future internet

MDPI

Article

Designing and Evaluating an Interactive Asymmetric 360-Degree Telepresence System for Long-Term Hospitalized Patients

Yuta Goto [1,*], Joe Hasei [2] and Tatsuo Nakajima [1,*]

[1] Department of Computer Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan
[2] Department of Medical Informatics and Clinical Support Technology Development, Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama 700-8558, Japan
* Correspondence: yuta.goto@dcl.cs.waseda.ac.jp (Y.G.); tatsuo@dcl.cs.waseda.ac.jp (T.N.)

Abstract

Some patients require long-term hospitalization, which restricts their opportunities for social connection and participation in important events or daily lives. To mitigate this issue, we developed an asymmetric 360-degree telepresence system requiring minimal hardware. At the remote site, the user wears a virtual reality head-mounted display to be immersed in the environment. At the local site, a 360-degree camera captures the surroundings, while a humanoid avatar displayed on a rotatable smartphone enables local users to perceive the remote user's gaze and gestures. In an exploratory, controlled laboratory evaluation ($N = 16$, 8 pairs), remote users reported positive preliminary spatial presence while social presence and non-verbal task completion times showed descriptive variations across visual feedback conditions without statistically significant differences, rather than demonstrating task efficiency gains or statistical equivalence to conventional video calls. Furthermore, in an exploratory single-case fieldwork ($N = 1$), a hospitalized adolescent used our prototype system to remotely view her graduation ceremony. In a post-event subjective questionnaire, the participant reported favorable personal impressions regarding the opportunity and experience to view the ceremony online, although severe cellular network congestion limited smooth video streaming to approximately 20% of the session, and standardized clinical or psychological outcomes were not measured. Overall, this work highlights the potential of lightweight asymmetric telepresence as an initial prototype, while identifying key technical, network, and methodological challenges that must be addressed for future real-world deployment.

Keywords: asymmetric telepresence; 360-degree camera; virtual reality; social presence; System Usability; pediatric telepresence feasibility

Check for updates

Academic Editors: Mara Nikolaidou, Teta Stamati, Cleopatra Bardaki, Georgia Dede, Alexandra Gasparinatou, Evangelia Filiopoulou and Elena Politi

Received: 6 September 2026
Revised: 30 October 2026
Accepted: 31 October 2026
Published: 3 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Hospitalized patients often face prolonged periods of clinical isolation depending on their symptoms and treatment protocols. In pediatric care, particularly for children diagnosed with serious conditions such as pediatric cancer, hospitalization can frequently span from eight months to over a year [1]. For instance, in Japan, leukemia is the most prevalent form of cancer in pediatric patients, accounting for approximately 38% of cases in children aged 0–14 and 24% in adolescents aged 15–19 [1]. Undergoing such long-term isolation from family, peers, and standard domestic routines causes severe psychological distress and significantly diminishes patients' mental well-being and quality of life (QOL) [2]. Although temporary home visits or outings are clinically beneficial for patients' mental health, these

institutional privileges are strictly constrained by pathological conditions and the discretion of primary care physicians. Therefore, implementing an accessible, intuitive telepresence technology that connects the hospitalized patients to their daily environments or events is highly warranted.

To address this challenge, researchers have explored various remote communication technologies to mitigate clinical isolation and support social participation. Existing interventions have actively investigated physical robotic avatars and virtual environments to maintain social connections and alleviate psychological distress for isolated patients [3,4].

In parallel, advances in spatial telepresence and neural rendering have enabled photorealistic reconstruction of remote physical environments with high degrees of freedom [5,6]. However, these state-of-the-art volumetric systems generally rely on complex multi-sensor arrays and high computational overhead [7], rendering them impractical for autonomous domestic deployment by patients' families.

To overcome these deployment hurdles while preserving social presence, inspired by emerging lightweight proxy approaches (e.g., IoT Avatars [8]), this study proposes an interactive asymmetric 360-degree telepresence system requiring minimal domestic hardware. In this asymmetric setup, only the remote user (the hospitalized child) wears a VR HMD to foster high spatial immersion, while the local users (e.g., the family at home) are intended to perceive the child's non-verbal cues (e.g., gaze, gestures) via a customizable humanoid avatar displayed on a smartphone screen without needing any specialized wearable devices. The avatar utilizes the standardized VRM format, ensuring interoperability with existing metaverse spaces [4,9], aiming to support a deeper sense of self-expression and identity continuity.

This asymmetric configuration is designed to satisfy two complementary psychological imperatives in remote communication: spatial presence (the subjective sensation of physically occupying the remote environment [10]) and social presence (the mutual perception of "being together with another" [11]). While the high-fidelity 360-degree video projection inside the HMD is intended to support the patient's spatial presence, the real-time, non-verbal feedback loops mediated by the virtual avatar and bidirectional audio aim to support mutual social presence. Thus, our proposed system is designed to maintain both of these dimensions of presence while reducing computational and physical deployment barriers for families.

To translate these theoretical principles into an initial prototype that aims to be practical in clinical situations, the proposed system is designed around three design goals:

1. High Spatial Immersion: The system aims to provide the remote user with spatial immersion via a live 360-degree view of the local environment.
2. Intuitive Co-presence: The system aims to allow local users to perceive the remote user's directional orientation and hand gestures via a mobile avatar proxy.
3. Deployment Simplicity: The system is designed with a guided setup interface to assist non-expert assembly, aiming to reduce technical intervention on either end.

To explore how this prototype performs and how much it satisfies these requirements, we formulated three exploratory research questions (RQs):

RQ1: Does the proposed system provide sufficient spatial presence to remote users, and how do their subjective social presence scores compare descriptively to conventional video conferencing?

RQ2: How do different visual feedback modalities (avatar-only, raw video-only, or combined representation) on the local side affect non-verbal interaction quality, such as gesture transmission and spatial navigation?

RQ3: How do local-side interaction modalities descriptively relate to both remote and local users' subjective social presence scores across conditions?

To address these questions and derive broader practical insights, this paper provides (1) a primary technical contribution comprising the architectural design and technical implementation of the lightweight asymmetric telepresence prototype and its gyroscope-based avatar alignment algorithm, accompanied by (2) secondary exploratory evaluations documenting preliminary observations and operational limitations from a controlled, exploratory laboratory study ($N = 16$, 8 pairs) and a single-case fieldwork ($N = 1$).

Specifically, we first evaluate the proposed system through an exploratory, controlled laboratory usability and presence study utilizing a 4 × 4 Latin square experimental design with 16 healthy participants (8 pairs). Usability is assessed quantitatively and qualitatively using the System Usability Scale (SUS) [12], while spatial and social presence are measured using the Igroup Presence Questionnaire (IPQ) [10] and the Networked Minds Social Presence Questionnaire (NMSPQ) [11], respectively. We also conducted an exploratory single-case field deployment ($N = 1$) in which a long-term hospitalized adolescent used the prototype to remotely view her graduation ceremony. Throughout this paper, we strictly separate what was empirically measured (usability, presence questionnaires, task completion times, and initial network telemetry) from exploratory observations and hypotheses, noting that standardized clinical or psychological outcomes, end-to-end latency, and device temperatures were not measured.

The remainder of this paper is structured as follows. Section 2 introduces the works related to our research. Section 3 describes the overview of the experience through our proposed system. Section 4 describes our materials and methods, including the system architecture and the unique gyroscope-based avatar alignment algorithm. Section 5 describes the experimental protocol for the exploratory laboratory study. Section 6 presents the statistical results of our exploratory laboratory evaluation and discusses the results and answers for the research questions. Section 7 describes the single-case fieldwork using our system and discusses the technical, operational, and practical challenges of introducing such systems into real-world pediatric care settings, addressing real-world limitations. Finally, Section 8 presents our conclusions and outlines future directions.

2. Related Work

2.1. Mitigating Social Isolation with Virtual and Robotic Avatars

Physical and social isolation in healthcare settings significantly impairs patients' emotional resilience and long-term quality of life (QOL). To bridge clinical gaps and restore agency, researchers have investigated robotic and digital proxies. In workplace contexts, Takeuchi et al. demonstrated that bedridden individuals could teleoperate physical robotic avatars ("OriHime-D") in public spaces to perform social labor, mitigating loneliness through societal re-integration [3]. Extending these efforts to educational domains, Page et al. synthesized international deployments of telepresence robots (TPRs), underscoring their vital role in sustaining classroom connectedness and academic inclusion for homebound and chronically ill students [13]. Grounded in this domain, Ahumada-Newhart and Olson conducted an extensive field study across 19 cases, identifying key design constraints and social dynamics when pediatric patients attend schools via mobile robots [14]. Complementing physical embodiment, virtual environments offer alternative psychosocial affordances. Hasei et al. deployed customized digital avatars within commercial virtual spaces for young cancer patients, showing that virtual proxies effectively lowered communication barriers and mitigated appearance-related anxiety caused by treatment [4]. While these physical robots and virtual avatars foster essential connectedness, deploying immersive, real-time omnidirectional telepresence presents distinct experiential opportunities for students participating in authentic milestone events.

2.2. Spatial Telepresence and Remote Collaboration

Spatial telepresence aims to produce the compelling cognitive illusion of "being there" inside a remote physical space, thereby evoking a robust sense of spatial presence. To conceptualize the cognitive mechanisms underpinning this spatial experience, Slater formalized virtual presence into two orthogonal illusions: Place Illusion (PI), the perceptual illusion of physically being in a remote place despite knowing otherwise, and Plausibility Illusion (Psi), the illusion that the observed events are genuinely occurring [15]. While PI is largely sustained through continuous visual and sensorimotor contingencies, Psi critically relies on the credibility of environmental interactions and dynamic responses.

