Effects of Locomotion and Viewpoint Control on Immersive Robot Telepresence for Navigation and Inspection

Ammaar Zaman, Neven ElSayed, Oliver Prentner, Mohamed Ellethy, Mostafa M. Lotfy, Maram Selsabila Mahmoudi, and Eduardo Veas

***Abstract*—In many robot telepresence interfaces, movement and viewing are managed through handheld inputs. This can divide attention between steering, looking, and inspection. We compared joystick versus omnidirectional treadmill locomotion and joystick-based versus head-guided camera control in a within-subject 2×2 study in which 17 participants used a virtual reality headset to guide a quadruped robot through a physical maze while locating marked wall symbols and capturing them as visual evidence. Mixed-effects analyses showed that head-guided camera control was associated with faster completion, lower wall-contact cost, lower workload, and more favorable presence, usability, technology acceptance, and flow than joystick-based camera control. The strongest overall objective performance was observed in the joystick locomotion plus head-guided camera condition, whereas the omnidirectional treadmill plus head-guided camera condition emerged as the strongest experiential compromise. Exploratory camera analyses showed that head-guided camera control reduced switching, increased reliance on the pan-tilt view, and supported a more stable viewing strategy across the maze. These strategy differences were consistent with part of the performance advantage, especially for wall-contact duration. Targets captured remained near ceiling and simulator sickness remained low across all conditions. Overall, the results indicate that viewpoint control was the stronger differentiator in this combined navigation and inspection task, while locomotion still shaped the objective and experiential tradeoff.**

***Index Terms*—Robot telepresence, immersive teleoperation, viewpoint control, virtual reality, body coupled interaction, omnidirectional locomotion, quadruped robot.**

I. Introduction

Mobile robot telepresence is increasingly relevant in scenarios where people must act in spaces that are distant, hazardous, or difficult to access directly. In these settings, the operator must do more than drive a remote platform from one point to another. The operator must also inspect the environment, notice relevant details, and make decisions through mediated perception. Telepresence has developed from simple mobile video access toward richer robotic systems that support navigation, inspection, collaboration, and remote action across domains such as industry, healthcare, and hazardous field work [1], [2]. At the same time, immersive interfaces have become more prominent in robot operation, with recent reviews showing growing interest in VR based interfaces for control, visualization, and interaction [3], [4]. This shift matters because immersive telepresence involves driving, inspection, and viewpoint management at the same time.

In many practical situations, operators must keep making progress while searching for cues, checking objects or structures, and documenting what is found. This creates a dual task burden in which navigation competes with visual inspection for attention and control. Work on multiple resource use and task switching suggests that performance can suffer when perceptual and control demands compete for overlapping resources [5], [6]. In telepresence, this burden is often expressed through viewpoint management. The operator must decide when to keep moving, when to reorient attention, and how to coordinate robot motion with the currently available view. Prior teleoperation research shows that camera and viewpoint control can directly shape workload, visual attention, usability, and operational performance [7]–[9].

These challenges motivate a shift from viewing telepresence only as remote control toward viewing it as body coupled interaction at a distance. Conventional telepresence interfaces control robot motion and viewpoint through handheld input. Such interfaces are convenient and familiar, but they can divide attention across multiple manual control loops. A more body coupled interface can translate body motion to robot control and head motion to viewpoint, linking movement and looking to bodily action. Prior telepresence work connects agency, self-location, and perceptual transparency to how naturally the operator can act through the remote system [10], [11]. Head-based viewpoint control has shown promise in teleoperated tasks, but the joint role of body coupled locomotion and camera control in immersive robot telepresence remains less explored [9], [12], [13].

We study this question in a robot navigation and inspection task using a within-subject 2×2 design. Participants wear a VR headset and teleoperate a quadruped robot through a physical foam-wall maze while searching for wall symbols and capturing them as visual evidence. The study compares two locomotion methods, omnidirectional treadmill walking and joystick-based locomotion, and two viewpoint control methods, head-guided and joystick-based. This configuration

Ammaar Zaman and Neven ElSayed are with Know-Center GmbH, Graz, Austria. E-mail: azaman@know-center.at; nelsayed@know-center.at.

Oliver Prentner and Eduardo Veas are with Graz University of Technology, Graz, Austria. E-mail: oliver.prentner@tugraz.at; eveas@tugraz.at.

Mohamed Ellethy is with Siemens Digital Industries Software Egypt. E-mail: mohamed.ellethy@siemens.com.

Mostafa M. Lotfy and Maram Selsabila Mahmoudi are with the American University in Cairo. E-mail: mostafamlotfy@aucegypt.edu; marammahmoudi@aucegypt.edu.

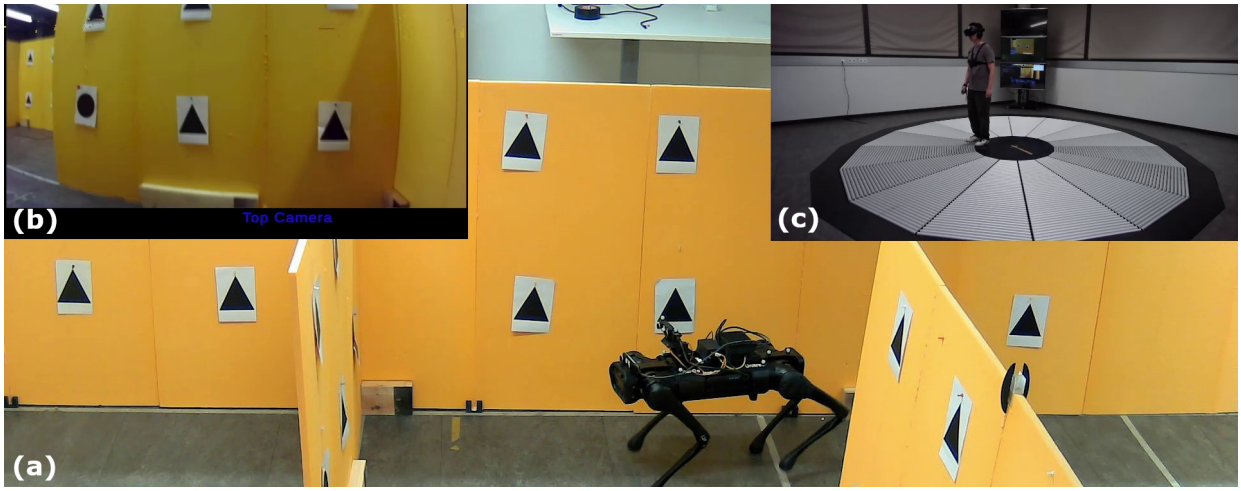


Fig. 1. Body coupled locomotion and viewpoint control. (a) A quadruped robot is navigated through a maze by a (c) user walking in an omnidirectional treadmill with head tracking mapped to pan-tilt camera control, capturing (b) a circle target.

pairs whole-body locomotion on an active omnidirectional treadmill with head tracking for loco-viewpoint control of a mobile quadruped robot. Because the robot does not provide a humanlike body, the study focuses on the relation between user action, camera control, robot response, and useful remote visual feedback [3], [14].

Across the mixed-model analyses, head-guided viewpoint control was associated with faster completion, lower workload, and more favorable presence, usability, technology acceptance, and flow than joystick-based camera control. Joystick locomotion with head-guided viewpoint showed the strongest objective performance, while omnidirectional treadmill locomotion with head-guided viewpoint offered the strongest experiential compromise. Target capture remained near ceiling and simulator sickness remained low across all four conditions. Overall, the results suggest that, in telepresence tasks combining movement and inspection, body coupled locomotion is most useful when the camera mapping keeps visual inspection easy to maintain.

II. Background and Related Work

Robotic telepresence combines sensing, communication, and robot control so that an operator can navigate, inspect, manipulate, or collaborate at a distance. Recent reviews describe a shift toward immersive systems that integrate virtual reality, wearable displays, and richer sensing pipelines [1], [2]. For the present study, the central issue is how the operator turns partial remote views into action while the robot and the camera system keep changing the available feedback. Situation awareness in remote operation depends on perceiving task relevant elements, understanding their meaning, and anticipating what may happen after the next command [5], [15]. In a navigation and inspection task, this includes judging clearance, maintaining route understanding, locating targets, and relating each camera view to the robot body.

A. From teleoperation to immersive robot telepresence

VR has become an increasingly relevant interface layer for robot operation. A systematic review of VR interfaces for controlling and interacting with robots organizes prior work around visualization, robot control and planning, interaction, and usability [3]. Across this literature, VR is used to improve 3D understanding, operator performance, situational awareness, and interface intuitiveness during remote tasks [3], [16]. Prior immersive systems have also shown benefits for remote exploration, mobile robot operation, and telemanipulation, especially when views remain spatially consistent and control mappings are closely coupled to the user's body and movement [17]–[19]. For navigation-oriented telepresence, the relevant question is how naturally the operator can move, look, and act through the remote platform.

B. Camera control and viewpoint management

A central issue in this literature is that camera and viewpoint control are not secondary interface details. They shape how operators divide attention between moving, searching, and acting. In many robot tasks, the operator must balance progress while deciding where to look next and interpreting the relation between the view and the robot's motion. This makes viewpoint management part of the task itself rather than a passive support layer [7], [20].

