Surrogate Avatar: Enhancing Situated Co-Presence and User Mobility in Symmetric Telepresence Conversations

SHENG-CIAN LEE, National Yang Ming Chiao Tung University, Taiwan
YI-LIEN CHANG, National Yang Ming Chiao Tung University, Taiwan
CHIU-HSUAN WANG, National Taiwan University, Taiwan
BING-YU CHEN, National Taiwan University, Taiwan
LIWEI CHAN, National Yang Ming Chiao Tung University, Taiwan

Fig. 1. Surrogate Avatar facilitates user mobility in symmetric telepresence conversations by strategically positioning avatars in socially and environmentally suitable locations within their local settings. This functionality is illustrated in a scenario involving User A and User B, depicted in the upper and lower row images respectively. User A, dressed in white, interacts with User B's avatar (B'). As User A moves within the environment—from (a) standing at a table to (b) sitting on a sofa, then (c) standing again, and finally (d) sitting at a desk—the avatar B' adjusts its position accordingly. Concurrently, User B, shown in the bottom images, dressed in black and seated at a desk, interacts with User A's avatar (A'), which remains seated opposite him for the duration of the conversation.

Authors' addresses: Sheng-Cian Lee, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, sclee.cs12@nycu.edu.tw; Yi-Lien Chang, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, alan881104.cs12@nycu.edu.tw; Chiu-Hsuan Wang, National Taiwan University, Taipei, Taiwan, chwang821014@ntu.edu.tw; Bing-Yu Chen, National Taiwan University, Taipei, Taiwan, robin@ntu.edu.tw; Liwei Chan, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, liweichan@nycu.edu.tw.

ACM 2573-0142/2025/9-ARTpn4492
https://doi.org/10.1145/3743716

We present Surrogate Avatar, an adaptive telepresence method that enhances user mobility and situated co-presence in symmetric avatar-mediated communication. The system enables a remote user's avatar to autonomously position itself in socially and environmentally appropriate locations within the local user's space—based on spatial affordances, interactional norms, and environmental constraints—supporting fluid interaction without requiring a shared environmental context. Through a formative study, we derived key adaptation objectives and implemented them using a distributed optimization framework based on the AUIT system. The framework distributes adaptation tasks across server and client to balance responsiveness and computational efficiency. A user study involving both stationary and nomadic scenarios demonstrated consistently high usability and presence, with some limitations observed under walking conditions. An additional exploratory field study in a semi-structured public setting demonstrated the system's viability beyond controlled lab conditions. These findings motivate future designs of mobile telepresence systems that dynamically adapt to spatial and conversational context while mitigating misunderstandings that can arise from asymmetric environmental awareness and supporting privacy-sensitive interaction.

CCS Concepts: • **Human-centered computing → Virtual reality**; *Interaction paradigms*; *User studies.*

Additional Key Words and Phrases: Mixed Reality Telepresence, Symmetric Telepresence, Nomadic Telepresence

ACM Reference Format:
Sheng-Cian Lee, Yi-Lien Chang, Chiu-Hsuan Wang, Bing-Yu Chen, and Liwei Chan. 2025. Surrogate Avatar: Enhancing Situated Co-Presence and User Mobility in Symmetric Telepresence Conversations. *Proc. ACM Hum.-Comput. Interact.* 9, 5, Article pn4492 (September 2025), 43 pages. https://doi.org/10.1145/3743716

1 INTRODUCTION

Telecommunication technology has continually enhanced long-distance communication. While telephones enabled audio conversations and mobile phones introduced mobility, recent advances in head-mounted displays (HMDs) now allow immersive 3D co-presence. This progression culminates in symmetric telepresence, where paired users simultaneously perceive each other as avatars situated within their respective environments [20, 43]. This stands in contrast to asymmetric telepresence, where one user establishes telepresence into the environment of another [37, 39], typically in scenarios where a remote expert provides guidance or assistance to a local worker.

In symmetric telepresence, basic co-presence can be achieved by placing the remote avatar at a fixed location or maintaining a constant spatial relationship to the user. To enable more immersive and socially coherent interaction, recent research has explored the concept of *situated co-presence*, in which avatars are not only colocated but also contextually embedded within the environment—such as by aligning with local furniture—to support attention, coordination, and task relevance [10, 20, 43]. However, this approach often relies on a shared context, established through predefined mappings between users' environments, such as matched furniture layouts or synchronized landmarks [10, 20, 33]. While shared context can facilitate communicative alignment, it also introduces rigid spatial constraints that limit mobility and confine interactions to predefined zones.

To address these limitations, we introduce the *Surrogate Avatar* system, which enables situated co-presence *without requiring pre-established shared context.* The system automatically positions the remote avatar within the local user's environment by continuously adapting to social and environmental cues—such as maintaining face-to-face orientation, aligning with eye level, and avoiding occlusions. Surrogate Avatar supports more flexible, authentic interactions and greater user mobility, albeit at the cost of forgoing environmental anchoring as a conversational aid.

This design decision reflects the nature of many remote interactions, which often begin without a shared spatial frame of reference. Instead, relevant situational cues typically emerge gradually as the conversation unfolds. For instance, an informal check-in may evolve into a goal-directed discussion, prompting a transition from a context-independent setup to one that leverages available environmental references [10, 20]. In some cases, effective cooperation may require transitioning into an asymmetric telepresence mode, where one participant is fully situated in the other's environment to establish a unified context [37, 39]. These examples highlight that shared context is not a static requirement but a dynamic, evolving aspect of communication. Telepresence systems should therefore support seamless transitions between context-independent and context-rich modes, allowing users to adjust the degree of environmental grounding as conversational demands shift. The Surrogate Avatar system supports this vision by serving as the default interaction mode for symmetric telepresence. It provides adaptive, context-aware avatar placement that does not depend on shared context and promotes spatial flexibility and user mobility.

Our approach draws on foundational theories of proxemics [12] and F-formations [25], which emphasize interpersonal distance and orientation as essential to socially coherent interaction. These principles have informed spatial behavior in telepresence, VR, and AR systems [13, 19, 34]. We extend their application to symmetric telepresence, with additional emphasis on environmental affordances—such as seating, obstacles, and line-of-sight—as factors influencing avatar positioning. To capture the nuanced expectations users hold in these dynamic spatial contexts, we conducted a formative study in which participants manually positioned remote avatars in varied settings and explained their rationale.

Insights from this study were synthesized into a set of objective positioning rules and implemented through a distributed adaptation framework inspired by the AUIT optimization algorithm. To address the computational limitations of HMDs and the dynamic demands of avatar control, our system delegates low-frequency, compute-intensive tasks (e.g., avatar positioning) to a remote server, while handling high-frequency, latency-sensitive tasks (e.g., avatar orientation) locally—ensuring responsive and lifelike avatar behavior.

We evaluated the Surrogate Avatar system in a controlled user study, where participant pairs engaged in conversations under both stationary and walking conditions. The goal was to assess whether the system could maintain adaptive avatar placement across changing social and environmental contexts. While the system generally supported coherent interaction, participants reported reduced responsiveness and spatial alignment during walking, revealing limitations in real-time adaptation. As the study lacked a baseline comparison, these findings offer preliminary evidence of the system's effectiveness. To further explore its practical applicability, we conducted an exploratory field study in a semi-structured public setting. This test surfaced additional challenges and revealed nuanced user behaviors, underscoring the need for robust avatar adaptation in less controlled, more dynamic settings.