To realize and balance these illusions in collaborative contexts, technical research has historically pushed the boundaries of visual display fidelity, spatial tracking, and interactive rendering. Early efforts, such as "TeleHuman," demonstrated that life-size, cylindrical displays supporting motion parallax significantly enhanced the accuracy of reading a remote partner's gaze, posture, and gestures, thereby reinforcing mutual social presence [16]. To establish collaborative shared workspaces, researchers integrated live 360-degree video with mixed reality headsets (e.g., "Augmented Virtual Teleportation") [5] or optimized photorealistic 360-degree spatial projections [17]. To scale these interactions over consumer networks, systems like "VR2Gather" introduced adaptive point-cloud tiling for multi-party holoconferencing [18], while multi-sensor fusion arrays combined 360-degree cameras with depth sensors to provide 6-DOF head-tracked viewing and suppress simulator sickness [19]. Most recently, neural rendering has unlocked room-scale photorealism; "GaussianNexus" utilized real-time 3D Gaussian Splatting combined with YOLO-based tracking for 6-DOF collaborative spaces [6], and "RT-FVHP" achieved ultra-fast holographic patient rendering for clinical telerehabilitation [7].

Despite achieving unparalleled spatial fidelity, these advanced 6-DOF pipelines require multi-camera grids, expensive GPU clusters, and expert-led spatial calibration [6,7], rendering them entirely impractical for autonomous, long-term deployment in domestic care settings. To address this technical bottleneck, HCI has shifted toward ambient and lightweight paradigms. Kida et al. introduced the "IoT Avatar" concept, demonstrating that remote presence could be projected seamlessly through everyday household items and furniture acting as low-cost physical proxies [8]. This philosophy aligns with the framework of "Passive Co-presence" articulated by Shakeri et al., who explored how smart home devices (e.g., picture frames or heating chairs) can subtly convey a remote loved one's presence in the background of a domestic space without demanding active, screen-focused conversation [20,21]. By organizing passive co-presence across four dimensions—activity, solitude, synchronicity, and spontaneity—they established that peripheral awareness of a remote partner's physical cues successfully supports mutual togetherness while respecting domestic routines and solitude [20].

Our proposed asymmetric 360-degree telepresence system is strategically positioned at the intersection of these two paradigms. Instead of forcing expensive 6-DOF volumetric rendering on families, we implement Kida's minimalist proxy design: the hospitalized child is actively immersed in a high-fidelity 3-DOF 360-degree video feed (maximizing the child's spatial presence), while the family passively monitors the child's non-verbal head gestures through a mobile-based virtual avatar. This asymmetric integration is conceptually motivated by ambient, low-demand passive co-presence using consumer-grade hardware, aiming for future practical use in domestic home environments, although long-term passive co-presence in actual home settings was not empirically tested in the present study.

2.3. Exploratory Psychological Support Options in Virtual Reality Environment

Beyond maintaining social ties, virtual reality serves as a powerful medium for psychological restoration and clinical stress reduction. Environmental psychology has long established that exposure to natural landscapes induces significant cognitive restoration and anxiety reduction; when physical mobilization is restricted, virtual reality (VR) acts as a highly effective surrogate. Sneed et al. demonstrated that immersive exposure to 360-degree virtual nature videos significantly enhanced participants' state of nature connectedness, replicating the psychological restoration curves of physical outdoor contact [22]. This is corroborated by Hejtmánek et al., who proved that exploring photogrammetry-based digital twins of forests successfully elevated positive affect and decreased subjective stress, provided that simulator sickness was minimized [23].

For fragile populations, including older adults and those with cognitive or physical impairments, Kalantari et al. demonstrated that virtual garden environments consistently improved emotional mood states (e.g., calmness, wakefulness, and general affect) [24], and that collaborative social–VR activities in virtual nature further stimulated social presence and meaningful remote connections [25]. These subjective emotional restorations are strongly supported by objective physiological measures; Sacchelli et al. integrated 360-degree forest VR with electroencephalography (EEG) and the Restorative Outcome Scale (ROS) to prove that brief immersive nature exposures physically reset brainwave activity to stress-free baselines [26]. Furthermore, a comprehensive scoping review of 165 publications by Cinnamon et al. underscored that 360-degree videos possess a unique, multi-disciplinary capability to evoke powerful restorative and emotional responses across clinical care, urban planning, and environmental rehabilitation [27].

While these studies motivate the long-term concept of streaming familiar environments (such as home interiors or family gardens), we explicitly note that the present study did not deploy the system in domestic home or garden settings, nor did it test psychological restoration or stress reduction; empirically evaluating any restorative or clinical psychological benefits requires dedicated longitudinal research beyond the scope of this study.

2.4. Evaluation Methodologies

Evaluating innovative telepresence systems requires bridging standardized laboratory metrics with ecologically valid field deployment methodologies.

To scientifically quantify user experience in controlled settings, researchers rely on established psychometric scales. Brooke established the System Usability Scale (SUS) as a standard instrument for assessing overall usability [12]. For spatial immersion, Schubert et al. developed the Igroup Presence Questionnaire (IPQ) to evaluate Spatial Presence, Involvement, and Realism [10], while Tran et al. compiled meta-analytic percentile distributions to contextualize raw IPQ scores against historical benchmarks [28]. Furthermore, to assess interpersonal connection beyond spatial immersion, Biocca et al. established the Networked Minds Social Presence Questionnaire (NMSPQ), conceptualizing social presence across co-presence, psychological involvement, and behavioral engagement [11]. In our laboratory study, combining the SUS, IPQ, and NMSPQ alongside Tran et al.'s distribution provides preliminary, exploratory indicators spanning usability, spatial presence, and social connectedness.

While laboratory evaluations provide necessary baseline metrics, transitioning systems into complex, authentic environments requires an adaptive paradigm. Eriksson et al. conceptualized the integration of Design-Based Research (DBR) in HCI, emphasizing that deploying systems as "technology probes" into real-world contexts generates intermediate-level knowledge that informs both system design and social workflows [29]. Building upon our laboratory findings, our single-case fieldwork adopts this DBR framework, uti-

lizing our system as a probe with external Peltier elements and Phase Change Materials (PCM) to identify bandwidth and hardware bottlenecks that must be resolved for future real-world deployment.

3. Proposed System

The proposed asymmetric 360-degree telepresence system is designed to provide an immersive spatial sharing experience between a remote user (e.g., a hospitalized child) and local users (e.g., family members at home) with minimal cognitive and physical burden as shown in Figure 1. The system establishes a bi-directional communication channel where the experiential value of each user is tailored to their specific environment.

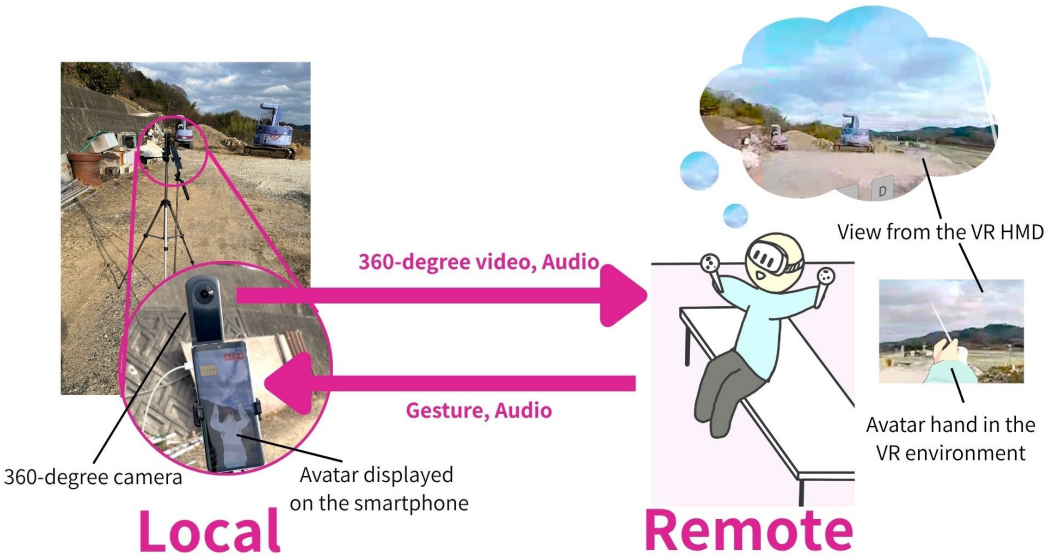


Figure 1. The user experience using our system.

For the remote user situated in a hospital room, the system is designed to offer high spatial immersion. By wearing a Virtual Reality (VR) Head-Mounted Display (HMD), the remote user is presented with a real-time 360-degree spherical video stream captured from the local domestic site. This wide field of view allows the remote user to look around the room naturally by physical head rotation, aiming to evoke a sense of being present in the local domestic site. In addition to high-fidelity spatial video, the remote user can actively express themselves using hand-tracked touch controllers. Their physical head and hand gestures are dynamically mapped to a humanoid avatar rendered in the local space, enabling non-verbal communication. The system also renders a first-person perspective of the avatar's body within the remote VR environment. This co-located visual feedback allows the remote user to monitor their own bodily movements and hand gestures in real time, intuitively understanding how their avatar is positioned and perceived by others at the local site.

For the local users in the domestic space, the system is designed to represent the remote user's directional presence within their physical environment without requiring them to wear any specialized devices. The local setup consists of a 360-degree camera and a smartphone mounted on a manually rotatable mobile stand. The smartphone screen displays the remote user's customizable humanoid avatar. A unique feature of the local interface is that it integrates the camera feed showing the back of the smartphone as an Augmented Reality (AR) background behind the avatar. This visual integration is intended to create the visual impression that the avatar is physically standing in the room. Since the smartphone stand is capable of 360-degree horizontal rotation, local users can physically move around

the setup and rotate the screen to maintain visual contact with the avatar from any direction in the room.

4. System Implementation

4.1. System Architecture

The system architecture employs an asymmetric structure specifically designed to delegate heavy computational rendering and video stitching tasks to the hardware best suited for them, thereby keeping the physical footprint on both ends as lightweight as possible. The system is conceptually divided into three main nodes: the Local (Broadcasting) Node, the Remote (VR) Node, and the intermediate WebRTC Network System as shown in Figure 2.