This issue becomes especially important in navigation-and-inspection tasks, where users must coordinate locomotion with ongoing visual selection. Viewpoint control can itself become a source of control burden because operators may need to decide whether to move the robot, reorient the camera, switch views, or pause to recover spatial understanding. Earlier comparisons of camera and motion control in teleoperated robots show that the way these mappings are assigned can change task performance and operator workload [21]. Recent work further links active camera control to perception-motion coupling strategies, cognitive load, visual attention, usability, and operational performance [7], [8]. Related studies comparing head movement, gaze, and other viewpoint control

techniques show that remote camera control can change both performance and subjective preference [9], [12]. Work on viewpoint transitions further suggests that deciding when to move between views or control modes is itself a meaningful part of teleoperation behavior [13].

Taken together, these studies suggest that camera control is both a visualization choice and part of the sensorimotor coupling between the operator, the interface, and the remote robot. This is directly relevant to the present work, where one of the key questions is whether a more body coupled viewpoint mapping can reduce the burden of managing visual perspective during robot navigation and inspection.

C. Spatial understanding and body coupled telepresence

Operators often have to infer layout, orientation, and future movement consequences through a limited set of remote views. Spatial understanding becomes fragile when visual and motor reference frames do not align well. Prior work has shown that remote and virtual immersive environments can differ from real environments in spatial presence, performance, and behavior [22]. Other teleoperation work similarly notes that restricted viewpoints, frame-of-reference conflicts, and limited sensory context reduce spatial awareness and increase control burden [15], [20], [23].

This problem is not limited to what is visible in the current frame. Spatial knowledge also involves maintaining route structure, object locations, heading, and relations among the user, the viewpoint, the robot, and the surrounding scene. Cognitive map research describes how spatial knowledge supports navigation, memory, and flexible future behavior [24]–[26]. In robot telepresence, that spatial knowledge has to be built from mediated camera views and updated while the robot moves.

For navigation and inspection tasks, these issues are especially important because users must keep moving while interpreting the scene and deciding when to reorient the camera or the robot. Here the idea of body coupled telepresence becomes relevant. Recent work argues that performance can improve when the operator experiences stronger agency, self-location, and body-referenced coupling with the remote system [10]. Broader telepresence and body related interaction literature also discusses how presence and bodily experience intersect in mediated systems and why body coupled control mappings matter for remote interaction [11], [27].

This perspective is not limited to strictly humanlike remote bodies. Research on agency and body-referenced coupling suggests that these experiences can extend beyond highly anthropomorphic forms, which makes this line of work relevant for humanoid teleoperation and for telepresence through mobile legged robots. Prior work on perspective and body related experience in virtual environments also shows that viewpoint, agency, and self-location can separate from one another. First-person views can support strong body related experience and accurate interaction, while third-person or alternating viewpoints can sometimes support broader spatial awareness [28], [29]. This distinction is important here because the robot body is not humanoid, and the experimental question concerns whether useful action feedback and viewpoint coupling can still support remote operation.

For this reason, the present system is best described as partial body coupled telepresence. The operator's head motion was mapped to a mechanical pan-tilt camera, and body motion on the treadmill was mapped to bounded quadruped velocity commands. The study examines whether these mappings provide useful action feedback through a non-human robot body during navigation and inspection.

D. Locomotion and omnidirectional treadmills

A second major design axis in immersive telepresence concerns how the user moves. Controller-based locomotion is common due to its simplicity and convenience, but it does not provide the same bodily cues as walking-based interfaces. VR locomotion research has shown that locomotion technique affects presence, cybersickness, spatial knowledge, and navigation behavior [30]–[32]. A broader typology of VR locomotion techniques emphasizes differences in movement continuity, body involvement, and control mapping [33]. Recent work further shows that locomotion choice influences spatial orientation and user experience during navigation tasks [34], [35].

This distinction is useful because travel is only one part of the wider navigation task. A locomotion technique may move the user efficiently, but the same technique must also preserve route continuity, support inspection, and leave enough attention for the next action. This is why locomotion cannot be interpreted only as a scale from less bodily to more bodily input. Its value depends on whether movement, visual feedback, and task demands fit together.

Omnidirectional treadmill systems also differ from one another. Active omnidirectional treadmills such as the Omnifinity Omnideck 5 move the walking surface to keep the user near the center while body motion is used as input. Passive or sliding omnidirectional treadmills rely more on low friction surfaces and user generated foot motion. Belt based systems such as CyberWalk use controlled belt motion to support walking within a mechanical platform [36]. These approaches all differ from unconstrained real walking because the device constrains gait, balance, speed, comfort, adaptation, and who can use the system safely [37], [38].

Within this literature, omnidirectional treadmills are interesting because they reintroduce walking-like input while keeping the user within a constrained physical space. Prior work has examined their usability, cybersickness, gait characteristics, and navigation performance in virtual environments [39]–[41], including gait and movement behavior during treadmill-based virtual walking [42]. However, most of this research has focused on virtual navigation rather than immersive robot telepresence. It therefore remains unclear whether the benefits often associated with more body coupled locomotion carry over directly to robot telepresence tasks that combine movement with inspection.

E. Gap and motivation

Taken together, prior work shows that viewpoint control affects teleoperation performance and experience, and that locomotion technique changes how users behave in immersive

environments [2], [3]. These lines of work are often studied separately. What remains less explored is how walking-based locomotion through an omnidirectional treadmill and head-guided camera control jointly affect a telepresence task that requires both movement and inspection. This gap is especially relevant for non-humanoid remote bodies such as quadruped robots, where body coupling, locomotion, and viewpoint control do not map cleanly onto a human avatar.

III. Research Focus and Contributions

When exploring an environment, people naturally use head motion and locomotion together while acquiring information. This paper studies that coordination in an immersive robot telepresence task. The study is guided by three expectations from the related work: head guided viewpoint control may reduce camera-management burden during inspection, body coupled locomotion may support presence and flow while adding physical effort and adaptation demands, and the interface that feels best may not be the interface that produces the fastest or safest task performance.

RQ1 In a telepresence task that combines navigation and inspection, how do body coupled and manual locomotion, combined with active or manual viewpoint control, affect task efficiency, wall contact safety, and target capture?

RQ2 How do locomotion and viewpoint control shape workload, presence, simulator sickness, usability, technology acceptance, and flow during immersive robot telepresence?

RQ3 When operators can use a robot body aligned view and an independently steerable view, what camera use strategies emerge, and how do those strategies relate to performance and experience?

This paper contributes a controlled comparison of locomotion and viewpoint control in immersive robot telepresence, empirical findings on their relative roles in performance and user experience, and behavioral evidence showing how camera-use strategies relate to those outcomes. Together, these contributions frame immersive telepresence as a coordination problem in which movement, viewpoint control, and remote action must work together.

IV. Experimental Setup

A. Robot platform

A Unitree AlienGo quadruped robot is used as the platform for the experiment. The robot has two onboard computers for low-level and high-level control. High-level motion control is handled by the onboard Mini PC, which has access to robot state information and implements the control interface used in this study, together with directional movement commands (forward: fwd, backwards: bwd, strafe left: lstr, strafe right: rstr, rotate left: lrot, rotate right: rrot). The robot was used in *walking* mode for active locomotion and received bounded target velocity commands for forward motion, lateral motion, and yaw. The command limits were 0.5 m/s for forward, backward, and lateral motion, and 1.5 rad/s for yaw. Robot control is realized over WiFi through the high-level control computer.

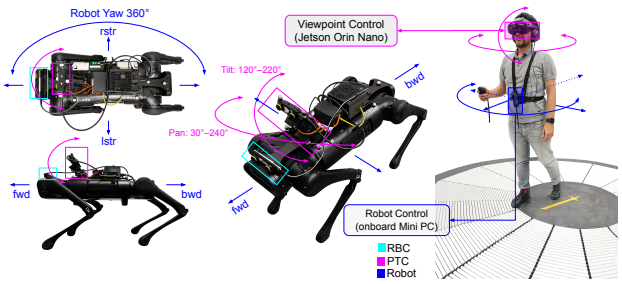


Fig. 2. Overview of the robot platform and telepresence interface.

For robot safety, the controller monitors IMU roll and pitch, gyroscope and accelerometer signals, foot-force unloading, and passive-drop conditions. On a safety trigger, the robot enters a safety routine with zero command velocity and transitions to damping mode if stability does not resolve.

The robot platform is extended with two stereo cameras (ZED Mini) offering two views: a lower fixed Robot Body Camera (RBC) mounted at the head of the robot and an independently controllable Pan-Tilt Camera (PTC) mounted above the torso on a pan-tilt system actuated by two MG996R servomotors mounted in a 3D printed frame. The cameras connect via USB 3.0 to an Nvidia Jetson Orin Nano viewpoint controller mounted on the robot torso and powered from the robot core. The pan-tilt system applied soft servo limits of 30–240° for pan and 120–220° for tilt. The two cameras are calibrated using a standard checkerboard-based stereo calibration procedure so that their forward views align to a common straight-ahead reference when centered. The PTC is positioned so that looking downward retains visibility of the robot body width. The camera heights and quadruped body geometry did not reproduce a human eye height or human body shape, so the apparatus provided partial body coupled telepresence rather than a humanoid telepresence setup. The viewpoint controller communicates through WiFi, receiving commands, updating pan-tilt status, and streaming video from both synchronized stereo cameras. The video pipeline operates at 15 Hz, capturing stereo frames at 2560 $\times$ 720, resizing them to 960 $\times$ 540, and compressing them with JPEG quality 70 for transmission. At any moment the system transmitted the active camera stream used by the operator rather than treating both camera views as separate experimental video channels.

