayeva. 2025. Understanding and addressing ambient illumination as a situational impairment for digital devices. In *Proceedings of the 37th Australian Conference on Human-Computer Interaction*. 556–566.

[89] Tianren Luo, Gaozhang Chen, Yijian Wen, Pengxiang Wang, Yachun Fan, Teng Han, and Feng Tian. 2024. Exploring the Effects of Sensory Conflicts on Cognitive Fatigue in VR Remappings. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[90] Azumi Maekawa, Kei Kawamura, and Masahiko Inami. 2020. Dynamic assistance for human balancing with inertia of a wearable robotic appendage. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4077–4082.

[91] Andrew Maimone, Xubo Yang, Nate Dierk, Andrei State, Mingsong Dou, and Henry Fuchs. 2013. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In *2013 IEEE Virtual Reality (VR)*. IEEE, 23–26.

[92] Kirsti Malterud, Volkert D. Siersma, and Ann Dorrit Guassora. 2016. Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research* 26, 13 (2016), 1753–1760. doi:10.1177/1049732315617444

[93] Junichi Nabeshima, MHD Yamen Saraiji, and Kouta Minamizawa. 2019. Arque: artificial biomimicry-inspired tail for extending innate body functions. In *ACM SIGGRAPH 2019 Posters*. 1–2.

[94] Shohei Nagai, Shunichi Kasahara, and Jun Rekimoto. 2015. Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. 113–116.

[95] Felix Nickel, Amila Cizmic, and Manish Chand. 2022. Telestration and augmented reality in minimally invasive surgery: an invaluable tool in the age of COVID-19 for remote proctoring and telementoring. *JAMA surgery* 157, 2 (2022), 169–170.

[96] Stefanos Nikolaidis, Suraj Nath, Ariel D Procaccia, and Siddhartha S Srinivasa. 2013. Human-robot mutual adaptation in collaborative tasks: Models and experiments. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 1–8.

[97] Matthias Nürnberger, Carsten Klingner, Otto W Witte, and Stefan Brodoehl. 2021. Mismatch of visual-vestibular information in virtual reality: is motion sickness part of the brains attempt to reduce the prediction error? *Frontiers in Human Neuroscience* 15 (2021), 757735.

[98] Ohan Oda and Steven Feiner. 2012. 3D referencing techniques for physical objects in shared augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 207–215.

[99] Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John Alejandro Castro-Vargas, Sergio Orts-Escolano, and Jose Garcia-Rodriguez. 2019. A Visually Plausible Grasping System for Object Manipulation and Interaction in Virtual Reality Environments. *Computers & Graphics* 82 (2019), 143–155. doi:10.1016/j.cag.2019.07.003

[100] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*. 741–754.

[101] Federico Parietti and Harry Asada. 2016. Supernumerary robotic limbs for human body support. *IEEE Transactions on Robotics* 32, 2 (2016), 301–311.

[102] Federico Parietti and Harry H Asada. 2013. Dynamic analysis and state estimation for wearable robotic limbs subject to human-induced disturbances. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 3880–3887.

[103] Federico Parietti and H Harry Asada. 2014. Supernumerary robotic limbs for aircraft fuselage assembly: body stabilization and guidance by bracing. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1176–1183.

[104] Federico Parietti, Kameron C Chan, Banks Hunter, and H Harry Asada. 2015. Design and control of supernumerary robotic limbs for balance augmentation. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5010–5017.

[105] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1716–1725.

[106] Christian Penaloza, David Hernandez-Carmona, and Shuichi Nishio. 2018. Towards intelligent brain-controlled body augmentation robotic limbs. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 1011–1015.

[107] Maria-Valeria Peppa, David Bell, and Thomas Komar. 2020. Urban 3D reconstruction from video clips for telepresence applications. *ISPRS International Journal of Geo-Information* 9, 5 (2020), 330. https://www.mdpi.com/2220-9964/9/5/330

[108] Praneeth Bimsara Perera, Hansa Marasinghe, Taiki Takami, Hiroyuki Kajimoto, and Anusha Withana. 2024. Integrating Force Sensing with Electro-Tactile Feedback in 3D Printed Haptic Interfaces. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers*. 48–54.

[109] Praneeth Bimsara Perera, Ravindu Madhushan Pushpakumara, Hiroyuki Kajimoto, Arata Jingu, Jürgen Steimle, and Anusha Withana. 2025. eTactileKit: A Toolkit for Design Exploration and Rapid Prototyping of Electro-Tactile Interfaces. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. 1–17. doi:10.1145/3746059.3747796

[110] Ivan Poupyrev, Mark Billinghurst, Stephen Weghorst, and Tadao Ichikawa. 1996. Go-go interaction technique: Non-linear mapping for direct manipulation in VR. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 79–80.