This work makes three key contributions:

- The introduction of Surrogate Avatar, an adaptive method designed to enhance user mobility in symmetric telepresence by dynamically adapting avatars to the local environment, ensuring their situated co-presence.
- An implementation of a distribution adaptation framework that allocates periodic and frequent adaptation cycles across server and client HMDs to ensure real-time avatar adaptation and preserve avatar liveness.
- A controlled lab study evaluating participant experiences with the Surrogate Avatar under stationary and nomadic conditions, complemented by an exploratory field study in a semi-structured public setting to identify additional challenges and behavioral nuances in a less controlled, more dynamic setting.

2 RELATED WORK

This section reviews prior research relevant to our study, highlighting advancements and current trends in telepresence, adaptation frameworks, and nomadic technologies.

2.1 Symmetric Avatar Telepresence

Early AR telepresence systems aimed to transport remote users into local physical environments through real-time 3D capture and reconstruction. For example, these systems allowed users to appear on a local sofa or near a table, complete with realistic occlusion, as viewed through AR HMDs. However, such systems lacked the ability to support free movement after teleportation. This limitation was addressed by Holoportation [29], which enabled high-quality teleportation with free movement, ensuring smoother interaction by aligning virtual and physical spaces. Room2Room [30] extended this line of work by enabling life-size projection of remote participants directly into physical rooms through spatial augmented reality. Their system reconstructed remote users in real time and projected them onto calibrated surfaces to maintain conversational formations. However, Room2Room required dense instrumentation of rooms with projector-camera units and manual labeling of seating affordances, limiting its scalability to dynamic or mobile settings.

With the development of consumer-grade AR and VR HMDs, telepresence has become more accessible. Platforms like Meta's Horizon Workrooms, rather than focusing on detailed 3D reconstruction, employ animated avatars to represent remote users. While this approach avoids the need for expensive custom environments, it introduces new challenges. Variations in physical spaces mean that directly copying a user's movements to their avatar in a different environment can lead to unnatural interactions and confusion. To address the challenge of differing physical environments, one approach is to establish a common virtual space where geometric differences between users' respective environments are minimized. In this shared space, mapped within their individual environments, users can meet and interact. This common space may be as simple as a shared empty area or can be defined around specific objects such as matching chairs [30], desks [14, 16], or whiteboards [10, 11]. Kim et al. [20] aimed to maximize this common space by mapping shared object clusters across both environments. The benefit of this method is that, when the common space is well-defined and geometric differences are minimized, face-to-face communication can occur with minimal manipulation of avatars. However, such direct communication remains restricted to the defined common space. To extend interaction beyond this limited common space, alternative approaches focus on modifying avatar positions and poses to preserve the communicative intent, rather than attempting to directly map the physical environments. Yoon et al. [43] proposed a learning-based method to position avatars in dissimilar remote spaces. Wang et al. [41] introduced a "predict-and-drive" technique to create smoother transitions for teleported avatars. Choi et al. [3] developed methods for retargeting locomotion and object interaction in morphologically different spaces, while Kang et al. [17] addressed the transmission of deictic gestures in spaces with different layouts. These approaches typically assume some level of layout consistency across spaces to calculate avatar positions and maintain the integrity of their communication intentions. Despite the varying methods for handling dissimilar environments, previous works focus on avatar telepresence within rooms, where knowledge of room layouts is necessary to support effective communication.

In contrast, our work introduces Surrogate Avatar, which removes the constraints imposed by physical borders. While we also adjust avatar positions and poses, we achieve this without considering the remote partner's space layout. Previous methods that depend on spatial layouts aim to preserve user-object relationships, typically focusing on object-centric scenarios, such as discussions involving a whiteboard or simulating a situation where users share a virtually connected living space. Our approach, however, removes the need for object-based interactions and instead prioritizes facilitating communication, particularly for conversations that maximize mobility and are independent of the physical environment.

2.2 Normadic Telepresence

Nomadic telepresence refers to the ability of telepresence technology to operate without being restricted by specific locations. Previous research has primarily explored this concept within the context of remote assistance, where a remote user can virtually inhabit the environment of a local host. The local host in a nomadic telepresence setup can either be a mobile robot or a user. When the host is a robot, the remote user experiences presence at the local site through the robot's capture capabilities [23, 32, 42]. Conversely, if the host is a (novice) user, the remote (expert) user provides assistance at the local site, a setup known as remote assistance via asymmetric telepresence. Systems like Loki [39] often rely on location-dependent assistance, where the host remains fixed, and remote assistance depends on this fixed location. The main challenge in asymmetric nomadic telepresence is to remove the local host's location dependency while preserving high-fidelity remote participation. A single head-mounted camera restricts the remote participant's view to the host's perspective, limiting the sense of presence. To address this, a 360-degree panoramic camera can be used [18, 35, 44, 45], allowing the remote participant to explore a full 360-degree view of the site, although still dependent on the host's viewpoint. Incorporating live 3D scene reconstruction [1, 37, 40] further enables the remote participant to explore the site freely, independent of the host's viewpoint.

Research on nomadic telepresence within the context of symmetric telepresence is limited. Our work introduces a novel approach by concurrently adapting avatars both socially and environmentally to users' local environments during telepresence conversations. This initiative represents a significant step toward advancing symmetric telepresence towards true nomadicity.

2.3 Adaptive Interfaces in XR

Adaptive UI and avatar placement in XR are critical for creating immersive and user-friendly experiences. As XR environments become increasingly complex and dynamic, the need for responsive systems that can adapt to varying contexts and user interactions grows. Effective adaptation not only enhances usability but also ensures that interactions remain intuitive and seamless. For instance, Oliveira and Araujo [28] developed a context-aware AR system that dynamically adjusts its interface based on changing contexts by applying adaptation rules to select appropriate UI patterns. To enhance the usability of XR applications, creators need to consider factors such as real-world geometry [9, 27] information density and cognitive load [24], ergonomics [5], and layout consistency across environments [4]. Additionally, Belo et al. [6] introduced an optimization framework called AUIT, which enables creators to experiment with adaptation goals derived from custom usability factors for UI placements. While these previous studies primarily focus on the placement and adaptation of UIs, our research extends the consideration to the placement of avatar positions and poses, emphasizing adaptation goals related to environmental and social contexts with the aim of maintaining the quality of face-to-face avatar conversations.

The adaptation of avatar positions and poses has been explored in various applications. For example, redirected walking [38] allows VR users to navigate a virtual space larger than the physical space through real walking, and avatar adaptation has been utilized to facilitate symmetric avatar telepresence, as discussed in Section 2.1. These approaches [20] aim to define a transfer function that maps between two spaces to maximize a shared virtual space. When considering complex user-object interactions between spaces, this transfer function can be learned through machine learning techniques, using adaptation data provided by users [43]. In parallel, Li et al. [22] proposed an interactive AR storytelling system that adapts virtual character behaviors and object placements to match scene semantics and user actions. Although their focus is narrative delivery, their consideration of spatial and social coherence through real-time adaptation resonates with our objective of enabling socially appropriate and context-sensitive avatar placements.

This work explores an alternative symmetric telepresence system centered on mobility, with the primary objective of ensuring high-quality face-to-face avatar conversations. Through a body-storming interview, we identified key adaptation goals for avatar placement in various environments during conversations. These goals were implemented using the AUIT framework [6], which automatically suggests avatar placements to improve face-to-face interactions in dynamic settings.