B. Telepresence platform

The telepresence application defines the interface for capturing the operator input for motion and viewpoint control and transferring them to the robot while rendering the robot viewpoints for the operator. It is implemented in Unity3D. The application receives the left and right streams from each camera separately and renders them on the corresponding left and right eye textures for stereo display in the headset. The system offers two locomotion modalities and two modalities for viewpoint control. Velocity and camera-control commands are emitted at 20 Hz. The four interface combinations are abbreviated as J+CJ for joystick locomotion with joystick camera control, J+HGC for joystick locomotion with head guided

camera control, O+CJ for Omnideck locomotion with joystick camera control, and O+HGC for Omnideck locomotion with head guided camera control.

Locomotion modalities. The telepresence application generates velocity commands according to the locomotion modality. Two modalities are used. In joystick locomotion, controller axes are converted into target robot velocities. In treadmill locomotion, tracked body motion on the omnidirectional treadmill is converted into target robot velocities. This was velocity control, not a one to one distance mapping between operator travel and robot travel. Velocity commands were used because they matched the robot high level walking interface and allowed both locomotion inputs to share the same robot command path. In both modalities the Unity application smoothed small changes, suppressed jitter near zero input, and constrained commands before sending them to the robot. The robot side used the most recent received command rather than executing a queued sequence of motions.

Locomotion through joystick In the joystick locomotion condition, VR controllers are used to control forward, lateral, and yaw motion directly. The left thumbstick controls forward motion and yaw, whereas the right thumbstick controls lateral motion. This gave the joystick and treadmill conditions the same robot command type after input processing, which made the locomotion comparison about the human input mapping rather than a different robot walking controller.

Locomotion through omnidirectional treadmill In the treadmill locomotion condition, robot motion is derived from tracked body translation and yaw on the treadmill. The left controller is attached to the torso with an action-camera body strap. This provides the body reference direction, while treadmill motion provides the basis for computing robot velocity updates. The mapping therefore used the participant's body motion as a continuous velocity input while the treadmill kept the participant near the center of the local walking area.

Viewpoint control modalities. Viewpoint control is tied to robot control for the RBC. Its viewpoint changes with robot position and body orientation. The right controller primary button switches between views. The viewpoint control modalities differ for the PTC.

Joystick viewpoint control uses the right joystick to control pan (left/right) and tilt (up/down). Small joystick inputs were ignored by a dead zone, and valid inputs updated the pan and tilt target within the servo limits.

Head viewpoint control uses head orientation tracking from the headset as control commands for the PTC. Pan and tilt are derived from headset orientation relative to a reference frame using a 1.5° dead zone and a recenter function. The pan and tilt controller smoothed commanded changes and clamped them to the same physical range used in the joystick camera condition. Recenter commands returned the PTC to its forward reference.

Mixed locomotion–viewpoint control modalities. In the joystick-locomotion plus joystick-camera condition, the left-controller Y button toggles between robot-control mode and pan-tilt-control mode. In the treadmill plus joystick-camera condition, no such mode switch is required because locomotion is controlled independently through treadmill motion. In the joystick-locomotion plus head-guided-camera condition, pan and tilt follow head motion directly and no locomotion-versus-viewpoint mode toggle is required. In the treadmill plus head-guided-camera condition, treadmill locomotion remains independent while pan and tilt follow head motion directly. The RBC remains body aligned in all conditions. Robot motion was body relative in all conditions and did not depend on the currently selected camera view. Participants could move the robot while viewing either the RBC or the PTC. Thus, J+CJ represents a fully manual baseline with shared controller resources. Other manual mappings with separate camera controls could reduce the cost of switching between robot and camera control.

C. Communication Architecture and Data Flow

The telepresence system is implemented as a distributed local-network architecture centered on a Dockerized Private Cloud (DPC) running on a dedicated local workstation. The DPC acts as the main on-premise orchestration layer for communication between distributed robotic components [43]. It decouples telepresence controls, transmitted through persistent Socket.IO connections to the DPC, from communication with ROS-based robot services. This allows Unity-based components and robot-side processes to exchange commands and telemetry within the same communication fabric without requiring a single shared implementation stack.

D. Robot Navigation and Inspection Task

A robot navigation and inspection task is implemented to study the effects of two locomotion modalities and two viewpoint-control modalities under a 2×2 design. Participants teleoperate a Unitree AlienGo quadruped robot through a physical foam-wall maze while wearing a VR headset. The task goal is to move the robot from start point to end point while capturing wall symbols as visual evidence. The task combines route completion, viewpoint control, visual search, and target documentation within the same trial.

The maze has a fixed outer footprint of $4.2\,\mathrm{m} \times 9.0\,\mathrm{m}$ and is built from $1.25\,\mathrm{m}$ high foam walls. Wall segments are modular, and the study uses $1.2\,\mathrm{m}$ and $1.8\,\mathrm{m}$ wide wall sections. Openings for entry, exit, and turns are $1.2\,\mathrm{m}$ wide, and all turns are 90 degrees. Three maze variants are created by shifting the position of openings relative to the start and end direction while preserving the same overall footprint and turn structure, so that participants cannot rely on memorizing a single repeated layout.

Targets are placed systematically across the maze walls according to a predefined schedule. Circle symbols serve as task-relevant targets to be captured in screenshots, whereas triangle symbols populate the remaining wall locations as non-target distractors. Each trial contains 5, 6, or 7 circle targets, balanced across the 12 trials completed by each participant. For target planning, the traversable maze is divided into five ordered sections from start to end, and the target schedule also balances upper-versus-lower placement across these sections. This varies target visibility across viewpoints and robot positions without favoring any one interface condition.

TABLE I
INTERFACE MAPPINGS AND COMMAND LIMITS USED IN THE FOUR EXPERIMENTAL CONDITIONS. THE TABLE SEPARATES THE OPERATOR INPUT, COMMAND BEHAVIOR, AND CONDITION ROLE FOR LOCOMOTION, VIEWING, ACTIVE VIEWPOINT CONTROL, CAMERA SWITCHING, AND RECENTERING.

Component	Operator input	Command behavior	Condition role
Joystick locomotion	Left thumbstick for forward motion and yaw. Right thumbstick for lateral motion.	Controller axes were mapped to bounded target velocity commands. Command limits were 0.5 m/s for forward, backward, and lateral motion, and 1.5 rad/s for yaw.	Used in J+CJ and J+HGC.
Omnideck locomotion	Tracked torso translation and yaw while standing on the Omnifinity Omnideck 5.	Body motion was mapped to target velocity commands rather than one to one distance travel. Commands were smoothed, thresholded near zero input, and saturated at the same velocity limits as joystick locomotion.	Used in O+CJ and O+HGC.
Robot Body Camera	Camera switch selected the fixed lower view.	The RBC was a view, not an active camera-control mode. It was aligned with the robot body and supported heading, wall clearance, and recovery of robot body orientation.	Available in all conditions. Robot motion remained body relative.
Joystick PTC control	Right thumbstick pan and tilt input when PTC control was active.	Joystick input used a dead zone and updated bounded pan and tilt targets within the 30–240° pan and 120–220° tilt soft servo limits.	Used in J+CJ and O+CJ. In J+CJ active robot control and active PTC control were serial because of the mode switch.
Head guided PTC control	Headset yaw and pitch relative to a reference frame.	Head motion was mapped to bounded pan and tilt targets within the same soft servo limits. Fast head turns could saturate or lag within the configured PTC range rather than producing unbounded camera motion.	Used in J+HGC and O+HGC. Locomotion and active PTC manipulation could occur together.
Camera switching and recentering	Right controller button for view switching and controller recenter commands.	The active view switched between RBC and PTC. Recenter commands returned the PTC to its forward reference.	Available in all conditions.

Two camera views are available: the lower fixed Robot Body Camera (RBC) and the independent Pan-Tilt Camera (PTC). Participants are free to use either for each navigation and target capture. The two views differ in geometric affordances. Being coupled to robot heading, what enters the RBC view depends directly on how the robot body is oriented and positioned relative to the maze wall. The RBC therefore provided a body aligned view for steering, wall clearance, and recovering the robot heading. This means that some targets can require additional body reorientation or larger viewing distance to become fully visible. By contrast, the PTC allows participants to redirect the viewpoint without changing robot heading in the same way. The PTC therefore served as an independent inspection view for search and target capture while the robot could remain oriented for navigation. As a result, it can reduce the need to rotate or reposition the robot body simply to bring a target into view, making viewpoint management part of the task itself rather than a separate support function. The task is a controlled navigation and inspection scenario focused on movement, active viewpoint management, and visual documentation in robot telepresence [2], [3].

Calibration, Safety, and Failure Handling. The task was implemented as an extension of the telepresence platform, including a virtual replica of the maze and robot in Unity, aligned to the physical environment through Vicon-based tracking. This allowed physical robot motion and maze contacts to be mirrored in the virtual scene and logged as collision events. Participants navigated from the live stereo robot camera feed rendered in the headset rather than from the simulated maze. The Unity replica supported alignment, triggers, logging, contact detection, experiment monitoring, and safety checks. Between trials, the physical maze was manually re-aligned to predefined floor markings and the robot was returned to the start position and orientation. Before each trial, participants were asked to verify that they could switch cameras, move the robot using the locomotion modality of the current condition, and capture a sample screenshot.