[111] Domenico Prattichizzo, Maria Pozzi, Tommaso Lisini Baldi, Monica Malvezzi, Irfan Hussain, Simone Rossi, and Gionata Salvietti. 2021. Human augmentation

by wearable supernumerary robotic limbs: review and perspectives. *Progress in Biomedical Engineering* 3, 4 (2021), 042005.

[112] Aswin K Ramasubramanian and Nikolaos Papakostas. 2021. Operator-mobile robot collaboration for synchronized part movement. *Procedia CIRP* 97 (2021), 217–223.

[113] Edgar Rojas-Muñoz, Maria E Cabrera, Chengyuan Lin, Daniel Andersen, Voicu Popescu, Kathryn Anderson, Ben L Zarzaur, Brian Mullis, and Juan P Wachs. 2020. The System for Telementoring with Augmented Reality (STAR): A head-mounted display to improve surgical coaching and confidence in remote areas. *Surgery* 167, 4 (2020), 724–731.

[114] Sebastian Rutkowski, Patryk Szary, Jerzy Sacha, and Richard Casaburi. 2021. Immersive virtual reality influences physiologic responses to submaximal exercise: A randomized, crossover trial. *Frontiers in physiology* 12 (2021), 702266.

[115] MHD Yamen Saraiji, Tomoya Sasaki, Kai Kunze, Kouta Minamizawa, and Masahiko Inami. 2018. Metaarms: Body remapping using feet-controlled artificial arms. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 65–74.

[116] MHD Yamen Saraiji, Tomoya Sasaki, Reo Matsumura, Kouta Minamizawa, and Masahiko Inami. 2018. Fusion: full body surrogacy for collaborative communication. In *ACM SIGGRAPH 2018 Emerging Technologies*. 1–2.

[117] Tomoya Sasaki, MHD Yamen Saraiji, Charith Lasantha Fernando, Kouta Minamizawa, and Masahiko Inami. 2017. MetaLimbs: Multiple arms interaction metamorphism. In *ACM SIGGRAPH 2017 emerging technologies*. 1–2.

[118] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10, 3 (2001), 266–281.

[119] Dong-Won Seo, Gyu-Tae Kim, Woojin Lee, Wan-Seop Rhee, and Seung-Hyun Rhee. 2016. Hybrid reality-based user experience and evaluation of a context-aware smart home. *Computers in Industry* 76 (2016), 11–23.

[120] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. 2011. Visual analogue scale to evaluate fatigue severity (VAS-F). In *STOP, THAT and one hundred other sleep scales*. Springer, 399–402.

[121] Hideki Shimobayashi, Tomoya Sasaki, Arata Horie, Riku Arakawa, Zendai Kashino, and Masahiko Inami. 2021. Independent control of supernumerary appendages exploiting upper limb redundancy. In *Proceedings of the Augmented Humans International Conference 2021*. 19–30.

[122] Rajinder S Sodhi, Brett R Jones, David Forsyth, Brian P Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188.

[123] Jan-Philipp Stauffert, Florian Niebling, and Marc Erich Latoschik. 2020. Latency and cybersickness: Impact, causes, and measures. A review. *Frontiers in Virtual Reality* 1 (2020), 582204.

[124] Anthony Steed, Ye Pan, Franziska Zisch, and William Steptoe. 2016. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*. IEEE, 67–76.

[125] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 265–272.

[126] Ryo Takizawa, Adrien Verhulst, Katie Seaborn, Masaaki Fukuoka, Atsushi Hiyama, Michiteru Kitazaki, Masahiko Inami, and Maki Sugimoto. 2019. Exploring perspective dependency in a shared body with virtual supernumerary robotic arms. In *2019 IEEE international conference on artificial intelligence and virtual reality (AIVR)*. IEEE, 25–257.

[127] Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2007. Videoarms: embodiments for mixed presence groupware. In *People and Computers XX—Engage: Proceedings of HCI 2006*. Springer, 85–102.

[128] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C Tang, and Bill Buxton. 2010. Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 271–280.

[129] Franco Tecchia, Leila Alem, and Weidong Huang. 2012. 3D helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*. 323–328.

[130] Nana Tian, Phil Lopes, and Ronan Boulic. 2022. A review of cybersickness in head-mounted displays: raising attention to individual susceptibility. *Virtual Reality* (2022). doi:10.1007/s10055-022-00638-2

[131] Adele Tong, Praneeth Perera, Zhanna Sarsenbayeva, Alistair McEwan, Anjula C De Silva, and Anusha Withana. 2023. Fully 3D-printed dry EEG electrodes. *Sensors* 23, 11 (2023), 5175.