3 FORMATIVE STUDY

We extend proxemics [12] and F-formation theory [25] to symmetric telepresence, emphasizing not only social spatial norms but also environmental affordances. To inform adaptive avatar placement, we conducted a formative study in which participants manually positioned remote avatars across various settings and explained their rationale. The findings revealed key spatial and contextual factors that shape natural, situated co-presence, guiding the design of our adaptation framework.

3.1 Apparatus

We developed a custom program to facilitate telepresence calls through augmented reality (AR) headsets, enabling an experimenter to join the participant's environment using avatar telepresence for the study. This system was developed using Unity and deployed on Meta Quest

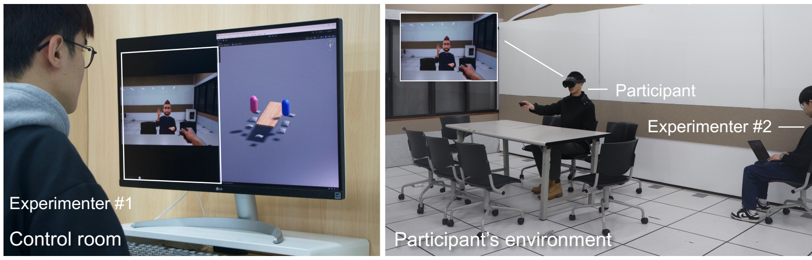


Fig. 2. In the formative study setup, Experimenter #1, located in the control room, joined the participant through an avatar representation in the participant's own environment. Concurrently, Experimenter #2 was physically present in the participant's environment to assist with the study.

Pro headsets, integrating Meta's Avatar System to capture realistic facial expressions and gestures. Networking was managed using Photon Unity Network, while voice communication was enabled through Photon Voice.

As depicted in Figure 2, participants equipped with AR headsets were able to visualize the experimenter's avatar within their environment. They could adjust the avatar's position and orientation using a controller. On the experimenter's side, a monitor interface displayed the layout of the participant's room and the positioning of avatars. This setup enabled the experimenter to observe and discuss the participant's choices regarding avatar placement.

3.2 Procedure

To enhance the diversity of environmental conditions, the study was carried out in the personal workspaces and living spaces of the participants, who voluntarily offered these settings for the research. The selection of workspaces was tailored to support various activities, including sitting at a desk, moving to a meeting table, and standing by a whiteboard. Similarly, the living spaces were chosen to accommodate actions such as sitting on a sofa, standing at the edge of the room, and walking around, thereby providing a range of interactive contexts for the study.

The study was facilitated by two experimenters: one operated remotely from the control room, acting as the remote user in the telepresence system, while the other provided on-site support to the participant. The on-site experimenter was responsible for setting up the experimental environment using the Meta Scene Capture tool[1], which involved labeling furniture and defining room geometry to ensure accurate interactions between the avatars and the environment. Additionally, this experimenter handled participant briefing, consent, demographic data collection, and technical support. During the telepresence sessions, the remote experimenter, represented by an avatar, guided the participants through various context switches that mimicked typical telepresence interactions, such as sitting on different pieces of furniture or standing in specific areas of the room. Participants were tasked with adjusting the avatar's position to fit each new context, and they discussed their placement choices with the experimenter to optimize the avatar's location within the virtual environment.

[1]https://developers.meta.com/horizon/documentation/unity/unity-scene-build-mixed-reality/

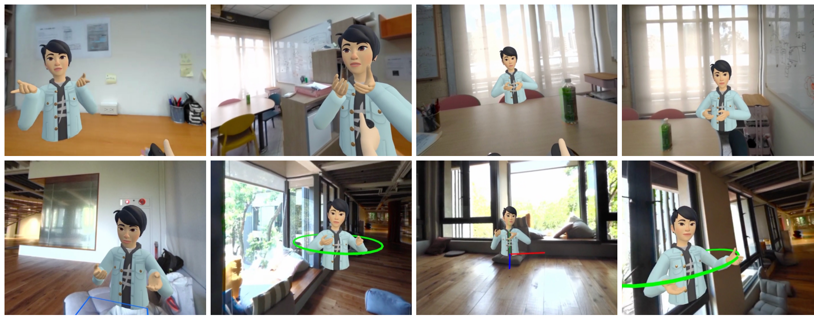


Fig. 3. Examples of avatar placement recorded from the participants' perspective include instances where the avatar was positioned at various points of interest within the environment, such as next to a work desk, beside a sitting area, or near an entryway to facilitate natural and contextually relevant interactions.

Following four context switches during which participants adjusted and discussed avatar placements, a semi-structured interview was conducted. This interview aimed to gather detailed insights from participants about their experiences with the avatar placement process. The entire session, including the context switches and interview, lasted approximately 1.5 hours.

3.3 Participants

We recruited 10 participants from a local university, ensuring an equal gender distribution (5 male, 5 female) with an average age of 23.7 years (SD = 2.3 years). Among these participants, two took part in both the work and living space scenarios, five participated exclusively in the living space scenario, and three only in the work space scenario. Nine participants had prior experience with AR or VR technology, and two had participated in AR/VR-based remote meetings. None of them were daily users.

3.4 Results and Discussion

Two authors analyzed the data, organizing it into 756 distinct segments and developing a framework to highlight key factors essential for maintaining co-presence between the participant and the avatar across different contexts. Examples of avatar placement recorded from the participants' perspective are shown in Figure 3.

3.4.1 Avatar Visibility (AV). Participants prioritized ensuring the visibility of the remote user's avatar, focusing on avoiding occlusion by environmental elements such as walls, furniture, and other barriers (N=10). All participants preferred to position the avatar directly in front of them or within their field of view. They also frequently adjusted the avatar's orientation to face them, enhancing interpersonal communication. As P4 noted, "In this way, I can see your facial expressions properly," underscoring the importance of visual cues in remote interactions. The feedback emphasized that clear and unobstructed visibility of the avatar made conversations feel more natural and engaging, whereas occluded or misaligned avatars diminished the sense of presence and responsiveness.

3.4.2 Task Visibility (TV). When navigating their environment, participants often repositioned the avatar near walls or corners to avoid obstructing their movement paths (N=10). During concurrent tasks, such as working on personal computers or reading, they preferred to place the avatar slightly to the side rather than directly ahead (N=5), allowing peripheral awareness without distracting from their primary activity. As P6 noted, "If you stay right in front, I can't concentrate on what I'm doing, but having you nearby feels reassuring." Participants emphasized that overly prominent avatar placements could interfere with task focus, highlighting the need for positioning strategies that maintain social presence while respecting users' attentional priorities.

3.4.3 Social Presence (SP). The sense of social presence, including co-presence and the feeling of being together, significantly influenced the placement of the remote collaborator's avatar. Many participants chose to position the avatar close and facing them to enhance the sensation of "communicating together" (N=7). When seated, participants often placed the avatar on a nearby chair or sofa to simulate the avatar "being seated" as well (N=10). Maintaining a similar eye-level between themselves and the avatar was also an important consideration; participants adjusted the avatar's height to avoid hierarchical impressions and to foster a sense of mutual respect. For instance, ten participants considered the remote user's feelings, with P4 stating, "If I place you (avatar) lower than me, it makes me feel that I'm not respecting you," highlighting how relative positioning and eye-level contribute to social presence during interactions.

3.4.4 Physical Affordance (PA). Participants tended to place the avatar where it matched the physical affordances of the environment, helping the avatar feel natural and integrated into the physical context. For example, avatars were often positioned on nearby chairs to prevent them from appearing to levitate, which participants found visually unnatural and socially awkward (N=9). Several emphasized that grounding the avatar—such as sitting on furniture or standing firmly on the floor—enhanced realism and supported the illusion of a co-located presence. As P9 remarked, "It (the floating avatar) makes me feel like it is just a model rather than a human," highlighting the negative impact of mismatched affordances. These findings suggest that leveraging environmental affordances is essential for sustaining the social believability and embodied realism of remote avatars.