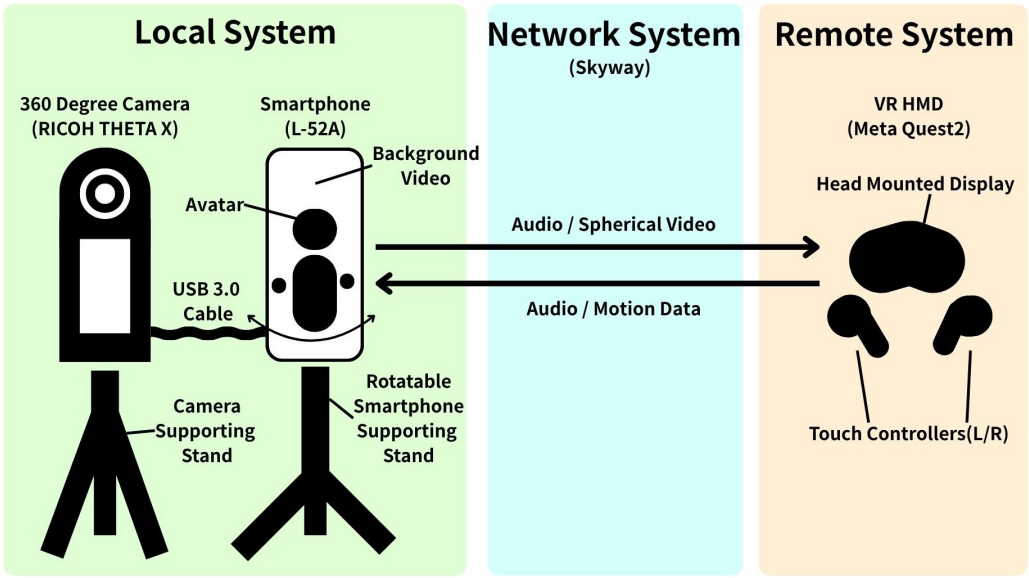


Figure 2. The system architecture of our system.

The local node is responsible for capturing high-resolution spherical video, encoding the multi-stream audio–video payload, and rendering the spatial avatar. The remote node handles the real-time decompression and rendering of the 360-degree skybox texture within the HMD, while continuously polling the high-frequency spatial tracking data (coordinate positions and rotational quaternions) of the user's head and hands.

To bridge these asymmetric nodes with minimal latency, WebRTC (Web Real-Time Communication) is utilized as the primary transmission protocol, which is highly optimized for peer-to-peer (P2P) streaming of high-bandwidth media. Communication is negotiated through NTT Docomo's SkyWay platform. The signaling and media routing topology prioritizes direct P2P connections to minimize latency. Due to hardware and logging constraints on the system, exact end-to-end latency was not quantitatively measured. Instead, interaction quality was examined through exploratory task completion times and qualitative feedback. Also, in restrictive hospital network environments where symmetric NATs or firewalls block direct P2P socket binding, the system dynamically falls back to a TURN (Traversal Using Relays around NAT) server hosted on the SkyWay infrastructure. While this hybrid signaling topology assists initial connection traversal across restricted firewalls, sustained streaming stability remains dependent on available network bandwidth and congestion levels.

4.2. Hardware Implementation

To ensure that the system remains accessible and easily deployable in domestic settings, we chose consumer-grade hardware. The physical components required for both nodes are outlined below:

The Local Node comprises:

- 360-Degree Spherical Camera: A RICOH THETA X camera, which performs real-time, on-device equirectangular stitching of dual fisheye lenses, outputting a high-resolution 2K video stream.
- Broadcasting Smartphone: An Android smartphone connected directly to the RICOH THETA X via a high-speed USB-C to USB-C cable for stable UVC data transfer. The model number is L-52A. The OS version is Android 11. SoC (System on Chip) is Snapdragon 765 G. CPU is Kryo 475 and GPU is Adreno 620. RAM has 6 GB. The battery capacity is 4300 mAh. The preferred environmental thermal range is from 5 degree to 35 degree.
- Smartphone Supporting Stand: A lightweight, portable tripod stand equipped with a manual 360-degree rotatable smartphone mount positioned directly in front of the THETA X lens array.
- Camera Supporting Stand: A portable stand which is capable of holding the spherical camera safely.
- USB Cable: A USB cable to connect the spherical camera and the smartphone to transmit videos in real time.

The Remote Node comprises:

- Head-Mounted Display: A Meta Quest 2 standalone VR HMD, which handles high-fidelity rendering of the 360-degree video sphere and native hand tracking.
- Touch Controllers: Meta Quest 2 Touch controllers to track the hand positions, rotational states, and physical gestures of the patient.

4.3. Software Implementation

The software on both the local and remote nodes was developed using the Unity engine (v2022.3.51f1). Real-time media streaming and metadata synchronization were implemented using the NTT Docomo SkyWay Unity SDK (v0.5.0).

As shown in Figure 3, the software pipeline manages three distinct real-time network streams between the Local and Remote nodes via SkyWay's P2P WebRTC infrastructure. Since we utilize WebRTC, the bitrate for the media streams is adaptive to the bandwidth.

1. Video Stream (Local to Remote): The 360-degree environment captured by the RICOH THETA X (1920 × 960 at 30 fps, 32 Mbps stream H.264 over USB) is ingested via `UVC4UnityAndroid` library (v0.4.0), transcoded into VP8 format at a target bitrate of 2800 kbps by the SkyWay's encoder, and transmitted to the VR HMD.
2. Audio Streams (Bi-directional): Monaural audio captured by both the camera microphone and the VR HMD microphone is compressed using the Opus codec at 32 kbps to ensure low-latency voice communication.
3. Avatar Motion Data Stream (Remote to Local): Spatial tracking metrics from the remote user are serialized into a lightweight binary data stream operating at 3.4 kbps, minimizing network overhead while maintaining constant pose updates.

To balance interactive responsiveness with system resource constraints, avatar pose processing is implemented with a dual-rate update strategy.

- Remote Self-Avatar (VR Node): On the remote VR headset, the patient's self-avatar kinematics are coupled directly to the `XRManager` render loop without frame-rate restrictions, which is intended to maintain responsive visual–motor synchronization (note that self-presence and simulator sickness were not measured in this study).

- Local Display Avatar (Network Transmission): To minimize network bandwidth consumption and reduce computational and thermal overhead on the local smartphone, outbound pose data is intentionally throttled to a transmission frequency of 5 Hz (sent once every 0.2 s). These compressed spatial positions and quaternions are transmitted via the SkyWay DataStream channel.
- Local Interpolation & Rendering (Local Node): On the receiving smartphone, incoming 5 Hz pose packets are parsed and smoothly interpolated within each Unity `Update()` loop, where the spatial orientation is combined with the local gyroscope data as described in Section 4.4.

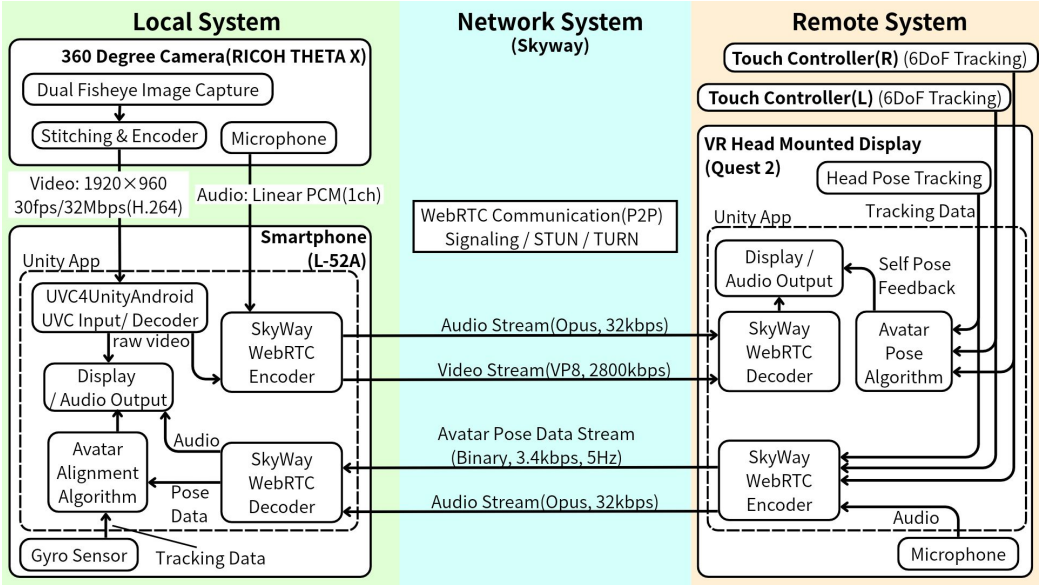


Figure 3. The detailed software architecture of our system.

A major technical challenge during the software implementation on the local Android platform was the hardware resource competition and thread collision between the external UVC (USB Video Class) camera driver and the WebRTC communication stack. If the WebRTC socket connection and the external 360-degree camera feed initialization are invoked simultaneously, the Android operating system frequently suffers from device hanging or immediate app crashes due to asynchronous resource locking. To resolve this critical issue, we implemented a serialized execution pipeline. During initialization, the application strictly prioritizes the hardware handshake and registration of the 360-degree camera via the `UVC4UnityAndroid` library (v0.4.0). The system polls the status of the USB driver until the high-resolution spherical video stream is securely bound to a target render texture in Unity. Only after confirming this visual binding does the system trigger the signaling connection of the WebRTC (SkyWay) client. A corresponding reverse sequence is enforced during system shutdown, where the WebRTC socket is cleanly closed before the local USB camera driver is released. This serialized thread management mitigated runtime device errors.

For the remote user's digital representation, we adopted the standardized `VRM` format. Because modern metaverse and peer-support platforms popular among young patients often utilize the open VRM standard, adopting this format is designed to allow future users to import pre-customized digital avatars into the system. While this capability is intended to support digital identity continuity in future applications, the psychological effect of personalized avatar import was not empirically tested in this study, which used a default avatar.

4.4. Avatar Rotation and Directional Alignment Algorithm

In an asymmetric telepresence setup, maintaining a natural spatial correspondence between the remote user's real-time gaze and the local avatar's orientation is essential for successful social interaction. Because the local smartphone mount is manually rotatable, family members can pivot the screen to face them. However, if the screen rotates physically while the avatar's internal transform remains static, the directional alignment between the remote user's gaze and the local user's physical position becomes mismatched.