During the study, a second experimenter monitored robot safety and intervened only when wall contact posed physical risk to the robot. Otherwise, wall contacts were tolerated, logged, and the foam maze was re-aligned after each trial. Participant safety on the treadmill relied on a spotter and an emergency push-stop button at the operator station. No harness was used.

Latency and Synchronization Logging. The system logged command and stream telemetry at multiple stages of the control pipeline. Before each block, a time-health script verified operating-system time services across all computers. Latency reporting was based on application-layer round-trip time (RTT). Glass to glass visual latency was not measured in this study. The latency values therefore characterize command acknowledgment and local receive to apply timing, not the full end to end visual delay experienced by the participant.

Latency was first summarized within each trial and then aggregated across logs for each channel (Table II). The table reports the median of trial medians, the IQR of trial medians, and the median trial $p95$. Robot receive-to-apply values were derived from matched robot-side logs, while the other channels were derived from Unity-side trial logs. Because the video stream operated at 15 Hz and glass to glass latency was not measured, end to end visual delay was not inferred from these RTT values. Condition-wise latency differences were small, so latency was treated as a stable system characteristic rather than an experimental factor.

TABLE II
TRIAL-LEVEL LATENCY CHARACTERIZATION BY LOGGED CHANNEL.

Channel	Median (ms)	IQR (ms)	Trial $p95$ (ms)
Robot command RTT	9.02	8.82–9.55	35.95
Robot receive-to-apply	1.02	0.99–1.04	1.91
Pan-tilt RTT	15.22	15.00–15.42	21.63
Pan-tilt receive-to-apply	3.00	2.00–3.00	6.00
Camera sender ACK RTT	19.53	18.98–21.00	30.70
Camera receive-to-apply	7.90	7.52–8.09	13.95

Apparatus. The task was implemented in Unity3D (v2021.3.15f1) and ran on a Meta Quest Pro headset with standard controllers over a Steam Link connection. Rendering used 4128 $\times$ 2096 resolution at 72 Hz. Interpupillary distance was adjusted before each session using the Quest fit procedure. Participants stood on the base platform of an Omnifinity Omnideck 5 active omnidirectional treadmill in all conditions and wore noise-cancelling headphones with monotonous white noise. The tracked laboratory setup used Vicon cameras with a custom marker configuration in Vicon Tracker.

V. EXPERIMENTAL DESIGN AND PROCEDURE

A. Participants and Ethical Considerations

A total of 17 participants completed the study. They were aged $M = 29.59$ years ($SD = 6.05$, $range = 20 - 48$), including 12 men and 5 women. All participants had normal or corrected-to-normal vision. Eleven participants reported prior experience with VR, five reported experience operating a robot or remote-controlled device, and six reported prior familiarity with omnidirectional VR treadmills. All 17 participants completed all four interface conditions and all 12 trials, giving 204 complete trials for the final analysis.

Participants were recruited through campus flyers, mailing lists, and student-group recruitment channels. Inclusion criteria were age 18 years or older, normal or corrected-to-normal vision. Exclusion criteria were pregnancy, self-reported susceptibility to motion sickness or VR discomfort, current use of medication likely to impair balance, alertness, or comfort during the study, or any self-reported condition that would make treadmill use unsafe. All participants received voucher compensation for their participation. All participants provided informed consent before the study. The study was approved by the institutional Ethics Committee.

B. Repeated-Measures Design

The study used a 2×2 within-subjects design with locomotion modality (Omni vs Joystick) and viewpoint (camera) control (Head vs Joystick) as independent variables. This produced the four interface combinations defined in Section IV-B, J+CJ, J+HGC, O+CJ, and O+HGC. Rather than as isolated component swaps, the conditions represent a systematic analysis of body coupling with different bodily links in viewpoint and locomotion [14].

Each participant completed four condition blocks, one per interface condition. Block order refers to the serial position of a condition block within the full session (1–4), counterbalanced across participants using a Latin-square schedule. Within each block, participants completed three trials, one in each maze variant; trial in block refers to the within-condition trial position (1–3). Maze refers to the three maze variants, and target count refers to whether a given trial required 5, 6, or 7 targets to be captured. Each participant therefore completed 12 trials in total. Maze variant and target count were part of the task schedule rather than primary independent variables. They were included to vary route geometry and inspection demand while preserving the main 2×2 interface comparison. A participant-specific counterbalancing schedule balanced condition order and maze order across the sample, while a predefined target-assignment schedule balanced inspection demand and upper-versus-lower target placement without favoring any one interface condition. In the final dataset, each interface condition contributed 51 trials, each maze variant appeared 68 times, and trials requiring 5, 6, or 7 targets each appeared 68 times.

C. Training and Familiarization

Before each condition block, participants completed a short training session designed to match the upcoming interface condition. Training consisted of five structured scenarios that required basic robot movement, camera switching, screenshot capture, and navigation while using the full range of motions supported by the current locomotion and viewpoint mappings. The scenarios included moving between predefined cardinal positions, switching between cameras, capturing sample screenshots, and navigating while capturing screenshots. Participants were also instructed to account for the robot body footprint while navigating. Training was criterion based. It typically lasted approximately 10 minutes per block and continued until the participant successfully completed at least four of the five scenarios. No participant required training beyond this criterion. This familiarization reduced unsafe or confused use, but it did not remove practice effects or the possibility that the four interfaces had different adaptation demands.

D. Experimental Procedure

Participants stood at the center plate of the omnidirectional treadmill in all conditions. Using the same standing posture across conditions kept the operator station, headset use, camera views, and safety procedure consistent. They wore the VR headset and noise-cancelling headphones with white noise during task execution. Control mappings and camera-operation behavior followed the interface configuration described in Section IV-B. Each trial began at the marked maze start point and ended when the robot reached the end point. Participants were instructed to complete the maze as fast as possible, avoid colliding with the walls, and capture all required wall targets within their intended maze sections using one target per screenshot. There was no time limit per trial. Because both views could support both navigation and target capture, participants were not instructed to assign different task roles to the two cameras. Instead, they were free to use the available

TABLE III
POST-TRIAL QUESTIONNAIRE MEASURES AND SCORING.

Measure	Items	Response range	Reported score
NASA-TLX workload	6	1 to 10	Raw item mean. Lower values indicate lower workload.
Simulator sickness symptom score	11 symptom ratings	0 to 3	Mean across symptom ratings. Lower values indicate fewer symptoms.
Presence questionnaire	6	1 to 10	Item mean. Higher values indicate higher presence.
System usability scale	10	1 to 5	Standard SUS conversion to a 0 to 100 score. Higher values indicate higher usability.
Technology acceptance	5	1 to 10	Item mean. Higher values indicate higher acceptance.
Flow	5	1 to 10	Item mean. Higher values indicate higher flow.

viewpoints in whatever way they found most effective for completing the route and documenting the targets. Scheduled breaks were provided between trials and condition blocks, typically lasting 5–10 minutes. Each condition block lasted approximately 50–55 minutes, and the full session lasted approximately 3.5 hours.

E. Screenshot Scoring and Questionnaires

After each trial, participants removed the headset and completed the post-trial questionnaire battery on a tablet. The questionnaire order was equipment comfort and technical issues, NASA-TLX items, simulator sickness symptom ratings, PQ items, SUS items, TAM items, Flow items, and task-specific camera/control items. The sickness items were administered after each trial only and averaged into a symptom score on the original 0 to 3 item scale.

Screenshot files were stored with filenames linked to trial metadata and the active camera. After the experiment, screenshot completion was derived against the predefined target schedule using a semi-automated workflow based on the stored filenames, the participant-specific target-assignment tables, and documented adjustments when a single screenshot contained more than one valid target. Triangle distractors were excluded from the target count. Repeated captures of the same target were counted once for completion. When one screenshot contained more than one scheduled circle target, each visible scheduled target was credited after manual validation. Screenshots that could not be matched to the target schedule were reviewed against the trial metadata and excluded from completion scoring when they showed distractors or unusable captures. In the analysis reported here, target-completion percentages were capped at the required target total for estimation. The resulting target capture measure is a trial level completion score for visual documentation, especially because performance was near ceiling.

F. Measures

Independent Variables The two primary within-subject independent variables were locomotion (omni vs joystick) and viewpoint modality (head vs joystick). The mixed-model analyses included maze variant (Maze 1–3), target count (5, 6, or 7 required targets), block order (condition-block position 1–4), and trial in block (within-block trial position 1–3) as additional fixed effects to account for route variation, inspection demand, and learning effects.

Objective Performance Measures. The primary objective performance outcomes were chosen to reflect efficiency, safety, and inspection success in the telepresence task. Completion time was defined as elapsed time from the maze start trigger to the maze end trigger. Targets captured was defined as trial level screenshot completion relative to the required target count for that trial, using the screenshot scoring described above. Safety was quantified using wall-contact count and wall-contact duration derived from the logged collision events. Together, these measures capture whether participants could complete the route efficiently, avoid wall contact, and document the required targets while navigating.

Subjective Measures. Subjective experience was assessed after each trial using post trial questionnaire scales for workload, simulator sickness symptoms, presence, system usability, technology acceptance, and flow. Lower values indicate more favorable workload and symptom outcomes, whereas higher values indicate more favorable presence, usability, technology acceptance, and flow outcomes. The item counts, response ranges, and scoring rules are summarized in Table III. Exploratory task-specific items captured perceived coordination, intuitiveness, comfort, and camera reliance, and a post-study questionnaire collected overall difficulty ratings, preference rankings, and feedback.