[132] Yuchuang Tong and Jinguo Liu. 2021. Review of research and development of supernumerary robotic limbs. *IEEE/CAA Journal of Automatica Sinica* 8, 5 (2021), 929–952.

[133] Vighnesh Vatsal and Guy Hoffman. 2017. Wearing your arm on your sleeve: Studying usage contexts for a wearable robotic forearm. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 974–980.

[134] Vighnesh Vatsal and Guy Hoffman. 2018. Design and analysis of a wearable robotic forearm. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 5489–5496.

[135] Miss P Vermunicht, Miss K Makayed, Mr P Meysman, K Laukens, Miss L Knaepen, Miss Y Vervoort, Doctor E De Bliek, Miss W Hens, E Van Craenenbroeck, Miss L Desteghe, et al. 2023. Validation of Polar H10 chest strap and Fitbit Inspire 2 tracker for measuring continuous heart rate in cardiac patients: impact of artefact removal algorithm. *Europace* 25, Suppl 1 (2023), euad122–550.

[136] Cecilia Vindrola-Padros. 2021. *Doing Rapid Qualitative Research*. SAGE Publications Ltd, London.

[137] Brandon Wagstaff and Jonathan Kelly. 2018. LSTM-Based Zero-Velocity Detection for Robust Inertial Navigation. *arXiv preprint arXiv:1807.05275* (2018). https://arxiv.org/abs/1807.05275

[138] Xueyang Wang, Kewen Peng, Chonghao Hao, Wendi Yu, Xin Yi, and Hewu Li. 2025. VR Whispering: A Multisensory Approach for Private Conversations in Social Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[139] Ziwei Wang, Yanpei Huang, Xiaoxiao Cheng, Pakorn Uttayopas, and Etienne Burdet. 2021. Shared Control for Bimanual Telesurgery with Optimized Robotic Partner. *arXiv preprint arXiv:2107.05531* (2021).

[140] Kevin Winata Wong. 2015. *HandsOn: a portable system for collaboration on virtual 3D objects using binocular optical head-mounted display*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[141] Thanatchai Wongjirad, Supawadee Chansangdee, Onnalin Arun, Chinnaphan Phoobuaphet, and Orapadee Joochim. 2024. Development of a Mobile Manipulator Robot for Human-Robot Collaboration. In *Proceedings of the 2024 7th International Conference on Robot Systems and Applications*. 16–20.

[142] World Labs. 2024. Marble: The World Model for Robots. https://marble.worldlabs.ai. Accessed: 2024-11-26.

[143] En-Huei Wu, Po-Yun Cheng, Che-Wei Hsu, Cheng Hsin Han, Pei Chen Lee, Chia-An Fan, Yu Chia Kuo, Kai-Jing Hu, Yu Chen, and Mike Y Chen. 2025. HeadTurner: Enhancing Viewing Range and Comfort of using Virtual and Mixed-Reality Headsets while Lying Down via Assisted Shoulder and Head Actuation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

[144] Yudan Wu, Shanhe You, Zixuan Guo, Xiangyang Li, Guyue Zhou, and Jiangtao Gong. 2023. MR. Brick: designing a remote mixed-reality educational game system for promoting children's social & collaborative skills. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[145] Bo Yang, Jian Huang, Xinxing Chen, Caihua Xiong, and Yasuhisa Hasegawa. 2021. Supernumerary robotic limbs: A review and future outlook. *IEEE Transactions on Medical Robotics and Bionics* 3, 3 (2021), 623–639.

[146] Lai Sum Yim, Quang TN Vo, Ching-I Huang, Chi-Ruei Wang, Wren McQueary, Hsueh-Cheng Wang, Haikun Huang, and Lap-Fai Yu. 2022. WFH-VR: Teleoperating a robot arm to set a dining table across the globe via virtual reality. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4927–4934.

[147] Jiakun Yu, Hasindu Kariyawasam, Shuying Wu, Sriram Subramanian, and Anusha Withana. 2025. Designing Multi-DoF Epidermal Bend Sensors Using Flexible Resistive Traces. *IEEE Sensors Journal* (2025).

[148] Jiakun Yu, Supun Kuruppu, Biyon Fernando, Praneeth Bimsara Perera, Yuta Sugiura, Sriram Subramanian, and Anusha Withana. 2024. IrOnTex: Using Ironable 3D Printed Objects to Fabricate and Prototype Customizable Interactive Textiles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–26.