To maintain directional consistency, we developed a sensor-fusion alignment algorithm utilizing the smartphone's built-in 3-axis gyroscope. The local Android application continuously samples the gyroscope data to compute the smartphone's physical horizontal yaw rotation angle, denoted as θ_{phone}. Simultaneously, the remote VR application tracks the HMD's global yaw rotation, denoted as θ_{HMD}. The local avatar's rotational transform is dynamically updated on the smartphone screen using the following mathematical formulation:

$$\theta_{avatar} = \theta_{HMD} - \theta_{phone} \quad (1)$$

In this formulation, angles are defined in Unity's left-handed coordinate system around the vertical Y-axis. Yaw angles are sampled in degrees at 60 Hz. To avoid magnetic distortion in indoor environments, initial alignment relies on relative gyroscope accumulation rather than geomagnetic sensors, requiring the smartphone screen to be initially oriented in the same direction as the omnidirectional camera's primary lens during setup. For continuous operation within a typical daily session, we recommend keeping the cumulative angle value within $\pm 10^5$ degrees to guarantee optimal floating-point numeric precision.

By continuously subtracting the smartphone's physical rotation offset from the remote user's tracked HMD gaze angle, the system cancels out any physical rotational displacement. Consequently, when the remote user turns their head to look at a specific object in the streamed 360-degree domestic scene, the local avatar physically faces the corresponding relative direction, allowing local users to perceive a natural spatial gaze and eye contact regardless of the stand's physical rotation. However, we note that mathematical yaw alignment alone does not automatically establish improved subjective social presence or interaction efficiency, which must be examined empirically.

4.5. Setup Assist Interface for Non-Expert Users

To fulfill the requirement of autonomous domestic deployment by non-technical family members, a comprehensive setup assistant user interface was integrated into the local Android application. When launched, the application initiates an interactive, sequential guide consisting of 9 physical and digital assembly steps:

1. Unfolding the Tripod: Guidance on expanding the tripod legs safely.
2. Extending Tripod Legs: Adjusting the height of the main support stand.
3. Leveling the Mount: Calibrating the camera mount platform to ensure a level horizon.
4. Mounting the Smartphone: Securing the Android device onto the rotatable clamp.
5. Securing the 360-degree Camera: Threading and locking the RICOH THETA X onto the tripod screw base.
6. Orienting the Devices: Aligning the physical screen of the smartphone to match the primary lens direction of the THETA X.
7. Powering the Camera: Turning on the RICOH THETA X via its physical power button.
8. Connecting the Cable: Establishing the physical USB-C connection between the camera and the smartphone, with visual confirmation when the system detects the USB connection.

9. Switching to LIVE Mode: Guiding the user through the THETA X on-screen menu to toggle the output mode from default camera storage to "LIVE streaming" mode, enabling active UVC output.

Each of these 9 steps is accompanied by short, loop-playable instructional video clips rendered within the interface. Upon completing the final step, a prominent "Start Stream" button is enabled on the interface, launching the serialized connection pipeline to initiate real-time communication.

5. Experiment

We conducted an exploratory, controlled laboratory study with 16 healthy participants (8 pairs, mean age 22.6 years, SD = 1.3, age range 21–25) to collect preliminary observations on usability, spatial presence, and social presence. In each pair, one participant was designated as the remote user (wearing a Meta Quest 2 HMD), and the other acted as the local user (situated in a separate laboratory room with the 360-degree camera system). The demographic background of the participants was carefully recorded to account for prior technology exposure. Among the 8 local broadcasting-side participants, 4 had prior experience using 360-degree cameras, and 5 had experience using tripods. All 8 remote-side participants had prior experience with VR hardware.

To examine how different local visual feedback representations operate during interaction, we implemented a 4 × 4 Latin square design across four experimental conditions:

- Condition 1 (Zoom Baseline): Conventional 2D video calling on a tablet.
- Condition 2 (Avatar-only): Asymmetric 360-degree telepresence where the local user sees only the remote user's VRM avatar superimposed on the camera feed.
- Condition 3 (Video-only): Asymmetric 360-degree telepresence where the local user sees a flat, 2D video window displaying the remote user's actual face on the smartphone screen.
- Condition 4 (Combined): Asymmetric 360-degree telepresence where the local user sees both the spatial VRM avatar and the adjacent 2D video window showing real-time facial expressions.

The participants completed the conditions in randomized sequences defined by the Latin square groups to eliminate order bias. Each condition lasted approximately five minutes and comprised three structured collaborative activities:

1. Gesture Guessing Game (3 min): The remote user performed 6 distinct physical gestures selected from a randomized pool of standard non-verbal gestures using HMD hand controllers, and the local user attempted to identify them. The elapsed transmission-to-recognition times were recorded.
2. Spatial Navigation Task (1 min): The remote user pointed towards a specific direction in their virtual field of view, and the local user physically moved around the camera to align themselves with the indicated direction. The time to complete the alignment was measured. Note that under Condition 1 (Zoom), spatial direction was communicated verbally, whereas under Conditions 2–4 it was communicated visually via pointing or gaze. Because this procedural difference conflates the visual feedback modality with the communication channel (verbal vs. visual), it directly affects the evaluated construct and prevents isolating the effect of the proposed interface relative to Condition 1.
3. Free Conversation (1 min): General conversation supported by randomized topic cards.

Subjective usability was measured post-trial using the System Usability Scale (SUS) [12]. Spatial presence for remote users was evaluated after their first exposure to the VR system using the 14-item Igroup Presence Questionnaire (IPQ) [10]. Social presence was measured

after each condition using the Mutual Awareness and Empathy subscales of the Networked Minds Social Presence Questionnaire (NMSPQ) [11].

6. Results and Discussion

6.1. System Usability Scale (SUS)

The subjective usability of both the remote (VR-side) and local (broadcasting-side) nodes was evaluated using the System Usability Scale (SUS) after the first operational trial. The individual and average SUS scores are summarized in Table 1.

Table 1. Individual and Mean System Usability Scale (SUS) Scores for the Local (Broadcasting) and Remote (VR) Nodes.

Participant Pair/Metric	Local Node (Broadcasting-Side)	Remote Node (VR-Side)
G1	67.5	92.5
G2	80.0	72.5
G3	82.5	92.5
G4	72.5	97.5
G5	72.5	80.0
G6	82.5	97.5
G7	52.5	97.5
G8	70.0	97.5
Mean ± SD	72.5 ± 9.9	90.9 ± 9.5
Cronbach's α	0.808	0.927

The internal consistency of the SUS evaluations was high for both nodes. The calculated Cronbach's alpha was $\alpha = 0.808$ for the local node and $\alpha = 0.927$ for the remote node, indicating robust reliability and consistent response patterns among the participants.

According to the industry benchmarks established by Lewis and Sauro [30], a SUS score of 68 represents the 50th percentile (average usability), while a score of 80.8 represents the 90th percentile (excellent usability). The average SUS score for the remote VR-side system was 90.9 ± 9.5, which represents an favorable usability profile, placing it well above the 90th percentile. The local smartphone-side system achieved an average SUS score of 72.5 ± 9.9, positioning it above the 50th percentile benchmark.

To rigorously confirm whether these usability scores significantly exceed standard benchmarks, statistical significance testing was performed. First, the distribution of SUS scores was assessed using the Shapiro–Wilk normality test. The local node scores were confirmed to be normally distributed ($p = 0.204 > 0.05$), whereas the remote node scores rejected the assumption of normality ($p = 0.0077 < 0.05$).

Based on these distribution profiles, a parametric one-sample t-test was conducted for the local node to compare its mean score against the 50th percentile baseline (68 points). The analysis showed no statistically significant difference ($t = 1.28, p = 0.12 > 0.05$), indicating that in this small sample (8 pairs), the local system did not significantly exceed the industry average usability threshold. For the remote node, a non-parametric one-sample Wilcoxon signed-rank test was performed with Bonferroni correction ($\alpha = 0.05 \div 2 = 0.025$). The results demonstrated that the remote HMD interface scored significantly higher than both the 68-point average benchmark ($V = 36, p = 0.0065 < 0.025$) and the 80.8-point excellent benchmark ($V = 33, p = 0.0198 < 0.025$).

Additionally, the post-trial questionnaire regarding the local setup process revealed that four out of eight broadcasting-side participants felt the entire physical assembly was perfectly straightforward. However, notable assembly difficulties were reported regarding specific hardware steps; three participants noted that securing the 360-degree camera

onto the tripod screw mount was somewhat difficult or confusing, and two participants highlighted that the written instructional manual required slightly more detail to prevent manual errors. These user reports demonstrate that the local system has not yet achieved robust, independent deployment readiness under realistic conditions, requiring mechanical refinement of the physical mount and instructional materials.

6.2. Spatial Presence Score (IPQ)

The spatial presence (the psychological sense of physical immersion in the remote space) experienced by the VR-side participants was measured immediately after their first exposure to the virtual domestic environment. The subjective evaluation was conducted using the 14-item Igroup Presence Questionnaire (IPQ) [10]. The calculated Cronbach's alpha for the overall IPQ scale was $\alpha = 0.671$. While this internal consistency lies near the lower acceptable threshold for exploratory research, it indicates that findings derived from the IPQ should be interpreted cautiously as preliminary indicators rather than definitive measures. The individual and collective spatial presence scores are detailed in Table 2.

Table 2. Individual and Mean Spatial Presence (IPQ) Scores for Remote VR Users.

Participant Pair	Spatial Presence Score (Range: −3.00 to 3.00)
G1	0.429
G2	1.29
G3	1.36
G4	1.86
G5	0.785
G6	1.21
G7	1.50
G8	2.07
Mean ± SD	1.31 ± 0.53

The average IPQ score across all remote users was 1.31 ± 0.53. For exploratory investigation, this result was compared against a comprehensive historical meta-dataset compiled by Tran et al. [28], which aggregates presence scores from 243 virtual reality studies spanning two decades. In this historical benchmark, the 50th, 75th, 90th, and 95th percentiles correspond to IPQ scores of 0.28, 0.73, 1.07, and 1.30, respectively. Although the observed mean score of 1.31 is marginally above the reported historical 95th-percentile value of 1.30, the 243 studies in Tran et al. remain heterogeneous in hardware, task, exposure duration, context, and participant characteristics. Therefore, given the modest scale reliability ($\alpha = 0.671$) and small sample size (8 pairs), this score does not establish that the proposed system belongs to the top 5% of VR systems in any comparative performance sense, serving solely as a preliminary descriptive reference.