Camera-Behavior Measures. The system also recorded camera-behavior measures to characterize how participants managed the RBC and PTC. The main trial-level camera outcomes analyzed in this paper were camera switching frequency, camera dwell allocation, capture camera share, use of both cameras, and camera viewing time. Camera switching, dwell allocation, use of both cameras, and camera-viewing-time measures were computed during active maze traversal between the start and end triggers. Capture camera share was computed from the target-capture screenshots saved for a trial, including valid pre-start captures. Each maze was divided into five ordered sections from start to end so that camera dwell allocation and camera switching frequency could also be examined across early, middle, and late parts of the route.

G. Statistical Analysis

The analyses used trial level mixed models because each participant contributed repeated observations across conditions, maze variants, and trial positions. Participant was in-

TABLE IV
OMNIBUS MIXED MODEL SUMMARY FOR THE MAIN OUTCOMES. SIGNIFICANT FIXED EFFECTS ARE SHOWN WITH TYPE III WALD TESTS. R^2_m AND R^2_c DENOTE MARGINAL AND CONDITIONAL R^2.

Outcome	Model	Interface effects	Task and order effects	R^2_m, R^2_c
Completion time	Gamma GLMM, log	Locomotion, $\chi^2(1) = 21.51$, $p < .001$ Camera control, $\chi^2(1) = 119.96$, $p < .001$	Block order, $\chi^2(3) = 34.26$, $p < .001$ Trial in block, $\chi^2(2) = 16.91$, $p < .001$	.304, .696
Target completion	Grouped binomial GLMM, logit	No fixed effect reached significance Locomotion, $\chi^2(1) = 3.44$, $p = .064$ Camera control, $\chi^2(1) = 1.14$, $p = .285$	Maze, $\chi^2(2) = 0.26$, $p = .878$ Target count, $\chi^2(2) = 2.84$, $p = .241$ Block order, $\chi^2(3) = 6.94$, $p = .074$ Trial in block, $\chi^2(2) = 2.81$, $p = .246$	.477, .830
Wall contact count	Negative binomial GLMM, log	Locomotion, $\chi^2(1) = 3.87$, $p = .049$ Camera control, $\chi^2(1) = 14.95$, $p < .001$ Locomotion by camera control, $\chi^2(1) = 12.64$, $p < .001$	Block order, $\chi^2(3) = 8.23$, $p = .041$	.161, .348
Wall contact duration	Tweedie GLMM, log	Camera control, $\chi^2(1) = 16.92$, $p < .001$	Block order, $\chi^2(3) = 14.00$, $p = .003$	.171, .473
NASA TLX	Gaussian LMM	Locomotion, $\chi^2(1) = 16.73$, $p < .001$ Camera control, $\chi^2(1) = 32.00$, $p < .001$	Block order, $\chi^2(3) = 21.46$, $p < .001$ Trial in block, $\chi^2(2) = 7.58$, $p = .023$	.091, .767
Presence	Gaussian LMM	Locomotion, $\chi^2(1) = 23.19$, $p < .001$ Camera control, $\chi^2(1) = 68.93$, $p < .001$ Locomotion by camera control, $\chi^2(1) = 13.14$, $p < .001$	Block order, $\chi^2(3) = 29.91$, $p < .001$ Trial in block, $\chi^2(2) = 16.13$, $p < .001$	.178, .772
Sickness symptom score	Gaussian LMM	No interface effect reached significance	Block order, $\chi^2(3) = 15.12$, $p = .002$	.022, .826
System usability	Gaussian LMM	Locomotion, $\chi^2(1) = 12.39$, $p < .001$ Camera control, $\chi^2(1) = 78.72$, $p < .001$ Locomotion by camera control, $\chi^2(1) = 12.57$, $p < .001$	Block order, $\chi^2(3) = 8.14$, $p = .043$ Trial in block, $\chi^2(2) = 8.85$, $p = .012$	.148, .763
Technology acceptance	Gaussian LMM	Locomotion, $\chi^2(1) = 3.96$, $p = .047$ Camera control, $\chi^2(1) = 58.60$, $p < .001$ Locomotion by camera control, $\chi^2(1) = 5.87$, $p = .015$	Block order, $\chi^2(3) = 9.09$, $p = .028$	.124, .706
Flow	Gaussian LMM	Locomotion, $\chi^2(1) = 7.04$, $p = .008$ Camera control, $\chi^2(1) = 27.51$, $p < .001$ Locomotion by camera control, $\chi^2(1) = 12.83$, $p < .001$	Block order, $\chi^2(3) = 16.71$, $p < .001$	.071, .788
Camera switching	Negative binomial GLMM, log	Camera control, $\chi^2(1) = 57.06$, $p < .001$	Block order, $\chi^2(3) = 12.97$, $p = .005$ Trial in block, $\chi^2(2) = 6.17$, $p = .046$	.310, .894
PTC dwell allocation	Gaussian LMM	Camera control, $\chi^2(1) = 38.49$, $p < .001$	No task or order effect reached significance	.104, .533
Section level PTC dwell	Gaussian LMM	Condition, $\chi^2(3) = 37.18$, $p < .001$ Condition by maze section, $\chi^2(12) = 40.71$, $p < .001$	Maze section, $\chi^2(4) = 18.96$, $p < .001$ Other task and order controls did not reach significance	.10, .86

cluded as a random intercept in all primary models. Locomotion and camera control were the primary fixed effects. Maze variant, target count, block order, and trial in block were included as task and order controls because they reflected route variation, inspection demand, and practice within the session. Gaussian outcomes were analyzed with linear mixed models. Positive continuous, count, bounded completion, and zero heavy outcomes were analyzed with generalized linear mixed models using Gamma, negative binomial, grouped binomial, or Tweedie families as appropriate. Model comparisons were used to decide whether to retain the locomotion by camera control interaction. Type III tests summarized fixed effects, and estimated marginal means on the response scale were used for condition summaries and Bonferroni adjusted pairwise comparisons.

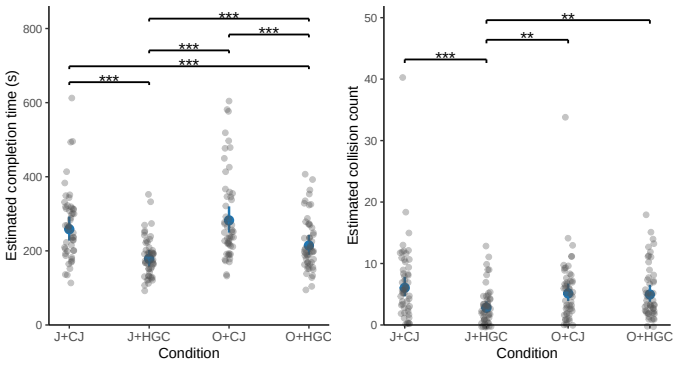


Fig. 3. Estimated marginal means (EMMs) by condition for (left) completion time and (right) wall-contact count. Error bars show 95% CI and gray points show raw trials. Asterisks indicate Bonferroni-corrected significance (* $p < .05$, ** $p < .01$, *** $p < .001$).

VI. RESULTS

Results are reported for the objective performance measures, questionnaire outcomes, and exploratory camera behavior measures. The abbreviations J+CJ, J+HGC, O+CJ, and O+HGC refer to the four interface conditions defined in Section IV-B. Omnibus model results are summarized in Table IV, and model based estimated marginal means on the response scale are summarized in Table V.

A. Objective performance

1) Completion time: Completion time was modeled with a Gamma GLMM using a log link. The omnibus model showed significant effects of locomotion, camera control, block order, and trial in block, while maze, target count, and the locomotion by camera-control interaction were not significant (Table IV).

Estimated marginal means on the response scale indicated that completion time was shortest in *J+HGC* ($M = 178.49$ s,

TABLE V
MODEL-BASED ESTIMATED MARGINAL MEANS (EMMS) ON THE RESPONSE SCALE FROM THE FINAL MIXED MODELS FOR THE OBJECTIVE AND VALIDATED SUBJECTIVE OUTCOMES.

Measure	J+CJ	J+HGC	O+CJ	O+HGC
Objective outcomes				
Completion time (s)	258.25	178.49	282.36	214.45
Targets captured (%)	99.32	98.80	99.78	99.61
Wall-contact count	6.05	2.79	5.13	4.96
Wall-contact duration (s)	19.49	11.59	23.91	14.22
Questionnaire outcomes				
NASA-TLX	4.25	3.76	4.61	4.12
Presence (PQ)	5.50	7.10	6.63	7.27
Sickness symptom score	0.159	0.155	0.167	0.163
System usability (SUS)	58.93	76.96	69.17	76.95
Technology acceptance (TAM)	6.03	7.74	6.78	7.67
Flow	6.56	7.61	7.30	7.50

$SE = 11.40$), followed by *O+HGC* ($M = 214.45$ s, $SE = 13.69$), *J+CJ* ($M = 258.25$ s, $SE = 16.49$), and *O+CJ* ($M = 282.36$ s, $SE = 18.03$), see Fig. 3. Bonferroni-corrected post hoc comparisons showed that *J+HGC* was faster than *J+CJ*, *O+CJ*, and *O+HGC* (all $p < .001$), and that *O+HGC* was faster than *J+CJ* ($p < .001$) and *O+CJ* ($p < .001$). *J+CJ* and *O+CJ* did not differ ($p = .190$).