[149] Jiakun Yu, Praneeth Bimsara Perera, Rahal Viddusha Perera, Mohammad Mirkhalaf Valashani, and Anusha Withana. 2024. Fabricating customizable 3-D printed pressure sensors by tuning infill characteristics. *IEEE Sensors Journal* 24, 6 (2024), 7604–7613.

[150] Neng-Hao Yu, Shih-Yu Ma, Cong-Min Lin, Chi-Aan Fan, Luca E Taglialatela, Tsai-Yuan Huang, Carolyn Yu, Yun-Ting Cheng, Ya-Chi Liao, and Mike Y Chen. 2023. DrivingVibe: Enhancing VR driving experience using inertia-based vibrotactile feedback around the head. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–22.

[151] Hongyu Zhou, Treshan Ayesh, Chenyu Fan, Zhanna Sarsenbayeva, and Anusha Withana. 2024. CoplayingVR: Understanding User Experience in Shared Control in Virtual Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–25.

[152] Hongyu Zhou, Yihao Dong, Masahiko Inami, Zhanna Sarsenbayeva, and Anusha Withana. 2025. A Survey on Methodological Approaches to Collaborative Embodiment in Virtual Reality. *arXiv preprint arXiv:2507.18877* (2025).

[153] Hongyu Zhou, Chia-An Fan, Yihao Dong, Shuto Takashita, Masahiko Inami, Zhanna Sarsenbayeva, and Anusha Withana. 2026. SRL Proxemics: Spatial Guidelines for Supernumerary Robotic Limbs in Near-Body Interactions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)* (Barcelona, Spain). Association for Computing Machinery. doi:10.1145/3772318.3790532

[154] Hongyu Zhou, Tom Kip, Yihao Dong, Andrea Bianchi, Zhanna Sarsenbayeva, and Anusha Withana. 2025. Juggling Extra Limbs: Identifying Control Strategies for Supernumerary Multi-Arms in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3706598.3713647

[155] Hongyu Zhou, Pamuditha Somarathne, Treshan Ayesh Peirispulle, Chenyu Fan, Zhanna Sarsenbayeva, and Anusha Withana. 2024. PairPlayVR: Shared Hand Control for Virtual Games. In *Proceedings of the Augmented Humans International Conference 2024*. 311–314.

[156] Yaonan Zhu, Keisuke Fusano, Tadayoshi Aoyama, and Yasuhisa Hasegawa. 2023. Intention-reflected predictive display for operability improvement of time-delayed teleoperation system. *ROBOMECH Journal* 10, 1 (2023), 17.

A Physiological Data Processing Details

Signal source and export. Heart activity was recorded continuously with a Polar H10 chest strap. We logged the device-provided inter-beat interval (IBI/RR) stream via Bluetooth Low Energy throughout each trial and exported the timestamped RR series for offline processing. Analyses were performed on normal-to-normal (NN) intervals derived from the RR stream.

Artefact handling and NN reconstruction. RR series were preprocessed following recommended HRV workflows [85]. We removed (i) biologically implausible intervals and dropouts (e.g., extremely short/long RR values indicating missed or extra detections) and (ii) transient artefacts identified as outliers relative to local beat-to-beat dynamics. Short gaps created by artefact removal were linearly interpolated to preserve continuity for window-based HRV estimation, while longer discontinuities were excluded from HRV computation. All HRV metrics reported in the paper were computed only from contiguous NN segments that satisfied the minimum window-length requirement described below.

Windowing for HRV. HRV was computed over active-task periods using contiguous analysis windows of at least 60 s (minimum duration per window), consistent with common practice for time-domain HRV when task segments are short [85]. When a trial segment exceeded 60 s, we computed HRV on consecutive windows and then aggregated to one value per participant × condition by averaging across valid windows within that segment.

RMSSD computation. We used RMSSD as the primary time-domain HRV index:

$$\mathrm{RMSSD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(NN_{i+1} - NN_i)^2}.$$

RMSSD was chosen because it is widely used in psychophysiology, is comparatively robust to slow trends in heart rate, and reflects short-term vagal modulation [85]. Higher RMSSD values indicate lower physiological stress/arousal.

Peak heart rate and event-triggered analysis. Peak heart rate was extracted within the same active-task segments as the maximum observed heart rate during the segment. For event-triggered analyses, we time-aligned the heart-rate stream to perspective-switch timestamps from system logs and summarized heart rate in a short post-switch window; these event-level summaries were then modeled as a function of switch direction (e.g., EMBEDDED ANCHORED VIEW→ OUT-OF-BODY VIEW vs. OUT-OF-BODY VIEW→ EMBEDDED ANCHORED VIEW) and role.