6.3. Social Presence Score (NMSPQ)

To capture the sense of co-presence and psychological connection between the interlocutors, we administered the Networked Minds Social Presence Questionnaire (NMSPQ) [11] to both remote and local users across all four experimental conditions. The scores for the Mutual Awareness and Empathy subscales are detailed in Table 3.

Table 3. Comparison of NMSPQ Social Presence Subscale Scores (Mean ± SD) across Baseline and Proposed Conditions.

Subscale & User Group	Condition 1 (Zoom)	Condition 2 (Avatar)	Condition 3 (Video)	Condition 4 (Combined)
Mutual Awareness				
Remote (VR) Users	5.88 ± 0.98	6.25 ± 0.74	6.02 ± 0.88	6.15 ± 0.60
Local (Home) Users	6.10 ± 0.44	5.69 ± 1.00	5.19 ± 1.11	6.21 ± 0.61
Empathy				
Remote (VR) Users	5.12 ± 0.91	5.65 ± 0.62	5.52 ± 0.81	5.60 ± 0.51
Local (Home) Users	6.06 ± 0.64	5.17 ± 1.37	5.52 ± 1.06	5.65 ± 1.10

The internal consistency of the NMSPQ was verified to be consistently high. For local users, the overall Cronbach's alpha was $\alpha = 0.924$ (Mutual Awareness: $\alpha = 0.864$, Empathy: $\alpha = 0.943$). For remote users, the overall coefficient was $\alpha = 0.843$ (Mutual Awareness: $\alpha = 0.758$, Empathy: $\alpha = 0.862$).

Since the Shapiro–Wilk tests rejected the assumption of normality for nearly all data subgroups ($p < 0.05$), non-parametric Wilcoxon signed-rank tests were performed to compare each of the proposed telepresence conditions (Conditions 2, 3, and 4) against the baseline Zoom setup (Condition 1). Because three comparative tests were executed within each dataset, a Bonferroni correction was applied, establishing an adjusted significance threshold of $\alpha = 0.05 \div 3 = 0.0167$.

The statistical tests revealed no significant differences in social presence scores between the proposed asymmetric telepresence conditions and the Zoom baseline for either user group. For instance, comparing the local user's Mutual Awareness in Condition 3 against Condition 1 yielded $p = 0.0234$, which did not meet the conservative Bonferroni-corrected threshold. However, because the small sample of eight pairs substantially limits statistical power and the precision of estimated effects, the absence of statistically significant differences must not be interpreted as evidence that the proposed system and conventional video conferencing are equivalent.

Descriptively, the mean scores exhibited a noticeable asymmetrical pattern between remote and local users. For Mutual Awareness, remote VR users reported slightly higher descriptive means across all three proposed conditions (6.25, 6.02, and 6.15) compared to the Zoom baseline (5.88). Conversely, local domestic users reported lower descriptive scores in the single-modality conditions (5.69 in Avatar-only and 5.19 in Video-only) compared to Zoom (6.10), reaching a similar descriptive level (6.21 ± 0.61) only in the Combined condition.

Regarding overall visual feedback preferences, local users were asked which of the three proposed telepresence feedback modalities they preferred, excluding the Zoom baseline (Condition 1) since this questionnaire intended to reveal which feedback system was ideal for the proposed system. As multiple selections were permitted for context-dependent preferences, the eight local users cast a total of 10 votes (one participant voted all three modalities depending on the situational context). The subjective votes were distributed as follows: three votes for Avatar-only feedback (Condition 2), three votes for Video-only feedback (Condition 3), four voted for Combined feedback (Condition 4). This distribution suggests that each visual representation possesses unique strength and limitations that may complement each other when combined.

6.4. Quantitative Task Performance

To objectively evaluate the quality of non-verbal communication, we recorded quantitative performance metrics for the collaborative activities under each experimental condition, as shown in Table 4.

Table 4. Quantitative Performance Metrics across Experimental Conditions (Mean ± SD).

Performance Metric	Condition 1 (Zoom)	Condition 2 (Avatar)	Condition 3 (Video)	Condition 4 (Combined)
Gesture Guessing Time (s)	69.0 ± 30.1	141.6 ± 48.1	103.0 ± 55.8	112.8 ± 58.7
Uncompleted Gestures (n)	0	2	2	2
Spatial Navigation Time (s)	18.1 ± 10.1 *	19.3 ± 8.9	23.9 ± 9.6	24.3 ± 20.5

* Verbal instruction used in Condition 1; confounded with communication channel and not directly comparable to Conditions 2–4.

In the Gesture Guessing Game, remote users attempted to transmit six standardized physical gestures within a three-minute time limit. Under the Zoom baseline (Condition 1), all pairs successfully completed all six gestures, achieving the fastest average transmission time of 69.0 ± 30.1 s. In contrast, Conditions 2, 3, and 4 each had two instances where a gesture could not be completed within the time limit (recorded as 180 s for analysis). The average guessing times for Conditions 2, 3, and 4 were 141.6 ± 48.1 s, 103.0 ± 55.8 s, and 112.8 ± 58.7 s, respectively.

To evaluate whether the visual feedback modality significantly affected gesture transmission speeds, a Shapiro–Wilk test was performed on the completion times. Since Condition 3 violated normality, we conducted a non-parametric Friedman test across Conditions 2, 3, and 4. The test showed no statistically significant differences in gesture guessing times among the three proposed modalities ($p = 0.394$).

For the Spatial Navigation Task, we measured the time required for the local user to physically align themselves with a direction pointed out by the remote user. Under the Zoom baseline, spatial direction was communicated verbally, serving as a reference, and is not comparable to Conditions 2–4. The average alignment times under the asymmetric telepresence conditions were 19.3 ± 8.9 s for Condition 2, 23.9 ± 9.6 s for Condition 3, and 24.3 ± 20.5 s for Condition 4. Since the Shapiro–Wilk test indicated that Condition 4 deviated from normality ($p = 0.0067 < 0.05$), a Friedman test was conducted across the three proposed conditions, showing no statistically significant difference in directional alignment times among the modalities ($p = 0.542$). Together with the gesture guessing results, these analyses do not demonstrate statistically significant differences among the three proposed modalities and provide limited evidence that any specific modality improves non-verbal communication performance.

6.5. Discussion: Evaluation of Research Questions from Laboratory Findings

The exploratory controlled laboratory evaluation yielded preliminary structured empirical data, allowing us to examine our three core research questions while clarifying the boundaries of our evidence.

Regarding RQ1 (spatial presence and social presence comparison), the proposed system provided exploratory indication of spatial immersion for remote users. Remote VR users reported a positive spatial presence score (IPQ score of 1.31 ± 0.53), suggesting that real-time communication, combined with a live 360-degree spherical camera feed, can support a cognitive cue of "being there" [10]. However, as noted in Section 6.2, because the historical VR benchmarks in Tran et al. [28] reflect heterogeneous hardware, tasks, and exposure durations—and given our scale reliability ($\alpha = 0.671$) and small sample size (8 pairs)—this

score serves as a preliminary exploratory indicator rather than evidence of comparative performance superiority. Regarding social presence, as detailed below for RQ3, subjective scores showed descriptive variations between user sides rather than uniform statistical equivalence to conventional video conferencing.

Regarding RQ2 (non-verbal interaction task performance across modalities), our quantitative findings revealed that task completion times did not differ significantly among the three proposed modalities for either gesture guessing ($p = 0.394$) or spatial navigation ($p = 0.542$, Table 4). While the static avatar (Condition 2) provided spatial directionality via gyroscope sensor fusion and the video feed (Condition 3) conveyed facial expressions, neither single modality significantly accelerated task completion compared to the Combined feedback (Condition 4). Furthermore, because the spatial navigation task relied on verbal instructions under the Zoom baseline but used visual gesture/pointing under the proposed conditions, this procedural difference conflates the communication channel with the visual feedback modality. Consequently, direct performance comparisons against Zoom cannot isolate the effect of the proposed interface, and our findings describe observational trends rather than evidence of superior task efficiency.

Regarding RQ3 (how local-side visual feedback modalities relate to remote and local social presence), the NMSPQ results (Table 3) revealed a notable asymmetrical pattern in mutual awareness. Remote VR users maintained consistently high descriptive scores across all proposed conditions (6.25, 6.02, and 6.15) relative to Zoom (5.88). We hypothesize that for the remote user, full-surround 360-degree spatial immersion combined with the knowledge that an avatar represents their physical presence in the room may provide a baseline sense of co-presence regardless of the local display mode. However, this underlying perceptual mechanism was not directly tested.

In contrast, local domestic users reported lower descriptive mutual awareness in single-modality conditions (5.69 for Avatar-only and 5.19 for Video-only) compared to Zoom (6.10), recovering only in the Combined condition (6.21 ± 0.61). Strictly separating what was measured from what is hypothesized, our empirical data are limited to these descriptive mean differences in self-reported NMSPQ scores, which did not reach statistical significance after Bonferroni correction. As an unverified post hoc hypothesis, rather than a tested cognitive mechanism, we speculate that this descriptive pattern may reflect a trade-off between spatial directionality (conveyed by the rotatable avatar in Condition 2) and facial expressions (conveyed by the 2D video in Condition 3), where presenting only one cue type might offer incomplete non-verbal context for local users.

Because our analysis compares descriptive means in a small sample ($N = 16$, 8 pairs) without measuring cognitive load or attention allocation, these results cannot establish causal mechanisms of social presence or prove equivalence to traditional video calling; rather, they suggest a testable hypothesis for future controlled experiments that integrating both spatial orientation and facial cues may help support local-side mutual awareness.

Methodological Limitations and Unmeasured Factors

A major experimental limitation is the asymmetry in task communication channels: the spatial navigation task relied on verbal instructions under the Zoom baseline but used visual gesture/pointing under the proposed conditions. Consequently, direct task performance comparisons between Zoom and the proposed system should be interpreted as exploratory baselines rather than controlled comparative benchmarks. Because the measured performance differences conflate the visual feedback modality with the communication channel, the study cannot isolate the effect of the proposed interface relative to Zoom. Furthermore, the small sample size ($N = 16$, 8 pairs) substantially limits statistical power and the precision of effect estimates; thus, the absence of statistically significant differences in NMSPQ or

task scores must not be interpreted as evidence of statistical equivalence between modalities or against video conferencing. Finally, the healthy adult student cohort does not reflect the physical, cognitive, or emotional constraints experienced by long-term hospitalized pediatric patients.