Order effects were evident. Estimated completion times were lower in Block 3 ($M = 219.36$ s, $SE = 14.00$) and Block 4 ($M = 204.14$ s, $SE = 13.04$) than in Block 1 ($M = 249.01$ s, $SE = 15.90$) and Block 2 ($M = 250.33$ s, $SE = 15.98$). Bonferroni-corrected comparisons showed that Block 1 was slower than Block 3 ($p = .014$) and Block 4 ($p < .001$), and Block 2 was slower than Block 3 ($p = .009$) and Block 4 ($p < .001$). Block 1 and Block 2, and Block 3 and Block 4 did not differ. This pattern indicates that participants became faster after the first two condition blocks. Within each condition block, Trial 1 ($M = 250.11$ s, $SE = 15.56$) was slower than Trial 2 ($M = 224.19$ s, $SE = 13.93$; $p = .007$) and Trial 3 ($M = 216.58$ s, $SE = 13.50$; $p < .001$), whereas Trial 2 and Trial 3 did not differ ($p = 1.0$).

Taken together, completion time shows a clear advantage for head-guided camera control together with a strong learning effect across blocks and within the first trial of each block.

2) Targets Captured: Targets captured was analyzed using a generalized linear mixed-effects model (GLMM) with a grouped-binomial response and logit link. The outcome was modeled as captured targets out of required targets after capping overcapture trials at the required total for estimation. Candidate binomial and beta-binomial mixed models were compared using AIC, and the additive binomial model provided the best fit. Performance remained near ceiling in all four conditions, ranging from 98.80% in J+HGC to 99.78% in O+CJ (Table V).

No fixed effect reached significance in the target-completion model, and no Bonferroni-corrected pairwise condition contrast was significant (Table IV). Participants captured nearly all required targets in every condition, leaving little evidence for reliable condition-related differences in target-capture success.

3) Wall contact: Wall-contact count was analyzed using a generalized linear mixed-effects model (GLMM) with a negative-binomial distribution (`nbinom1`) and log link. Candidate Poisson and negative-binomial mixed models were compared by AIC, and the `nbinom1` interaction model provided the best fit.

The wall-contact count model showed significant effects of locomotion, camera control, block order, and the locomotion by camera-control interaction. Maze, target count, and trial in block were not significant (Table IV).

Estimated marginal means on the response scale indicated that collision count was lowest in *J+HGC* ($M = 2.79$, $SE = 0.46$), followed by *O+HGC* ($M = 4.96$, $SE = 0.70$), *O+CJ* ($M = 5.13$, $SE = 0.71$), and *J+CJ* ($M = 6.05$, $SE = 0.80$), see Fig. 3. Bonferroni-corrected post hoc comparisons showed that *J+HGC* yielded fewer collisions than *J+CJ* ($p < .001$), *O+CJ* ($p = .001$), and *O+HGC* ($p = .002$). The remaining condition contrasts were not significant.

Estimated counts decreased across condition blocks from Block 1 ($M = 5.73$, $SE = 0.78$) to Block 4 ($M = 4.05$, $SE = 0.59$), with Block 2 ($M = 4.42$, $SE = 0.64$) and Block 3 ($M = 4.20$, $SE = 0.61$) in between. However, no Bonferroni-corrected pairwise block-order contrasts remained significant, indicating a gradual downward trend rather than one isolated contrast. Overall, the count model indicates that head-guided camera control, especially when paired with joystick locomotion, was associated with fewer collisions than the camera-joystick conditions.

Wall-contact duration was modeled with a Tweedie GLMM. The model showed significant effects of camera control and block order, while locomotion, maze, target count, and trial in block were not significant (Table IV).

Collision duration was lowest in *J+HGC* ($M = 11.59$ s, $SE = 2.10$), followed by *O+HGC* ($M = 14.22$ s, $SE = 2.51$), *J+CJ* ($M = 19.49$ s, $SE = 3.38$), and *O+CJ* ($M = 23.91$ s, $SE = 4.12$), as summarized in Table V. Bonferroni-corrected comparisons showed that *J+HGC* had lower collision duration than *J+CJ* ($p < .001$) and *O+CJ* ($p < .001$), and that *O+HGC* also had lower collision duration than *O+CJ* ($p < .001$). This mirrors the count result: head-guided camera control reduced both how often participants contacted walls and how long those contacts lasted. Collision duration also decreased across condition blocks: Block 1 was higher than Block 2 ($p = .013$), Block 3 ($p = .043$), and Block 4 ($p = .012$), whereas the differences among Blocks 2–4 were not significant. A sensitivity refit changed the significance of locomotion and trial in block; interpretation therefore focuses on the stable effects of camera control, block order, and the condition-level contrasts.

B. Subjective measures

All questionnaire outcomes were analyzed with Gaussian linear mixed-effects models, and condition means are summarized in Table V. NASA-TLX showed significant effects of locomotion, camera control, block order, and trial in block. Workload was lowest in *J+HGC* ($M = 3.76$, $SE = 0.27$) and highest in *O+CJ* ($M = 4.61$, $SE = 0.27$). The sick-

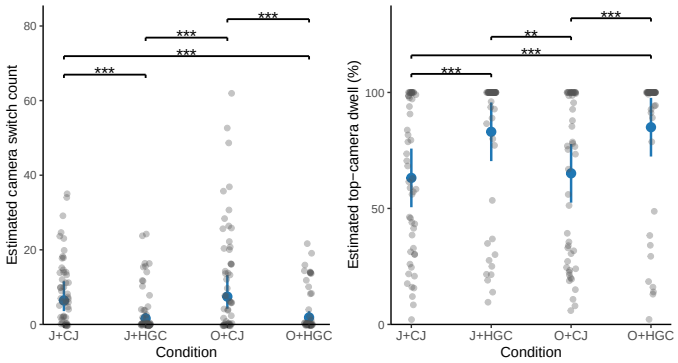


Fig. 4. Estimated marginal means (EMMs) by condition for (left) camera switching frequency and (right) camera dwell allocation, expressed as the percentage of trial time spent in the PTC view. The RBC percentage is the complement of the PTC quantity. Error bars show 95% confidence intervals and gray points show raw trials. Asterisks indicate Bonferroni-corrected pairwise differences (* $p < .05$, ** $p < .01$, *** $p < .001$).

ness symptom score showed no reliable condition effect and remained low across all interfaces (range = 0.155–0.167).

Presence, flow, system usability, and technology acceptance all showed significant camera-control effects. Presence, flow, usability, and technology acceptance also showed locomotion by camera-control interactions. Across these scales, *J+CJ* was consistently weakest, both head-guided conditions were highest or near-highest, and *O+CJ* was generally intermediate. Presence was highest in *O+HGC* ($M = 7.27$, $SE = 0.41$) and *J+HGC* ($M = 7.10$, $SE = 0.41$). Flow was highest in *J+HGC* ($M = 7.61$, $SE = 0.41$) and *O+HGC* ($M = 7.50$, $SE = 0.41$), while usability and technology acceptance were highest in the two head-guided conditions (SUS: both $M \approx 76.95$; TAM: J+HGC $M = 7.74$, O+HGC $M = 7.67$). Together, the questionnaire measures closely matched the objective pattern: head-guided camera control improved the felt quality of remote operation, while *O+HGC* provided the strongest experiential compromise.

C. Camera behavior

1) Camera switching frequency: The following camera-behavior outcomes were analyzed exploratorily to help interpret the condition differences in performance and experience. Camera switching frequency was modeled with a negative-binomial GLMM. The model showed significant effects of camera control, block order, and trial in block. Locomotion, maze, and target count were not significant (Table IV).

Estimated marginal means indicated that switching was least frequent in the head-guided camera conditions, with *J+HGC* showing the fewest switches ($M = 1.62$, $SE = 0.53$) and *O+HGC* remaining similarly low ($M = 1.88$, $SE = 0.61$). Both joystick-camera conditions showed substantially higher switching, with *J+CJ* at $M = 6.42$ ($SE = 1.93$) and *O+CJ* at $M = 7.45$ ($SE = 2.18$), see Fig. 4. Bonferroni-adjusted post hoc comparisons showed that both head-guided camera conditions differed from both joystick-camera conditions, whereas *J+CJ* and *O+CJ* did not differ, and *J+HGC* and *O+HGC* did not differ. This pattern shows that head-guided camera control reduced camera switching across both locomotion modes.

Switching frequency also declined across condition blocks, from Block 1 ($M = 4.63$, $SE = 1.45$) to Block 3 ($M = 2.16$, $SE = 0.72$), with Block 2 ($M = 4.26$, $SE = 1.35$) and Block 4 ($M = 3.42$, $SE = 1.09$) in between. Bonferroni-adjusted tests showed two significant block-order contrasts. Within each condition block, switching was highest in Trial 1 ($M = 4.35$, $SE = 1.31$) and decreased to Trial 2 ($M = 3.56$, $SE = 1.08$) and Trial 3 ($M = 2.71$, $SE = 0.85$), with one Bonferroni-adjusted pairwise contrast significant. Sensitivity refits excluding flagged trials changed only the significance of target count and did not alter the main conclusion that head-guided camera control substantially reduced switching frequency.