3.4.5 Continuity (C). As participants moved through various contexts during the study, they considered the continuity of avatar placement during successive context switches (N=7). For instance, when P8 began to walk around, they positioned the avatar on the left, consistent with its previous placement on a chair to the left, avoiding unnecessary adjustments and ensuring smoother transitions.

3.4.6 Alignment with Spatial Interaction Theories. Notably, these adaptation goals align with established theories of spatial behavior. Specifically, the objective to maintain appropriate distances reflects Edward Hall's proxemic zones [12], ensuring avatars occupy socially comfortable spaces that facilitate interaction. Likewise, the emphasis on sustaining direct facing and flexible reconfiguration resonates with Marshall et. al.'s F-formation structures [25], which describe how participants cooperatively manage spatial arrangements to maintain shared conversational spaces. By embedding these spatial theories into our adaptation framework, we aim to support socially coherent and naturalistic co-presence even in dynamic and nomadic interaction scenarios.

3.5 Summary

Building on participants' placement behaviors identified in the formative study, we operationalized these insights into a structured set of adaptation objectives and embedded them within a multi-objective optimization framework based on the AUIT system, as detailed in Section 5. This framework enables the surrogate avatar to dynamically balance competing spatial and social demands in real time.

Specifically, participants' emphasis on avatar visibility (AV) informed the User Field of View and Environmental Occlusion objectives, ensuring that avatars remain clearly visible and unobstructed within the environment. Social presence (SP), reflected in preferences for comfortable distance, mutual eye-level, and facing direction, guided the Social Distance, Maintain Eye-Level, and Head-toward-User objectives. The need for environmental grounding (PA) motivated the Sitting Affordance and Avoid Levitation objectives, ensuring that avatars integrate naturally into the physical context without visual inconsistencies such as floating or improper seating. To support fluid transitions during user mobility, the Continuity (C) theme informed the Spatial Consistency objective, promoting stable and predictable avatar repositioning. In addition to these user-driven objectives, we incorporated the Body-toward-Head objective to support posture ergonomics and preserve avatar naturalness during dynamic interactions. While Task Visibility (TV) was highlighted by participants as a potential concern, addressing it would require real-time task recognition and attentional modeling, introducing a layer of complexity beyond the scope of the current system. Therefore, task-aware adaptation is deferred to future work.

4 SURROGATE AVATAR IN SYMMETRIC TELEPRESENCE

Our system leverages an adaptation process derived from the AUIT [6] toolkit, originally developed to facilitate the creation of adaptive user interfaces for extended reality applications. The conventional AUIT framework is primarily oriented towards the placement of static user interfaces, necessitating adaptation primarily in response to changes in the user's environment or shifts in focus within the current setting. We aim to extend the AUIT framework to enable dynamic and persistent adaptation of avatars that accompany the user as they navigate within the environment. We selected the AUIT framework because of its high flexibility for multi-objective optimization. Unlike prior frameworks that often rely on fixed sets of adaptation rules tailored to specific environments (e.g., predefined chair mappings or fixed collaboration zones) [2, 7, 21], AUIT allows designers to flexibly add, adjust, or reweight multiple adaptation objectives, facilitating scalability across diverse conversational contexts. This flexibility is crucial for supporting situated and mobile telepresence, where environmental layouts and social expectations dynamically change.

4.1 Design Challenges

Applying the AUIT framework to avatar adaptation introduces three distinct challenges:

(1) Objectives Tailoring for Avatar Telepresence Adaptation: Our formative research has pinpointed specific challenges in maintaining appropriate social and environmental contexts during avatar telepresence communications. These insights have been incorporated into the objectives for avatar adaptation, introduced later. This customized set of objectives is designed to address

the unique demands of avatar telepresence, ensuring that the avatar behaves in ways that are contextually appropriate and socially engaging during tele-communication.

(2) Complex Spatial Behaviors and Dynamic Nature of Avatars: Unlike static user interfaces, avatars require continuous adjustments not only in their location but also in their body and head orientations. This dynamic nature significantly increases the complexity of the adaptation parameters, posing particular challenges for implementation on HMDs. Moreover, while user interface elements can remain unchanged until triggered by a new adaptation condition, avatars demand frequent updates to preserve a sense of liveliness, often necessitating high frame rates (e.g., 90 fps). These requirements complicate the straightforward application of the AUIT framework without modifications.

(3) Adaptation for Stationary and Walking States: In contrast to the AUIT's focus on stationary settings, our system ensures that the avatar remains continuously beside the user during movements through the local environment. To prevent the avatar from erratically shifting locations along the user's walking path, we implement specific adaptations. For stationary users, the avatar's positioning relies on the world coordinate system. Conversely, for walking users, we adjust the avatar's alignment to the user's coordinates, ensuring smooth movement and maintaining a stable, consistent companion.

To address the computational demands of real-time avatar adaptation—especially under mobility constraints—we implemented a distributed adaptation mechanism. While offloading computationally intensive tasks to the server is a common practice in interactive systems, this setup was essential to balance responsiveness and complexity in our design. In this setup, parameters that necessitate less frequent updates and involve more intricate computations, such as the avatar's position needing broad search within the environment around the user while considering diverse environmental contexts are processed on a remote server. This approach accommodates the broad search space and diverse environmental contexts. In contrast, parameters requiring more frequent updates, such as body and head orientations, are handled locally on the client HMDs due to their simpler, one-dimensional calculations and less demanding objectives. Additionally, our framework adjusts to both stationary and walking states of the user, ensuring smooth avatar movement and a consistent presence by aligning with the user's coordinates during movement. In the following sections, we first introduce the distributed adaptation framework and then detail the implementations that address each of the three challenges.

4.2 The distributed adaptation framework

Figure 4 illustrates the network architecture of the distributed adaptation framework. When a client pair (Client A and Client B) initiates a telepresence call via the server, a series of initial setup processes are triggered. The server creates an adaptation thread encompassing an AUIT component for each client for calculating the position of the remote avatar (e.g., Avatar B) in the client's (e.g., Client A) local environment. The component requires the client's registration of context objects capturing both environmental (e.g., spatial information about furniture) and social contexts (e.g., the position of Client A) in their local environment. Concurrently, each client operates a local AUIT component within their HMDs, which calculates the orientation of the remote avatar, relying solely on social context data (e.g., the position of Client A).

Throughout the session, both clients continuously update their respective server-side AUIT components with their current positions within the local environment. They also stream voice

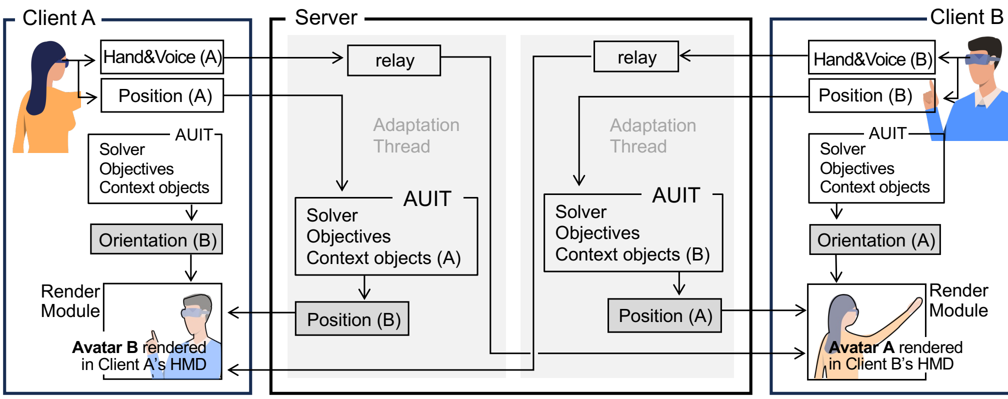


Fig. 4. The network architecture of the distributed adaptation framework.

and avatar postures from their HMDs to the server, which forwards this data to the remote client. This enables the animation of hand motions and voice-activated facial expressions of the avatar.