To clearly delineate the boundaries of our empirical evidence, we note what was not measured or tested in this study across both the laboratory evaluation and the fieldwork. First, regarding system and network performance, we did not quantitatively measure exact end-to-end (glass-to-glass) latency, internal device temperatures of the smartphone or 360-degree camera, CPU/GPU utilization, or power consumption. Second, regarding perceptual and human-factors mechanisms, we did not measure objective attention allocation (e.g., via eye tracking), cognitive load, perceived self-presence, simulator sickness, or the psychological effect of importing personalized VRM avatars. Third, regarding clinical and psychological outcomes, we did not administer standardized psychometric or clinical scales (e.g., for depression, anxiety, stress restoration, or quality of life), nor did we evaluate long-term usage in actual domestic home environments. Consequently, all discussions of cognitive trade-offs, thermal throttling, and emotional support in this paper represent exploratory observations or hypotheses rather than empirically demonstrated effects.

7. Fieldwork

7.1. Overview of the Case and Preparation

By utilizing our system, we conducted an exploratory fieldwork in March 2026 to support a long-term hospitalized patient in attending her graduation ceremony remotely in case her physical condition precluded in-person participation. She was 15 years old and had been undergoing treatment for two years and she spent a total of one year in the hospital.

We first had a discussion with physicians at Okayama University whether our system could be utilized in the medical field. Through these dialogues, the medical team identified this patient as someone facing a potential barrier to attend her graduation ceremony. In response, we proposed a contingency plan to enable remote attendance should her health condition preclude onsite participation.

To ensure the patient and her family could make a well-informed, unpressured decision, an informational pamphlet disclosing full details of the proposal and system overview was provided via her attending physician. Following a deliberate consideration period within the family, the patient and her parents expressed a strong desire to participate, granting their initial verbal informed consent and assent to try the system as an optional safety net (the subsequent post-event written consent for the questionnaire and the overarching institutional ethics approval obtained in July 2026 are detailed in the Institutional Review Board and Informed Consent Statements). Upon confirming their intent to participate, we consulted the school authorities and secured formal permission from the school authorities to deploy the system in the school gymnasium.

To accommodate the day-to-day volatility of her health, we collaboratively established three operational scenarios: (1) full onsite attendance if her condition permitted, (2) a hybrid model (e.g., attending the formal ceremony remotely and joining the homeroom gathering in person), and (3) full remote participation via our telepresence system. The operational setup was flexibly designed to adapt to the patient's immediate physical readiness on the day of the ceremony.

7.2. Rehearsal Phase: System Verification and Initial Immersion

Approximately two weeks before the graduation ceremony, we conducted a rehearsal to verify if our system works well between the school and the hospital and if our system is adequate for the patient to use for her graduation ceremony. In the patient's classroom,

our "local system", a set of 360 degree camera and a smartphone mounted on tripods, was installed. In the hospital, the patient wore the VR HMD to immerse herself in the school environment. One of the authors (Y.G.) set up the equipment for the rehearsal with the cooperation of the classroom teacher. One of the authors (J.H.), a physician, supported the patient to wear the VR HMD with the patient's parents in the hospital.

During this five-minute rehearsal, bidirectional audio-visual streaming and avatar orientation operated without noticeable interruption between the school and the hospital, confirming basic technical connectivity for the upcoming event. To prioritize the patient's physical condition and the graduation ceremony itself, the rehearsal interactions were kept to a minimum, and feedback regarding the rehearsal was collected retrospectively through an optional questionnaire after the ceremony.

7.3. Graduation Ceremony Deployment

7.3.1. Deployment Setup and Mobile Web Client

After the rehearsal, a new option as "(2) a hybrid model" to watch the ceremony in the car going to school in the morning to attend the classroom physically emerged considering her physical condition. Since it was hard to wear VR HMD in the car, we created a light-weight mobile web version of the remote system which receives and shows the omnidirectional video, as well as the local ambient sound. The web application utilized gyro sensor in the smartphone, so that the user can watch anywhere by just rotating the smartphone to the direction the user wants to see. The application also required a password since this 360-degree video streams many students in the ceremony. The password was told to the patient and her parents securely using an A4-size paper with QR code for the web application on it.

This web application was implemented with HTML5 and Javascript. A-Frame (v1.4.0) was used to show omnidirectional video and to control viewpoint orientation via WebXR. Skyway Javascript SDK (v2.4.1) was used for real-time media streaming. This application was deployed using github pages.

7.3.2. Ceremony Operation and Telepresence Experience

On the ceremony day, the second scenario, "(2) a hybrid model", watching the ceremony in the car and also at home was adopted as a streaming method. The web application was utilized in the iOS smartphone to watch the ceremony.

We installed the local streaming system as shown in Figure 4 in the school gymnasium. We put our system at the patient's seat in order for the patient to feel as if she was present there. We used mobile WiFi for the network environment. The seat and table for the researchers were set in front of the system so that we could fix the troubles immediately. A personal computer was set on the researchers' table and connected Zoom streaming for backup in case our system went into troubles. We also utilized power supply to keep the battery alive for the entire ceremony. To mitigate potential thermal buildup on the smartphone and spherical camera during continuous streaming (which we hypothesized could contribute to process stalls, although internal device temperatures were not quantitatively logged), we attached a Peltier cooler to the smartphone and applied a phase change material (PCM) pack (melting point: 28 degree) to the camera. All these settings' positions were decided through the discussion with teachers at school not to prevent other students' movement and vision. The streaming time lasted for 45 min.

7.4. Post-Event Evaluation and User Feedback

We requested a completely optional feedback for the whole experience of this exploratory field study. We asked the patient and patient's parents to read and answer the questionnaires

together including written informed consent. The written informed consent was obtained from patient and patient's parents prior to these questionnaires. To minimize the burden of the patient, all the items were optional to answer.

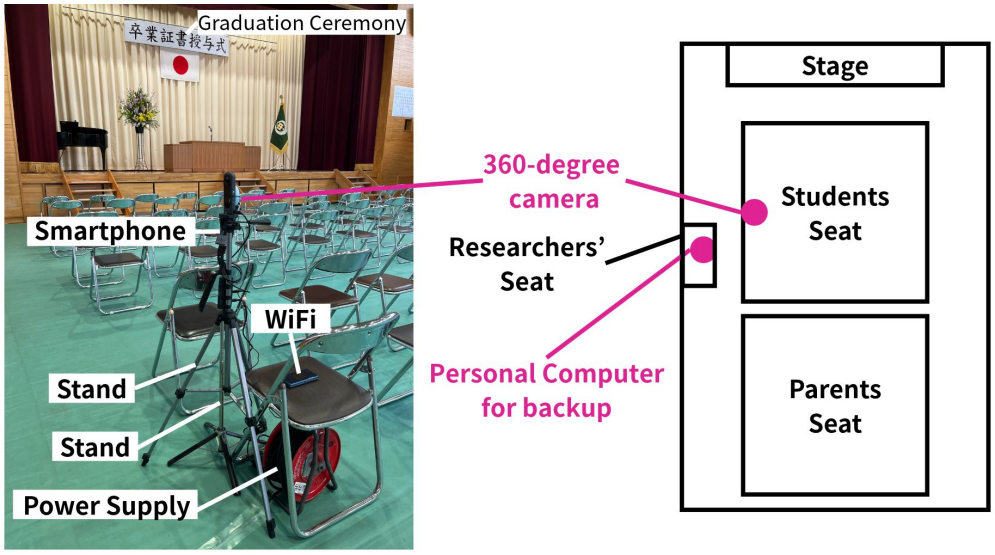


Figure 4. Settings for the online graduation ceremony.

The questionnaires consisted of six sections: (1) written informed consent with a checkbox, (2) basic information without identifiable personal data (e.g., age, gender, duration of treatment and hospitalization), (3) subjective pre-event impressions from the initial proposal to the event day (4) feedback on the rehearsal, (5) feedback on the online graduation ceremony, and (6) overall experience. Questionnaires consisted of 5-point scale questionnaires and free comments.

7.4.1. Pre-Event Subjective Impressions

When evaluating the patient's subjective impressions during the preparation period, specifically from the initial proposal up to the day of the ceremony, the participant selected the highest rating, reporting that she "felt extremely positive." In the open-ended response explaining this rating, the patient noted that simply having the option to view the ceremony online made her feel somewhat better when in-person attendance was uncertain:

> *Even though I couldn't attend the graduation ceremony in person, I felt a little better when I found out I could watch it online.*

While this remark is meaningful as an individual account of how the participant perceived the option to participate in the ceremony online, a single-case retrospective report without standardized psychometric scales cannot establish a causal psychological benefit, a sustained improvement in well-being or clinical effectiveness.

7.4.2. Feedback on the Rehearsal Phase

In the post-event survey regarding the rehearsal, all four items evaluating telepresence were rated within the top two positive tiers: two items received "Agree" ("felt like being at school" and "felt like being with the teacher"), while the other two achieved the highest rating, "Strongly Agree."

Notably, the participant marked "Strongly Agree" both for experiencing a heightened sense of presence compared to conventional video conferencing and for feeling that their

presence was successfully conveyed to the teacher. This latter finding is particularly notable from a telepresence standpoint. In the client VR interface, the patient only observed the omnidirectional live stream of the classroom overlaid with their own avatar from a first-person perspective, without any self-view monitor indicating how they appeared on the classroom's display. Despite the absence of visual feedback confirming her own outward appearance, the participant reported feeling that her presence was acknowledged by the teacher during the brief rehearsal.

This individual observation suggests that first-person avatar viewing was acceptable to this participant without a self-view monitor, while a single five-minute rehearsal cannot establish that this configuration fosters superior reciprocal social presence compared to conventional tools like Zoom.

Furthermore, an open-ended inquiry on future applications revealed an unexpected finding: while the authors initially hypothesized a demand for continuous home telepresence, the participant expressed a stronger personal preference to attend collective school events such as sports festivals and cultural festivals. Although a single participant's preference cannot establish the general socio-emotional needs of long-term hospitalized adolescents as a population, it suggests collective school milestones as a relevant context for future investigation.