2) Camera dwell allocation: Camera dwell allocation was analyzed as the percentage of trial time spent in the PTC view, with the RBC percentage interpreted as its complement. Inferential testing used the PTC share because the two camera shares sum to 100%. The model showed a significant effect of camera control, while locomotion, maze, target count, block order, and trial in block were not significant (Table IV).

Estimated marginal means showed greater PTC dwell in the head-guided camera conditions. *J+HGC* spent $M = 83.05\%$ ($SE = 6.13$) of trial time in the PTC view and *O+HGC* spent $M = 85.02\%$ ($SE = 6.13$), leaving corresponding RBC shares of roughly 17–15%. The joystick-camera conditions showed more mixed allocation, with PTC dwell at *J+CJ* $M = 63.16\%$ ($SE = 6.13$) and *O+CJ* $M = 65.14\%$ ($SE = 6.13$), corresponding to RBC shares of roughly 37–35% (Fig. 4).

Bonferroni-adjusted pairwise contrasts showed that both head-guided camera conditions differed from both joystick-camera conditions. *J+CJ* and *O+CJ* did not differ, and *J+HGC* and *O+HGC* did not differ. This pattern indicates that head-guided camera control shifted viewing toward sustained use of the PTC and reduced reliance on the RBC across both locomotion modes. Absolute viewing-time descriptives followed the same direction, but relative allocation is more interpretable because viewing time changes with overall trial duration. Sensitivity refits excluding flagged trials did not change the omnibus conclusions.

The target-capture screenshots showed a similar viewing preference. The share of captures taken from the PTC was higher in the head-guided conditions (*J+HGC*: $M = 93.34\%$, *O+HGC*: $M = 94.17\%$) than in the joystick-camera conditions (*J+CJ*: $M = 81.74\%$, *O+CJ*: $M = 83.77\%$). This model showed a significant main effect of camera control and no locomotion effect.

Use of both cameras within the same trial showed the opposite pattern. Trials were much more likely to use both cameras in the joystick-camera conditions (*J+CJ*: $M = 94.01\%$, *O+CJ*: $M = 90.28\%$) than in the head-guided-camera conditions (*J+HGC*: $M = 25.23\%$, *O+HGC*: $M = 16.64\%$). This model also showed a significant main effect of camera control and no locomotion effect. Together, these outcomes indicate that head-guided camera control reduced continued view negotiation during both navigation and target capture.

3) Camera-use strategy across maze sections: Camera dwell allocation was analyzed at the section level using a

Gaussian linear mixed-effects model (LMM). Each trial was divided into five ordered maze sections from start to end using the maze-boundary checkpoints. For interpretation, Sections 1–2 are treated as the beginning of the maze, Section 3 as the middle, and Sections 4–5 as the end. Relative to the general model structure described above, these section-level models added *maze section*; for dwell allocation, the *condition* $\times$ *maze-section* interaction was also tested. Random intercepts for *participant* and *trial* accounted for repeated section rows within trials.

The section-level dwell model showed significant effects of condition, maze section, and their interaction. Maze, target count, block order, and trial in block were not significant (Table IV).

Collapsing across sections, PTC share was lower in the joystick-camera conditions than in the head-guided-camera conditions. The estimated means were J+CJ $M = 62.39\%$ ($SE = 6.54$), O+CJ $M = 64.34\%$ ($SE = 6.56$), J+HGC $M = 81.99\%$ ($SE = 6.56$), and O+HGC $M = 85.30\%$ ($SE = 6.51$). Bonferroni-adjusted post hoc tests showed that both head-guided camera conditions differed from both joystick-camera conditions, whereas the two joystick-camera conditions did not differ from each other and the two head-guided camera conditions did not differ from each other.

Across the route, the head-guided camera conditions kept PTC dwell high and stable. J+HGC changed only slightly from Section 1 (80.8%) to Section 5 (79.8%), and O+HGC remained similarly high from Section 1 (84.5%) to Section 5 (87.0%). The joystick-camera conditions showed lower PTC allocation and a stronger late-maze shift toward the RBC, especially in J+CJ, where PTC dwell decreased from 70.6% in Section 1 to 55.3% in Section 5. Sensitivity refits excluding flagged section rows did not change the omnibus significance pattern.

A complementary section-level switching model supported the same interpretation. Collapsing across sections, switching was lower in the head-guided camera conditions (J+HGC $M = 0.13$, O+HGC $M = 0.17$) than in the joystick-camera conditions (J+CJ $M = 0.80$, O+CJ $M = 0.84$). Descriptively, switching remained especially elevated in the middle and later maze sections for O+CJ, whereas both head-guided camera conditions stayed low throughout. Together, the section-level dwell and switching results indicate that HGC supported a stable camera strategy across the route, while joystick-based camera control continued to demand view management deeper into the maze.

4) Exploratory camera-strategy models: We also examined whether camera-use strategy was consistent with part of the condition effects on performance. Each primary performance model was compared with an augmented model that added total section switches, whether the last switch occurred in the late maze, and mean late-maze PTC dwell. These exploratory models reduced the omnibus condition effect by 31.7% for completion time, 28.9% for wall-contact count, and 57.8% for wall-contact duration.

Across the three models, more switching predicted slower completion, more collisions, and longer collision duration. Late-maze PTC dwell predicted fewer collisions, whereas late switching predicted slower completion. These analyses do not establish causal mediation, but they support the interpretation that stable camera-use strategy was associated with part of the advantage of the head-guided conditions.

VII. Discussion

Teleoperation work treats viewpoint design as part of the control loop because camera and motion control mappings change what the operator can attend to, how quickly the remote scene can be understood, and how much effort is spent maintaining a usable view [7]–[9], [21]. The present study extends this argument to a task where navigation, inspection, and visual documentation had to happen together. The main result is not that one input device was uniformly better than another. Instead, the task exposed where coordination costs entered immersive robot telepresence. Viewpoint control shaped performance and experience more consistently than locomotion modality, which shows that the camera mapping was part of the task structure rather than a secondary display choice.

This contribution becomes clearer in the camera behavior results. Accounts of transparent telepresence emphasize that control input and visual feedback must remain easy to relate during action, otherwise the interface itself becomes part of the task [10], [14]. In this study, HGC changed how participants used the available views and reduced the need for repeated view negotiation. In a navigation and inspection task, every camera transition adds a small interpretive cost. The operator must judge whether the current view is sufficient, switch if needed, recover the scene from the new viewpoint, and then continue moving or inspecting. The section-level results and exploratory strategy models point in the same direction. Viewpoint transition work is useful here because switching views is a control decision that can interrupt how the operator understands the current scene and prepares the next action [13]. HGC appears to have reduced this interruption by supporting a more stable viewing strategy across the route.

The locomotion findings are best read through the same task fit logic. VR locomotion research has shown that more body involvement can support experience and spatial updating, but it can also introduce balance, regulation, learning, or control demands when the wider task does not fit the method [44], [45]. Our results follow that tradeoff. Omnidirectional treadmill locomotion did not reliably improve task efficiency, yet O+HGC produced favorable experiential ratings and a stable camera-use pattern. This suggests that body coupled locomotion can improve the felt quality of remote operation when it is paired with low-overhead viewpoint control, but it cannot compensate for a demanding camera mapping.

The quadruped platform clarifies the scope of this telepresence claim. Telepresence accounts that emphasize agency, self-location, and transparent action through a remote system do not require the design argument to rest only on humanoid morphology [10], [11], [14]. Here, the robot body, gait, camera height, and motion limits differed from the operator's body, and both head motion and treadmill motion were mapped to bounded robot mechanisms. The claim is therefore limited

to partial body coupled telepresence. Even so, the interface provided task relevant body cues. The RBC supported heading and wall-clearance judgments, the PTC could return to a forward reference, and looking downward preserved visibility of the robot body width. What mattered was whether the operator could connect their control actions to useful remote visual feedback during movement and inspection.

Several limits define the scope of these conclusions. The sample size was fixed at 17 participants, which limits sensitivity to smaller effects. Target completion was near ceiling, so that outcome could not strongly distinguish the four interfaces. The task was a controlled laboratory navigation and inspection scenario with foam-wall contacts under low speed supervision. Practical systems would need collision avoidance, speed adaptation, shared autonomy, or locomotion assistance for less controlled spaces. Glass to glass visual latency was not measured, although the system logged round-trip timing information. The four conditions also combined posture, body coupling, and controller allocation, and seated joystick or swivel chair operation may be more realistic for some deployments. Treadmill based telepresence adds physical effort and accessibility constraints. Finally, the camera strategy models are exploratory and support a mechanism interpretation, but they do not establish causal mediation.

The practical implication follows from both previous camera-control work and the present robot task. Prior studies connect camera control and viewpoint transitions to workload, attention, and remote action quality [7]–[9]. Our results show that this connection remains critical when the operator is also using a body coupled locomotion interface and controlling a real quadruped robot. For navigation and inspection tasks, omnidirectional locomotion therefore needs a camera-control method that keeps visual inspection easy to maintain while moving.

VIII. Conclusion

In this immersive robot telepresence task, the way the remote viewpoint was coupled to the user mattered more consistently than locomotion modality alone. Head-guided camera control reduced camera-management overhead, supported a stable viewing strategy, and was associated with better objective performance and more favorable subjective experience. These effects emerged on a quadruped robot, suggesting that effective immersive telepresence depends less on morphology alone than on how naturally user actions couple to useful remote visual feedback. For telepresence design, viewpoint-control mapping is a central design decision, and body coupled locomotion is most promising when paired with equally low-overhead viewpoint control.