The server-side AUIT component frequently recalculates and updates the appropriate positions for the remote avatar based on the context data of the associated client and sends these updates back to the client to adjust the avatar's position locally. Simultaneously, the client-side AUIT component updates the orientations of the avatar. Put together, the render module in the client HMD manages the rendering of the remote avatar through three distinct update cycles:

- Continuous Updates: The remote avatar's hand motions and facial expressions are updated continuously. This information is relayed from the remote client through the server.
- Periodic Updates: The position of the avatar is updated every 200 milliseconds. These updates are determined by the server-side AUIT component.
- Frequent Updates: The orientations of the avatar are updated every 30 milliseconds, based on calculations performed by the AUIT component running on the client's HMD.

4.3 The AUIT component

The specifications of the AUIT components are distinct on the server and client sides. In the following, we introduce the overall goal of each modules in the AUIT component, followed with how they are set differently for the server and the client. For the pseudocode of the server, the client, and their respective objectives, please see the appendix.

4.3.1 The solver and evaluator. The solver's function is to explore the parameter space to identify the optimal parameter values that result in the lowest cost, as defined by our objectives.

For illustration, consider the server-side configuration and its corresponding pseudocode, where the goal of the adaptation is to determine the optimal avatar position. In this scenario, the solver's parameter space encompasses the three-dimensional space of the environment. The initial position of the avatar, denoted as current_pos, is set in front of the user if the avatar has not yet been assigned a position. Once a parameter setting is in place, the evaluator calculates the cost by computing a weighted sum of costs across all objectives, expressed as e(current_pos). The update process begins, which involves repeatedly verifying whether the quality of current_pos meets the specified suitability threshold. If the cost function, denoted as e(current_pos), surpasses this threshold, it signifies that current_pos is suboptimal and requires adaptation. The objective is to adjust to a new position where the cost is below the threshold.

Adaptation involves tweaking the objectives. The process selects one objective at random and adjusts current_pos towards a new position, new_pos, using the heuristic associated with that objective. If e(new_pos) is found to be lower than e(current_pos), then current_pos is updated to new_pos. If not, new_pos is discarded as it does not offer a cost reduction. This objective tweaking is repeated up to 1500 times or until a current_pos is identified that meets the suitability threshold, allowing the adaptation process to conclude early. If a suitable current_pos is found, it is used to update the rendering of the avatar's position. If not, the current_pos remains unchanged, and the system proceeds to the next update cycle.

We established the limit of 1,500 iterations to strike a balance between achieving high-quality adaptations and maintaining reasonable computation cost. This threshold ensures that the solver performs sufficiently thorough searches within the parameter space while preventing excessive processing that could delay system responses and impact user experience. In the client-side configuration, the adaptation process focuses on determining optimal orientations for the avatar's head and body. Here, the solver operates within a two-dimensional parameter space that includes the directions the avatar's head and body are facing. To balance effectiveness and efficiency, the number of iterations for objective tweaking is set at 750.

4.3.2 The objectives and context objects. The objectives for avatar adaptation differ between the server and client sides, tailored to the specific requirements of avatar positioning and orientation. On the server side, objectives focus on environmental interaction and social appropriateness. The objectives include:

- Environmental Occlusion: Ensuring the avatar is not obscured by objects in the environment.
- Social Distance: Maintaining an appropriate distance from other avatars or users to reflect social norms.
- User Field of View: Positioning the avatar within the user's visible area to ensure visibility.
- Sitting Affordance: Placing the avatar in a position that suggests it can sit if the scenario requires.
- Avoid Levitation: Ensuring the avatar remains grounded and does not appear to float.
- Maintain Eye-Level: Aligning the avatar's eyes with the eye level of the user to facilitate natural interaction.
- Spatial Consistency: Ensuring the avatar's position remains stable without abrupt changes from its previous location.

On the client side, the objectives are designed to enhance interpersonal engagement during interactions with the user and include:

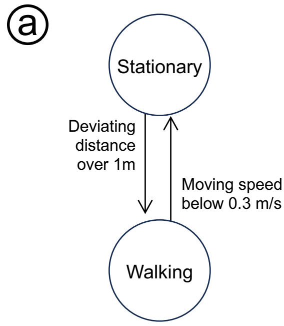
- Head-toward-User: Orienting the avatar's head towards the user to simulate engagement and attentiveness.
- Body-toward-Head: Orienting the avatar's body in the direction of the avatar's head to maintain a natural avatar posture

	Objective	user avatar	Heuristic
Server side	**Environmental Occlusion**: Prevent avatar merging with the environment		Relocate the avatar away from obstructive geometries.
	Social Distance: Position the avatar at a comfortable social distance.		Adjust the avatar's position to within a predefined social distance.
	User Field of View: Keep the avatar within a ±45° angle of the user's central field of view.		Move the avatar within the 90° field of view interval.
	Sitting Affordance: Ensure the avatar sits on a nearby chair.		Position the avatar on the nearest chair or a randomly selected chair within the environment.
	Avoid Levitation: Ensure the avatar stays grounded		Place the avatar on the closest surface, adding a random offset to accommodate large surface areas.
	Eye-Level: Align the avatar's eye level with the user.		Adjust the avatar's height to match the user's eye level.
	Spatial Consistency: Ensure the avatar's position without abrupt changes from its previous location	prior position	Relocate the avatar toward the previous location
Client side	**Head-toward--User**: Orient the avatar's head towards the user		Rotate the avatar's head towards the user.
	Body-toward-Head: Orient the avatar's body toward the direction of the avatar head		Rotate the avatar's body toward aligning the avatar's head.

Fig. 5. The objectives and the associated heuristic that dictate the adaptation algorithm implemented respectively on the server and client sides.

The objectives for avatar adaptation, along with their associated optimization heuristics, are illustrated in Figure 5. The pseudocode detailing the implementation of these objectives is provided in Appendix. These objectives operationalize what we define as "socially and environmentally suitable" locations. Social suitability is reflected in objectives such as Social Distance, Eye-Level, and Head-toward-User, which aim to preserve norms of proxemic spacing, mutual orientation, and interpersonal attentiveness. Environmental suitability is ensured by objectives like Sitting Affordance, Avoid Levitation, and Environmental Occlusion, which anchor avatars to appropriate surfaces, prevent unnatural floating, and avoid visual obstruction. These heuristics are grounded in spatial interaction theories and informed by our formative study findings, ensuring that avatar placement aligns with both social expectations and environmental constraints.

Context objects, which include environmental and social contexts, are critical for calculating the costs associated with various adaptation objectives in our system. The environmental context details the spatial positions and characteristics of objects like chairs and tables, essential for objectives such as Environmental Occlusion, Avoid Levitation, and Sitting Affordance. On the other hand, the social context involves tracking data from the client, specifically the position and orientations of the head. This information is essential for objectives focused on Social Distance, Field-of-View, Eye-Level, Head-toward-User, and Body-toward-Head, ensuring accurate social interaction of the avatar within the user's personal space. The objective of Spatial Consistency does not utilize any context object.