7.4.3. Perceptions of the Live Graduation Ceremony

In the evaluation of the actual ceremony, the participant reported a generally favorable experience, rating overall satisfaction, usability, and the heightened sense of presence compared to video calls all within the second-highest tier ("Agree").

However, an accompanying qualitative estimate indicated a severe technical limitation: the participant noted that smooth, freeze-free video streaming was maintained for only approximately 20% of the entire session. Despite such substantial streaming interruptions and the lightweight viewing format (i.e., accessed via a mobile web browser rather than an HMD), both overall satisfaction and perceived spatial presence remained consistently positive.

One contextual factor behind this positive evaluation is reflected in qualitative feedback from the parent, who commented that the participant "seemed delighted to be able to actively change her viewpoint from her seat to look for her friends." This suggests that intermittent 360-degree views still offered meaningful situational agency and partial ambient awareness. However, because smooth streaming was maintained for only approximately 20% of the ceremony, this single-case observation ($N = 1$) represents a preliminary proof of concept rather than evidence of reliable technical performance or emotional superiority over traditional video conferencing. The technical root causes of these interruptions—spanning network stability, thermal constraints, and operational telemetry—are critically analyzed in the following subsection.

7.5. Design Insights and Field Challenges

Grounded in the methodology of Design-Based Research (DBR), this exploratory fieldwork was not merely a passive system test, but a responsive deployment conducted in complex, authentic environments (hospital and school). In collaboration with physicians, educators, and the family, the design evolved iteratively—shifting from an initial VR HMD setup during the rehearsal to a lightweight mobile WebXR client to accommodate the physical volatility of the patient on the ceremony morning. From this deployment cycle, we extract critical engineering challenges and resulting design principles.

7.5.1. Pre-Event Network Profiling and Carrier Selection

To mitigate cellular variability in the gymnasium, we conducted a multi-carrier throughput benchmark on-site prior to the ceremony, recording a total of 11 measurements between 8:20 AM and 9:13 AM. We evaluated mobile Wi-Fi routers across three major commercial carriers (SoftBank, $n = 5$; NTT DOCOMO, $n = 5$; and au, $n = 1$).

The initial trial with the au link was immediately discarded due to insufficient uplink throughput for high-resolution 360-degree transmission. Between the remaining two candidates, the SoftBank link demonstrated the highest consistency; with the exception of the initial baseline run, it sustained an uplink throughput of approximately 20 Mbps, which comfortably satisfied the required baseline for the WebRTC stream. Consequently, SoftBank was selected as the primary uplink.

7.5.2. Field Challenges and Failure Dynamics

Despite the robust pre-event benchmark (20 Mbps uplink), real-time transmission suffered severe degradation during the live ceremony, leading to an unprecedented frequency of video freezing that contrasted with prior continuous runs of approximately one hour. This degradation coincided with the physical arrival of hundreds of attendees (parents and guests) in the gymnasium, which rapidly saturated the local cellular base stations.

To understand the operational breakdown under this heavy network contention, we analyzed the WebRTC telemetry during the initial seven minutes of the ceremony stream:

- Bitrate Oscillations and Jitter: Continuous jitter hovered between 20 ms and 43 ms, accompanied by severe wave-like oscillations in video bitrate fluctuating between 513.4 kbps and 3209.2 kbps, alongside bursty packet loss peaking at 0.6%.
- Frame-rate Depletion: The received frame rate, which initially held at approximately 20 fps, experienced sharp drops down to 13 fps. While it briefly recovered after the initial drop, a subsequent drop to 15 fps failed to recover, culminating in stream termination with the bitrate bottoming out at approximately 700 kbps.

While definitive causal attribution is constrained by the absence of local device temperature logs, these telemetry signatures point to a plausible dual-factor failure mode. On one hand, persistent jitter and aggressive bitrate throttling indicate that WebRTC congestion control was struggling to cope with sudden cellular bandwidth depletion. On the other hand, real-time 360-degree encoding imposes severe computational and thermal strain on mobile hardware. One possible explanation is that excessive packet retransmission cycles under fluctuating network conditions compounded the local encoding overhead, pushing the capturing hardware beyond its thermal or operational threshold and triggering recurrent process stalls despite the physical Peltier and PCM cooling units. Because local device temperature logs were not recorded, thermal throttling and computational overload are presented strictly as a plausible hypothesis rather than a confirmed failure mechanism.

7.5.3. Design Principles for Hospital–School Telepresence

Reflecting on these challenges through the DBR lens, we formulate three core design principles for future rapid-deployment telepresence:

1. Resilient Multi-Path Uplink: Relying on a single commercial cellular link creates an acute vulnerability in public gatherings. Future architectures must integrate multi-SIM network bonding or prioritize ultra-low-bandwidth spatial fallback streams.
2. Asymmetric Feedback and Social Presence: In the brief rehearsal, the participant felt acknowledged without a self-view monitor, suggesting that future studies with larger samples should examine whether first-person avatar alignment can support perceived co-presence while reducing client-side interface clutter.

3. Adaptive Multi-Modal Client Flexibility: As clinical situations shift rapidly, telepresence systems must support agile transitions across display modalities (e.g., from immersive HMDs to mobile WebXR via standard browsers) without imposing software installation barriers on the patient.

8. Conclusions and Future Work

This paper presented the design and technical implementation of an interactive, asymmetric 360-degree telepresence prototype system aiming to reconnect long-term hospitalized pediatric patients with their domestic and educational environments alongside exploratory evaluations from a laboratory study ($N = 16$, 8 pairs) and a single-case graduation ceremony deployment ($N = 1$). In the laboratory study ($N = 16$, 8 pairs), remote users exhibited positive preliminary spatial presence scores (1.31 ± 0.53 on the IPQ), providing an initial descriptive indication of perceived "being there" in our laboratory setting, though without establishing comparative performance superiority over historical VR literature given the modest internal consistency ($\alpha = 0.671$), small sample size, and heterogeneity of historical VR studies. For social presence, while no statistically significant differences were found across conditions, an asymmetrical descriptive trend emerged; remote users reported descriptive mutual awareness means of 6.25, 6.02, and 6.15 across the proposed conditions (vs. 5.88 in Zoom), whereas local users reported lower descriptive scores in the single-modality conditions (5.69 in Avatar-only and 5.19 in Video-only vs. 6.10 in Zoom), reaching a comparable descriptive mean (6.21 ± 0.61) only in the Combined modality. Regarding non-verbal task performance, completion times did not differ significantly among the three proposed modalities for either gesture guessing ($p = 0.394$) or spatial navigation ($p = 0.542$), and direct spatial navigation comparisons against Zoom were confounded by the verbal versus visual communication channel, providing limited evidence that any specific modality improves non-verbal communication efficiency. In terms of usability, remote users rated the HMD interface very favorably (90.9 ± 9.5), while the local setup demonstrated a mean SUS score (72.5 ± 9.9), which did not significantly exceed the 68-point industry average benchmark ($p = 0.12$), with four of eight local participants reporting camera mounting or instructional difficulties that must be resolved before independent deployment.

In the single-case fieldwork ($N = 1$), the remote participant reported favorable subjective ratings in a post-event questionnaire regarding the opportunity to view her graduation ceremony online, despite severe cellular network congestion that limited smooth streaming to approximately 20% of the session. Importantly, because this study did not measure standardized clinical or psychological outcomes, end-to-end latency, or device temperatures, these exploratory observations do not establish statistical equivalence to conventional video conferencing, prove causal psychological or clinical efficacy, or demonstrate independent deployment readiness; rather, they serve as an initial feasibility probe that clarifies critical technical, operational, and network boundaries for future deployment.

These empirical and ecological findings define three critical directions for future work:

- Bridging Unrepeatable School Milestones: Motivated by the participant's individual interest in collective school events (e.g., sports or cultural festivals), future studies with larger samples should examine the technical and interaction requirements for supporting remote participation in group educational contexts.
- Resilient Network Bonding and Low-Overhead Pipelines: To mitigate the severe cellular congestion observed during large-scale public gatherings, future architectures will incorporate multi-SIM bonding, device temperature logging, and adaptive fallback mechanisms, coupled with streamlined native mobile capture and simplified camera mounting.

- Enriching Reciprocal Social Presence: To address the asymmetric social presence observed on the local side and rigorously isolate modality effects, future controlled experiments will standardize task communication channels across conditions and explore facial expression tracking directly on the avatar, enriching emotional fidelity without adding visual clutter or computational overhead on mobile clients.

By addressing these identified technical and operational challenges, we aim to refine the proposed asymmetric telepresence system toward robust and verifiable deployment readiness. Establishing such technical reliability and operational usability will be a necessary prerequisite before conducting future longitudinal studies to evaluate whether the system can meaningfully support remote social and educational participation for hospitalized patients.

Author Contributions: Conceptualization, Y.G., J.H. and T.N.; Methodology, Y.G.; Software, Y.G.; Validation, Y.G. and T.N.; Formal Analysis, Y.G.; Investigation, Y.G. and J.H.; Resources, J.H. and T.N.; Writing—Original Draft Preparation, Y.G.; Writing—Review & Editing, Y.G., J.H. and T.N.; Visualization, Y.G.; Supervision, T.N.; Project Administration, Y.G., J.H. and T.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The laboratory usability study did not require formal ethical approval as it was a non-invasive, minimal-risk evaluation conducted entirely within the laboratory environment involving research lab members as voluntary participants. The single-case field deployment in March 2026 was conducted as an urgent, compassionate support initiative to facilitate the patient's remote participation in her graduation ceremony. Using non-invasive telecommunication technology similar to standard remote video communication, this deployment was implemented in consultation with the attending medical team, with pre-event verbal assent/consent and post-event written consent for the optional survey and reporting obtained from the patient and her family. The comprehensive research protocol covering the telepresence setup, observational methods, and questionnaire administration was reviewed and approved by the Ethics Committee of Okayama University Hospital (protocol code: 2607-022; date of approval: 24 July 2026). The anonymized qualitative feedback and observational records from the March 2026 deployment are reported in line with the ethical and privacy standards established by this approved protocol.