References

[1] A. Kristoffersson, S. Coradeschi, and A. Loutfi, "A review of mobile robotic telepresence," *Adv. Hum. Comput. Interact.*, vol. 2013, pp. 902 316:1–902 316:17, 2013.

[2] K. Youssef, S. Said, S. A. Kork, and T. Beyrouthy, "Telepresence in the recent literature with a focus on robotic platforms, applications and challenges," *Robotics*, vol. 12, p. 111, 2023.

[3] M. Wonsick and T. Padir, "A systematic review of virtual reality interfaces for controlling and interacting with robots," *Applied Sciences*, vol. 10, no. 24, p. 9051, 2020.

[4] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht, "Immersive telepresence and remote collaboration using mobile and wearable devices," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 1908–1918, 2019.

[5] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159–177, 2002.

[6] S. Monsell, "Task switching," *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 134–140, 2003.

[7] T.-C. Lin, A. Unni Krishnan, and Z. Li, "Perception-motion coupling in active telepresence: Human behavior and teleoperation interface design," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 3, pp. 1–24, 2023.

[8] H. Liu, X. Wang, C. Or, J. Pan, R. Jia, W. Wang, and L. Yang, "Operational performance, cognitive load, visual attention, and usability of fixed-, manual-, and autonomous-camera control in single- and multiple-camera telemanipulation systems," *Applied Ergonomics*, 2025.

[9] S. Yoon, M. Park, and C. R. Ahn, "Comparing dynamic viewpoint control techniques for teleoperated robotic welding in construction," *Automation in Construction*, vol. 172, p. 106053, 2025.

[10] A. Toet, I. A. Kuling, B. N. Krom, and J. B. F. van Erp, "Toward enhanced teleoperation through embodiment," *Frontiers in Robotics and AI*, vol. 7, p. 14, 2020.

[11] N. Nostadt, D. A. Abbink, O. Christ, and P. Beckerle, "Embodiment, presence, and their intersections: Teleoperation and beyond," *ACM Transactions on Human-Robot Interaction*, vol. 9, no. 4, pp. 28:1–28:19, 2020.

[12] D. Zhu, T. Gedeon, and K. Taylor, "Head or gaze? controlling remote camera for hands-busy tasks in teleoperation: A comparison," in *Proceedings of OZCHI 2010*, 2010, pp. 300–303.

[13] S. Yoon, M. Park, and C. R. Ahn, "Learning viewpoint control from human-initiated transitions for teleoperation in construction," *Advanced Engineering Informatics*, vol. 68, p. 103665, 2025.

[14] I. A. Aguilar, M. Suomalainen, S. M. LaValle, T. Ojala, and B. E. Riecke, "From "skype on wheels" to embodied telepresence: a holistic approach to improving the user experience of telepresence robots," *Virtual Reality*, vol. 29, no. 4, p. 161, 2025.

[15] C. Mutzenich, S. Durant, S. Helman, and P. Dalton, "Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles," *Cognitive Research: Principles and Implications*, 2021.

[16] P. Stotko, S. Krumpen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, and M. Weinmann, "A vr system for immersive teleoperation and live exploration with a mobile robot," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[17] J. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters*, 2017.

[18] C.-Y. Kuo, C.-C. Huang, C.-H. Tsai, Y.-S. Shi, and S. Smith, "Development of an immersive slam-based vr system for teleoperation of a mobile manipulator in an unknown environment," *Computers in Industry*, 2021.

[19] R. Schwenk and S. Smith, "An immersive spatially consistent multimodal augmented virtuality human-machine interface for telerobotic systems," *Computers in Industry*, 2025.

[20] L. Almeida, P. Menezes, and J. Dias, "Interface transparency issues in teleoperation," *Applied Sciences*, 2020.

[21] G. Doisy, A. Ronen, and Y. Edan, "Comparison of three different techniques for camera and motion control of a teleoperated robot," *Applied ergonomics*, vol. 58, pp. 527–534, 2017.

[22] N. Khenak, J. Vézien, and P. Bourdot, "Spatial presence, performance, and behavior between real, remote, and virtual immersive environments," *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[23] E. Fabris, V. A. Sangalli, L. P. Soares, and M. Pinho, "Immersive telepresence on the operation of unmanned vehicles," *International Journal of Advanced Robotic Systems*, 2021.

[24] R. A. Epstein, E. Patai, J. Julian, and H. Spiers, "The cognitive map in humans: spatial navigation and beyond," *Nature Neuroscience*, 2017.

[25] M. Peer, I. K. Brunec, N. Newcombe, and R. A. Epstein, "Structuring knowledge with cognitive maps and cognitive graphs," *Trends in Cognitive Sciences*, 2020.

[26] T. Behrens, T. H. Muller, J. C. R. Whittington, S. Mark, A. Baram, K. L. Stachenfeld, and Z. Kurth-Nelson, "What is a cognitive map? organizing knowledge for flexible behavior," *Neuron*, 2018.

[27] M. González-Franco and T. C. Peck, "Avatar embodiment. towards a standardized questionnaire," *Frontiers in Robotics and AI*, vol. 5, p. 74, 2018.

[28] G. Gorisse, O. Christmann, A. Etienne-Armand, and S. Richir, "First- and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users," *Frontiers in Robotics and AI*, vol. 4, pp. 1–12, 2017.

[29] H. G. Debarba, S. Bovet, R. Salomon, O. Blanke, B. Herbelin, and R. Boulic, "Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality," *PLoS ONE*, vol. 12, no. 12, p. e0190109, 2017.

[30] C. Boletsis and J. E. Cedergren, "Vr locomotion in the new era of virtual reality: An empirical comparison of prevalent techniques," *Advances in Human-Computer Interaction*, vol. 2019, pp. 1–15, 2019.

[31] E. Langbehn, P. Lubos, and F. Steinicke, "Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking," in *Proceedings of the Virtual Reality International Conference - Laval Virtual*, 2018, pp. 1–9.

[32] F. Buttussi and L. Chittaro, "Locomotion in place in virtual reality: A comparative evaluation of joystick, teleport, and leaning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 1, pp. 1–12, 2021.

[33] C. Boletsis, "A typology of virtual reality locomotion techniques," *Multimodal Technologies and Interaction*, vol. 6, no. 9, p. 72, 2022.

[34] H. Kim, S.-B. Jeon, and I.-K. Lee, "Locomotion techniques for dynamic environments: Effects on spatial knowledge and user experiences," *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[35] P. Hořejší, A. Lochmannová, V. Jezl, and M. Dvořák, "Virtual reality locomotion methods differentially affect spatial orientation and cybersickness during maze navigation," *Scientific Reports*, 2025.

[36] J. L. Souman, P. Giordano, M. C. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. De Luca, H. H. Bülthoff, and M. O. Ernst, "Cyberwalk," in *ACM Transactions on Applied Perception (TAP)*, vol. 5, no. 3, 2008, pp. 1–22.

[37] V. Wehden *et al.*, "Gait patterns during overground and virtual omnidirectional treadmill walking," *Journal of NeuroEngineering and Rehabilitation*, 2023.

[38] S. Chakraborty, A. Kane, H. C. Gagnon, T. P. McNamara, and B. Bodenheimer, "Comparative effectiveness of an omnidirectional treadmill versus natural walking for navigating in virtual environments," in *ACM Symposium on Applied Perception*, 2024, pp. 1–9.

[39] K. Hooks, W. Ferguson, P. Morillo, and C. Cruz-Neira, "Evaluating the user experience of omnidirectional vr walking simulators," *Entertain. Comput.*, vol. 34, p. 100366, 2020.

[40] J. Lohman and L. Turchet, "Evaluating cybersickness of walking on an omnidirectional treadmill in virtual reality," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 4, pp. 803–812, 2022.

[41] S. Chakraborty, A. Kane, H. C. Gagnon, T. P. McNamara, and B. Bodenheimer, "Comparative effectiveness of an omnidirectional treadmill versus natural walking for navigating in virtual environments," in *ACM Symposium on Applied Perception 2024*, 2024.

[42] M. M. Lewis, C. Waltz, L. Scelina, K. Scelina, K. Owen, K. Hastilow, E. M. Zimmerman, A. Rosenfeldt, M. M. Koop, and J. L. Alberts, "Gait patterns during overground and virtual omnidirectional treadmill walking," *Journal of NeuroEngineering and Rehabilitation*, vol. 21, no. 1, pp. 1–13, 2024.

[43] M. Ellethy, M. George, A. Abouzeid, A. Shaaban, A. Elgammal, R. Adham, M. Abdelsalam, and T. ElBatt, "A digital twin architecture for automated guided vehicles using a dockerized private cloud," in *2023 IEEE Smart World Congress (SWC)*. IEEE, 2023, pp. 1–8.

[44] Y. Wang, J. Eckkrammer, M. Kocur, and P. Wintersberger, "Investigating walking performance and experience with different locomotion technologies in vr," in *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, 2024.

[45] L.-O. Wehden, F. Reer, R. Janzik, W. Y. Tang, and T. Quandt, "The slippery path to total presence: How omnidirectional virtual reality treadmills influence the gaming experience," *Media and Communication*, vol. 9, no. 1, pp. 180–191, 2021.