ⓑ

Objcetives	Stationary	Walking
Environmental Occlusion	0.12	0.15
Social Distance	0.3	0.2
User Field of View	0.24	0.25
Sitting Affordance	0.12	0
Levitation	0.08	0
Eye-Level	0.12	0.3
Spatial Consistency	0.02	0.1
Face User	0.5	0.5
Head Direction	0.5	0.5

Fig. 6. (a) The state diagram for detecting stationary and walking activities. (b) The weights applied in adaptation respectively for stationary and walking activities.

4.4 Adaptation for Stationary and Walking States

Our system differentiates between these stationary and walking states and creates corresponding adaptive behaviors. When detecting users are in walking state, the adaptation shifts from the world coordinate to the user's coordinate and the objective weightings tailored to the walking state are employed.

4.4.1 Detection of stationary and walking. To adapt avatar behavior based on user movement, we employ a straightforward state machine, depicted in Figure 6a, which determines whether the user is in a stationary or walking state. In the Stationary state, the system continuously monitors the distance between the user and a reference point—defined as the user's position upon entering the Stationary state. If the user moves more than 1 meter away from this point, it signals significant movement, prompting the system to update the state to Walking. Conversely, in the Walking state, the system evaluates the user's speed every second. If the speed falls below 0.3 m/s, indicating the user has stopped or significantly slowed, the state reverts to Stationary, and the system establishes the current location as the new reference point for this state.

4.4.2 State-specific objective weighting. Each state has specific considerations, necessitating tailored avatar behaviors. To address this, we implement state-specific objective weighting to customize the avatar's behavior for each state. Detailed weightings are provided in Figure 6b. For example, in the Walking state, the weight for the Sitting Affordance Objective is reduced to zero, as sitting does not occur during movement. This released weight is then redistributed to objectives that are more relevant to the walking state, promoting a balanced adaptation. The configurations for these weights were developed through extensive testing and evaluations to ensure they accurately reflect real-world interaction dynamics. Additionally, objectives utilize different cost function parameters depending on the user's state. For example, in the User Field of View objective, placing the avatar in front of the user has a lower cost value in Stationary mode, aligning with natural interaction. Conversely, in Walking mode, positioning the avatar directly in front incurs a higher cost value to discourage placements that could obstruct the user's forward path.

4.4.3 Stationary and Walking: World vs. User Coordinate Systems. In the Walking state, the avatar is anchored to the user's coordinates, maintaining a consistent spatial relationship with the user and moving in tandem. As outlined in Figure 5, the objectives during the Walking state focus on avoiding environmental occlusion and maintaining social engagement. This ensures that as the avatar moves in alignment with the user, it remains visible and interacts appropriately within the environment.

4.5 Implementation

We developed a symmetric telepresence system enhanced with surrogate avatars using Unity 2022.3.11LTS. The server was run on the Windows PC (Intel Core I7 11800H, 32GB RAM,NVIDIA Geforce RTX 3070 Laptop), incorporating clients on Meta Quest Pro headsets for both users. This setup leveraged the headsets' pass-through functionality to merge virtual and physical environments effectively. To represent the remote user's avatar, we utilized the Meta Avatar system along with its inverse kinematics capabilities for accurate motion rendering.

For networking and data exchange, we implemented the Unity Transport Package, utilizing UDP sockets to facilitate communication between the server and client. We incorporated the Photon PUN 2 framework to manage multi-user networking and posture data transmission efficiently. Additionally, Photon Voice was integrated to enable real-time voice communication among users. The environment setup was captured and integrated using the Meta room setup, enhancing the overall telepresence experience by aligning the virtual and physical spaces.

5 USER STUDY

This study aims to evaluate if our adaptation facilitates effective telepresence conversations across different contexts. Using a within-subject design with mobility as the central variable, we compared user experiences in two scenarios: stationary and nomadic. In the stationary scenario, participants stayed seated at a desk within an office. In the nomadic scenario, they moved between sitting, standing, and walking in a room designed for movement. We assessed system usability, spatial, and social co-presence among participants. *Our primary hypothesis is that if the adaptation appropriately positions the avatar, there will be no significant differences in measurements between the two scenarios.*

5.1 Apparatus

The symmetric telepresence system described in Section 4 was employed. It included two Meta Quest Pro headsets, each connected to a Server PC via Wi-Fi 5G 6e. To simulate conversations under different environmental conditions, we established two distinct room setups, as shown in Figure 7. The "Nomad" room, a spacious 7-meter by 7-meter area, was designed to represent users in a nomadic mode, offering ample space for movement and interaction. Conversely, the smaller "Office" room, measuring 2 meters by 3 meters, simulated users in a stationary mode, mimicking a more confined, typical office environment.

5.2 Task

The study was structured to simulate conversations between participant pairs under various mobility conditions, combining a discussion task with a locomotion task.

Fig. 7. In the User Study, on the left: the office room, measuring 2 meters by 3 meters, represents the stationary scenario. On the right: the 7-meter by 7-meter nomad room embodies the nomadic scenario.

Discussion Task: This involved a two-phase approach based on a modified desert problem. In the preparation phase, participants individually considered a ranking question presented on paper, forming their rankings and rationales within two minutes without discussing with their partner. The collaborative discussion phase followed, where participants shared their thoughts, exchanged ideas, and aimed to agree on a unified ranking. Locomotion task conducted during the discussion task.

Locomotion Task: Tailored to the specific conditions used in the study. In the stationary condition, participants in the Office room were static, remaining seated throughout the discussions. In contrast, in the nomadic condition, the Nomad room provided a more dynamic environment where participants were required to be mobile. To ensure experimental control and consistency across participants, we deliberately employed a structured locomotion pattern in the Nomad room. This design allowed all participants to experience comparable environmental transitions and mobility states, enabling systematic evaluation of the system's adaptation performance under controlled conditions. Specifically, they followed a designated path that included sitting, standing, and walking, as detailed in Figure 7. They engaged with features of the environment—such as sitting on sofas, standing at tables, and leaning against walls—at four designated areas. The task consisted of eight activities in total—four sitting and four standing—distributed across these areas and connected by seven walking transitions. Participants were instructed to spend at least two minutes at each activity location and to walk slowly during transitions.

5.3 Procedure

Two study protocols were implemented to ensure avatar anonymity and reliable posture tracking in our study. (1) *Avatar Anonymity:* To prevent biases associated with avatar representation and the labor-intensive process of customizing individual avatars, we implemented a protocol for avatar anonymity. Participants were unable to meet in person or see their own avatars, allowing the use of just four preselected avatars (two male and two female) from Meta's system, based on the gender pairing of participants. This approach maintained the authenticity of the study while simplifying the avatar setup process. (2) *Reliable Avatar Posture Tracking:* Our system uses Meta's Avatar Inverse Kinematics (IK) to derive body posture from head and hand tracking data captured by HMDs. However, posture estimation becomes unreliable when hands move out of the HMD's tracking range, such as when resting on the lap. To counter this issue, participants were instructed to keep their hands visible to the HMDs throughout the telepresence conversation, ensuring accurate tracking and posture representation.