Informed Consent Statement: For the laboratory usability study, written informed consent was formally obtained from all participating laboratory members prior to the experiments. For the single-case field deployment, a comprehensive two-stage consent process was rigorously implemented. Prior to the event, an informational pamphlet detailing the system, procedural considerations, and potential operational risks (including the possibility of unsuccessful coordination with the school) was provided via the medical team. The materials explicitly emphasized that the telepresence system served merely as an optional safety net—with the patient's in-person attendance remaining the foremost goal—ensuring that consenting to the trial imposed no obligation to utilize the remote system should onsite participation become viable. The materials also stated that participation, future evaluation questionnaires, and academic publication were entirely voluntary, with the explicit right to withdraw or cancel at any time without prejudice. Following a joint deliberation period within the family, initial verbal consent from the parents and verbal assent from the patient were secured. Following the ceremony, written informed consent for the evaluation questionnaire and publication was formally obtained. To maximize privacy protection, an anonymized format was adopted where the parents confirmed agreement via a checkmark rather than personal signatures, following a flexible deliberation period of approximately one week with the patient, during which no strict deadline was enforced. All questionnaire items were clearly marked as optional, and all documentation explicitly guaranteed that no personally identifiable information would be published.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request, subject to ethical and privacy restrictions concerning participant confidentiality.

Acknowledgments: During the preparation of this manuscript, the authors utilized Gemini (Google, versions 3.6, 3.7, and 3.8 Flash) for Japanese-to-English translation support, structure mapping, and draft compilation, as well as NotebookLM (Google) for cross-referencing and verifying source literature. The authors have reviewed and edited the output and assume full responsibility for the content of this publication. The authors would like to express their deepest gratitude to Hirokazu Tsukahara and Hisashi Ishida at Okayama University for their invaluable clinical advice and support. In particular, we are deeply grateful to Ishida for his clinical coordination, specifically in facilitating the delivery of the system pamphlets and informational materials to the pediatric patient and her family. The authors would also like to thank the teachers at the patient's school for accepting our fieldwork, helping us set up the equipment, and supporting this practical deployment. The authors would like to express their profound gratitude to Kazuma Tora for connecting the authors and facilitating the initiation of this collaborative research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Katanoda, K.; Shibata, A.; Matsuda, T.; Hori, M.; Nakata, K.; Narita, Y.; Ogawa, C.; Munakata, W.; Kawai, A.; Nishimoto, H. Childhood, Adolescent and Young Adult Cancer Incidence in Japan in 2009–2011. *Jpn. J. Clin. Oncol.* **2017**, *47*, 762–771. [CrossRef] [PubMed]
2. National Cancer Center Hospital. Pediatric Blood and Lymph Cancers. 2025. Available online: https://www.ncc.go.jp/jp/rcc/about/pediatric_leukemia/index.html (accessed on 5 September 2026).
3. Takeuchi, K.; Yamazaki, Y.; Yoshifuji, K. Avatar Work: Telework for Disabled People Unable to Go Outside by Using Avatar Robots "OriHime-D" and Its Verification. In *Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, HRI '20*; ACM : New York, NY, USA, 2020; pp. 53–60. [CrossRef]
4. Hasei, J.; Ishida, H.; Katayama, H.; Maeda, N.; Nagano, A.; Ochi, M.; Okamura, M.; Iwata, S.; Ikuta, K.; Yoshida, S.; et al. Utilizing the Metaverse to Provide Innovative Psychosocial Support for Pediatric, Adolescent, and Young Adult Patients with Rare Cancer. *Cancers* **2024**, *16*, 2617. [CrossRef] [PubMed]
5. Rhee, T.; Thompson, S.; Medeiros, D.; dos Anjos, R.; Chalmers, A. Augmented Virtual Teleportation for High-Fidelity Telecollaboration. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 1923–1933. [CrossRef] [PubMed]
6. Huang, X.; Frehlich, D.; Xia, Z.; Gholami, P.; Xiao, R. 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, Busan, Republic of Korea, 28 September–1 October 2025 ; pp. 1–18. [CrossRef]
7. Cao, S.; Zhao, J.; Hu, F.; Gan, Y. Real-Time, Free-Viewpoint Holographic Patient Rendering for Telerehabilitation via a Single Camera: A Data-Driven Approach With 3D Gaussian Splatting for Real-World Adaptation. *IEEE Trans. Vis. Comput. Graph.* **2025**, *31*, 7311–7323. [CrossRef] [PubMed]
8. Kida, Y.; Chiba, T.; Kinoshita, T.; Mikami, A.; Yakoh, T.; Ogi, T. IoT Avatar: Turning Various Objects into Avatars. In *Proceedings of the Advances in Network-Based Information Systems*; Barolli, L., Ed.; Springer: Cham, Switzerland, 2024; pp. 398–405. [CrossRef]
9. VRM Consortium. 3D Humanoid Avatar File Format for VR|VRM. 2022. Available online: https://vrm.dev/en/ (accessed on 5 September 2026).
10. Schubert, T.; Friedmann, F.; Regenbrecht, H. The Experience of Presence: Factor Analytic Insights. *Presence Teleoperators Virtual Environ.* **2001**, *10*, 266–281. [CrossRef]
11. Biocca, F.; Harms, C.; Gregg, J.L. The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. In Proceedings of the International Workshop on Presence, Philadelphia, PA, USA, 21–23 May 2001.
12. Brooke, J. SUS—A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*; Taylor & Francis: Abingdon, UK , 1996; Volume 189, pp. 189–194.
13. Page, A.; Charteris, J.; Berman, J. Telepresence Robot Use for Children with Chronic Illness in Australian Schools: A Scoping Review and Thematic Analysis. *Int. J. Soc. Robot.* **2021**, *13*, 1281–1293. [CrossRef]
14. Ahumada-Newhart, V.; Olson, J.S. Going to School on a Robot: Robot and User Interface Design Features That Matter. *ACM Trans.-Comput.-Hum. Interact.* **2019**, *26*, 25. [CrossRef] [PubMed]
15. Slater, M. Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3549–3557. [CrossRef] [PubMed]
16. Kim, K.; Bolton, J.; Girouard, A.; Cooperstock, J.; Vertegaal, R. TeleHuman: Effects of 3d Perspective on Gaze and Pose Estimation with a Life-Size Cylindrical Telepresence Pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*; ACM: New York, NY, USA, 2012; pp. 2531–2540. [CrossRef]

17. Gunkel, S.N.; Dohmen, M.D.; Stokking, H.; Niamut, O. 360-Degree Photo-realistic VR Conferencing. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 946–947. [CrossRef]
18. Viola, I.; Jansen, J.; Subramanyam, S.; Reimat, I.; Cesar, P. VR2Gather: A Collaborative, Social Virtual Reality System for Adaptive, Multiparty Real-Time Communication. *IEEE MultiMed.* **2023**, *30*, 48–59. [CrossRef]
19. Sumigray, A.; Laidlaw, E.; Tompkin, J.; Tellex, S. Improving Remote Environment Visualization through 360 6DoF Multi-sensor Fusion for VR Telerobotics. In Proceedings of the Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, Boulder, CO, USA, 8–11 March 2021; pp. 387–391. [CrossRef]
20. Shakeri, H.; Yuan, Y.; Axtell, B.; Geiskkovitch, D.Y.; Neustaedter, C. Designing Smart Home Technology For Passive Co-Presence Over Distance. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS '24*; ACM: New York, NY, USA, 2024; pp. 3389–3406. [CrossRef]
21. Shakeri, H.; Neustaedter, C. Passive Co-presence: Exploring How Peripheral Devices Connect People Over Distance. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24*; ACM: New York, NY, USA, 2024; pp. 1–6. [CrossRef]
22. Sneed, J.C.; Deringer, S.A.; Hanley, A. Nature Connection and 360-Degree Video: An Exploratory Study With Immersive Technology. *J. Exp. Educ.* **2021**, *44*, 378–394. [CrossRef]
23. Hejtmánek, L.; Hůla, M.; Herrová, A.; Surový, P. Forest Digital Twin as a Relaxation Environment: A Pilot Study. *Front. Virtual Real.* **2022**, *3*, 1033708 . [CrossRef]
24. Kalantari, S.; Bill Xu, T.; Mostafavi, A.; Lee, A.; Barankevich, R.; Boot, W.R.; Czaja, S.J. Using a Nature-Based Virtual Reality Environment for Improving Mood States and Cognitive Engagement in Older Adults: A Mixed-Method Feasibility Study. *Innov. Aging* **2022**, *6*, igac015. [CrossRef] [PubMed]
25. Kalantari, S.; Xu, T.B.; Mostafavi, A.; Kim, B.; Dilanchian, A.; Lee, A.; Boot, W.R.; Czaja, S.J. Using Immersive Virtual Reality to Enhance Social Interaction Among Older Adults: A Cross-Site Investigation. *Innov. Aging* **2023**, *7*, igad031. [CrossRef] [PubMed]
26. Sacchelli, S.; Grilli, G.; Capecchi, I.; Bambi, L.; Barbierato, E.; Borghini, T. Neuroscience Application for the Analysis of Cultural Ecosystem Services Related to Stress Relief in Forest. *Forests* **2020**, *11*, 190. [CrossRef]
27. Cinnamon, J.; Jahiu, L. 360-Degree Video for Virtual Place-Based Research: A Review and Research Agenda. *Comput. Environ. Urban Syst.* **2023**, *106*, 102044. [CrossRef]
28. Tran, T.Q.; Langlotz, T.; Young, J.; Schubert, T.W.; Regenbrecht, H. Classifying Presence Scores: Insights and Analysis from Two Decades of the Igroup Presence Questionnaire (IPQ). *ACM Trans.-Comput.-Hum. Interact.* **2024**, *31*, *61*. [CrossRef]
29. Eriksson, E.; Elif Baykal, G.; Torgersson, O. Design-Based Research in Human–Computer Interaction: A Scoping Review. *Interact. Comput.* **2025**, *advance online publication*, iwaf030 . [CrossRef]
30. Lewis, J.R.; Sauro, J. Item Benchmarks for the System Usability Scale. *J. Usability Stud.* **2018**, *13*, 158–167.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.