Participants were recruited in pairs from individuals who were not previously acquainted and were kept from meeting each other throughout the study to preserve anonymity. Each pair was briefed separately in different rooms and introduced to the symmetric telepresence system, followed by a practice session to familiarize themselves with the system's features. Following this briefing, participants were assigned to either the Nomad or Office room based on their study condition. In their assigned rooms, they first underwent an embodiment process to establish a sense of connection with their avatars. Their real-time hand movements, captured by the headsets' hand-tracking system, were directly mapped to the avatar's limb movements, allowing their physical hands to fully overlap with the avatar's hands in the visual field. This calibration step was intended to help participants develop a sense of embodiment before proceeding to the main tasks. Subsequently, participants were guided through discussion and locomotion tasks. Participants in the Nomad room were additionally given time to navigate and familiarize themselves with the environment before beginning the assigned study condition.

Following the tasks, participants completed a questionnaire that included a 7-point Likert scale. This scale evaluated the System Usability Scale (SUS) and aspects of Social and Spatial Presence from the Temple Presence Inventory (TPI). We adapted the original SUS to a 7-point format to improve scale sensitivity and reduce response interpolation, following prior recommendations [8] suggesting that 7-point Likert scales provide more accurate and reliable subjective measures than the original 5-point format. Reported SUS scores reflect the mean item ratings without rescaling to the traditional 0–100 SUS format. Additionally, participants responded to five custom questions designed to assess different facets of the participants' experiences during telepresence conversations. In the nomadic condition, participants provided ratings for each of the custom questions while in three different states: sitting, standing, and walking, respectively. Conversely, in the stationary condition, which only involved sitting activities, participants were asked to rate the custom questions solely for the sitting state. These custom questions assessed (1) participants' ability to maintain the verbal aspects of the conversation, (2) the smoothness of their conversational experience, (3) the partner's awareness of their environment, and whether (4) the spatial and (5) temporal aspects of the adaptation met their expectations. Displayed in Figure 9, these questions evaluate overall conversation quality (Q1 and Q2), the effectiveness of avatar positioning in conveying environmental awareness (Q3), and the accuracy and timeliness of spatial and temporal adaptations (Q4 and Q5). Please refer to the Appendix for the complete questionnaire.

For the second condition, participants switched rooms without encountering each other to maintain anonymity, completed the tasks, and filled out the same questionnaire. After both conditions, the participants met face-to-face for the first time to discuss their responses and experiences in an interview. We encouraged participants to articulate their ratings and overall impressions of the system, and to identify potential applications for the telepresence system. Additionally, we asked them to pinpoint areas in real life where miscommunications could arise due to avatar adaptation. The total study duration was 90 minutes per participant.

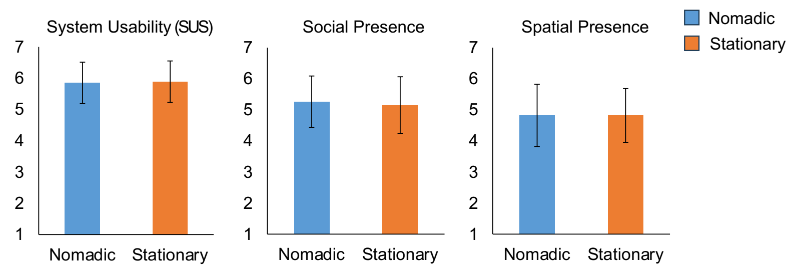


Fig. 8. User ratings for stationary and nomadic conditions on system usability (SUS), social presence, and spatial presence. Error bars indicate standard deviations.

5.4 Participants

Eighteen participants were recruited through public online college forums (e.g., Facebook) and physical campus advertisements for the study, comprising 5 males and 13 females, with an average age of 22.9 years ($SD = 3.36$). Participants were university students from academic backgrounds, including engineering, design, and communication studies. All participants reported prior experience with VR gaming headsets (e.g., Meta Quest, HTC Vive) or video passthrough AR features (e.g., Meta Quest Pro). Regarding usage frequency, two participants reported using VR/AR headsets approximately five times per week, while the remaining participants reported using them once per week, typically for gaming or casual exploration. Despite this experience, none of the participants identified themselves as heavy or daily users of VR/AR devices.

5.5 Results

The results were organized into three subsections: usability using SUS, social and spatial presence using TPI, custom questions, and qualitative insights.

5.5.1 System Usability. SUS scores were consistently high for both conditions (Stationary: $M = 5.89$, $SD = 0.66$; Nomadic: $M = 5.85$, $SD = 0.66$), indicating high usability. A Wilcoxon Signed-Rank Test found no significant difference in SUS scores between the conditions ($p > 0.05$). The results suggest that the participant pair was able to effectively use the system in both the static and nomadic scenarios.

5.5.2 Social and Spatial Presence. Social presence scores were high across both conditions (Stationary: $M = 5.15$, $SD = 0.91$; Nomadic: $M = 5.26$, $SD = 0.83$), with a Wilcoxon Signed-Rank Test showing no significant difference between them ($p > 0.05$), indicating a consistent and strong sense of social presence under both static and nomadic conditions. Spatial presence scores were above the midpoint for both stationary and nomadic conditions (Stationary: $M = 4.82$, $SD = 0.87$; Nomadic: $M = 4.82$, $SD = 1.00$), with no significant differences found between the two.

5.5.3 Custom Questions. Figure 9 displays the average ratings for the custom questions, comparing nomadic users in sitting, standing, and walking states with stationary users in the sitting state. A Friedman Test was conducted for each question, consistently showing significant differences across conditions. Notably, the Nomadic-Walking condition demonstrated lower ratings

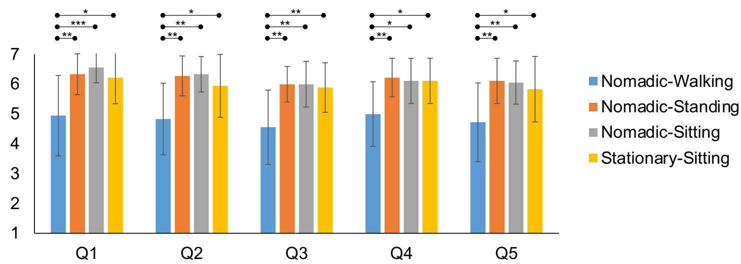


Fig. 9. User ratings for stationary and nomadic conditions on the five custom questions. Error bars indicate standard deviations. The one-asterisk(*), two-asterisk(**), and three-asterisk(***) symbols indicate p<0.05, p<0.01, p<0.001 significant differences, respectively.

compared to the non-walking conditions: Nomadic-Standing, Nomadic-Sitting, and Stationary-Sitting.

Regarding participants' perception of avatar awareness of their local environment (Q3), participants rated Nomadic-Walking significantly lower compared to other conditions (Nomadic-Walking: $M = 4.56$, $SD = 1.25$; Nomadic-Standing: $M = 6.0$, $SD = 0.59$, $Z = 3.68$, $p = 0.001$, $r = 1.23$; Nomadic-Sitting: $M = 6.0$, $SD = 0.77$, $Z = 3.61$, $p = 0.002$, $r = 1.20$; Stationary-Sitting: $M = 5.89$, $SD = 0.83$, $Z = -3.29$, $p = 0.006$, $r = 1.10$). This indicates that avatar positioning adaptation was less effective in the walking condition.

Participants also evaluated the spatial (Q4) and temporal (Q5) aspects of adaptation. Significant differences were noted, with Nomadic-Walking consistently rated lower than the other conditions. For the spatial aspect, mean scores were: Nomadic-Walking at 5.00 ($SD = 1.08$), Nomadic-Standing at 6.22 ($SD = 0.65$, $Z = 3.49$, $p = 0.003$, $r = 1.16$), Nomadic-Sitting at 6.11 ($SD = 0.76$, $Z = 2.84$, $p = 0.027$, $r = 0.95$) and Stationary-Sitting at 6.11 ($SD = 0.76$, $Z = -2.71$, $p = 0.04$, $r = 0.90$). For the temporal aspect, scores were: Nomadic-Walking at 4.72 ($SD = 1.32$), Nomadic-Standing at 6.11 ($SD = 0.76$, $Z = 3.55$, $p = 0.002$, $r = 1.18$), Nomadic-Sitting at 6.06 ($SD = 0.73$, $Z = 3.29$, $p = 0.006$, $r = 1.10$), and Stationary-Sitting at 5.83 ($SD = 1.10$, $Z = -2.71$, $p = 0.04$, $r = 0.90$). These findings indicate that participants found the adaptation to be less spatially and temporally congruent in walking compared to sitting and standing scenarios.

5.5.4 Qualitative Feedback. While our quantitative findings showed no statistically significant differences in usability, spatial presence, or social presence across conditions—consistent with our expectation that adaptation would support comparable user experiences between conversation partners across different locomotion scenarios—the qualitative feedback from participants revealed rich and nuanced perspectives on their interactions.

For SUS, most participants agreed that the system facilitated smooth telepresence communication across diverse scenarios. One participant highlighted its usability, noting, "The system made it easy to maintain face-to-face communication even while navigating the environment."

Regarding Social Presence, participants remarked, "Both modes provided sufficiently smooth experiences, and overall, the experiences in both modes were nearly identical." Additionally, a strong sense of co-presence was reported by most participants (N = 15). One participant commented, “During the conversation, their presence felt so strong that I perceived them as a real person and interacted with them accordingly,” while another added, “I will look into the avatar’s eye, even though I receive no response.”

Although Spatial Presence scores were above the midpoint, some participants criticized the avatar’s linear movement trajectory, describing it as mechanical. One participant noted, “Because it always moves in a straight line, it’s harder for me to perceive it as a real person moving through space.” Despite this, most participants agreed that the system’s adaptation positively influenced their discussions with the partner. As one participant stated, “It appeared in an appropriate position, and I could naturally engage in a conversation with it.”

Custom questions from the study highlighted participant preferences for sitting and standing scenarios over walking. In both nomadic and stationary conditions, participants praised the avatar’s effective use of environmental affordances. One participant remarked, “When I am sitting, the avatar naturally occupies the opposite chair, creating a realistic face-to-face discussion setting.” Another noted, “When I was standing against a standing table, the avatar positioned itself slightly to the side, which felt natural and avoided direct confrontation.”

However, the walking condition presented several challenges and received lower ratings. Participants reported issues such as delayed avatar response, visual occlusion during navigation, and inappropriate avatar movements through obstacles, all of which negatively impacted their experience. Specific comments from participants included: "The avatar took a bit longer to react when I started walking, which broke the flow of interaction," "While I was walking, the avatar appeared in my view, obscuring my navigation to the chairs," and "It was jarring to see the avatar pass through a table as I approached to sit down, which diminished the realism of the interaction." These issues highlight the need for enhancements in avatar behavior during walking to better accommodate mobility and adaptability. We further elaborate on these enhancements in the discussion section.

The above contrast indicates that, although overall ratings remained stable, participants still encountered meaningful challenges and expressed distinct preferences, particularly in nomadic scenarios. These qualitative insights provide important guidance for refining avatar behavior, transition strategies, and system responsiveness.

6 DISCUSSION

Our findings support the hypothesis that adaptive avatar positioning can maintain consistent user experiences across different mobility scenarios, as reflected in stable usability and presence scores. However, further analysis of user feedback revealed that adaptation effectiveness varied by context—particularly during walking, where participants reported issues with responsiveness and realism. Drawing on post-study interviews, we examine how factors such as avatar placement, mobility, and privacy concerns influence user experience and suggest directions for improving future symmetric telepresence systems.

6.1 Catering to Coupling Dynamics in Adaptation

In real-world one-on-one conversations, participants often shift their attention between their conversational partner and external tasks or objects. Our experimental design simulates this dynamic by incorporating two distinct phases: preparation and collaborative discussion. These phases correspond to different tight coupling styles, as classified in [26], which involves "Sharing the same information but using different physical displays," and DISC, defined as "Active discussion between two or more persons about the data or task."

Our findings indicate that preferences for avatar placement significantly vary based on the coupling style. During the preparation phase, which adheres to the SIDD style, participants preferred the avatar to be positioned outside their direct line of sight. One participant explained, "When I'm reading a document, having the avatar in front of me sometimes distracts me from reviewing the document. Ideally, when I am focused on my own tasks, I would prefer the avatar to keep a bit more distance." In contrast, during the DISC phase, which involves more collaborative discussions, participants were comfortable with the avatar placed directly within their field of view.

This variation underscores the importance of adaptive strategies that respond to the level of collaboration required. For instance, in loosely-coupled telepresence scenarios where individuals primarily want to coexist while engaging in occasional conversation, the adaptation can be minimized. Future research could enhance this approach by incorporating the concept of designated regions within the environment, such as living rooms or kitchens. In tightly-coupled conversations, the avatar would maintain close proximity to the user. For moderately-coupled interactions, the avatar would reposition only when the user moves to a different designated region. Conversely, in loosely-coupled scenarios, the avatar would remain static as long as the user remains within their environment. The coupling condition can be determined by analyzing the ongoing conversation or by allowing users to manually adjust the desired level of adaptation.

6.2 Catering to Mobility in Adaptation

While the system facilitates mobility without geographical constraints, subjective feedback indicated decreased satisfaction with its performance during walking compared to stationary use. This decline in satisfaction can be attributed to challenges encountered at different stages of the walking process.

At the Start of the Walk: Initially, the system detects the transition from stationary to walking based on the user's movement away from a reference point. This detection often results in a delayed activation of the walking mode adaptation, where the avatar's behavior adjusts to follow the user's coordinates. This delay contributes to reduced user satisfaction, as the avatar's response does not immediately align with the user's movements.

During the Walk: While walking, they must simultaneously manage navigation and obstacle avoidance, increasing the complexity of multitasking. Although our system strives to position the avatar within the user's coordinate frame in a socially appropriate location while minimizing obstruction to their frontal view, it can still interfere with navigation. To better support user mobility, future adaptations should more explicitly account for users' attention to their surroundings, ensuring that the avatar does not obscure critical visual information needed for safe and effective movement. One promising direction, inspired by prior work on avatar visualization during remote collaboration [31], is to render the avatar smaller during walking phases. This approach maintains avatar visibility while reducing occlusion of the environment. This approach warrant further investigation to determine effective strategies for avatar visualization while walking.

At the End of the Walk: As users conclude a walk, such as approaching a table to sit down, the avatar may need to navigate around or through obstacles like the table to properly position itself across from the user. However, this process often leads the avatar to move through obstacles like tables, which can diminish the sense of realism and negatively impact user satisfaction [36]. This issue arises from the lack of trajectory planning between positional updates; the current system transitions the avatar using direct or linear paths, which participants described as abrupt and mechanical. To enhance spatial plausibility and preserve immersion, future work should incorporate environmental affordances into transition planning—such as avoiding obstacles and following more natural approach trajectories. In scenarios where navigating around obstacles is impractical or overly time-consuming, a possible mitigation is to temporarily render the avatar transparent during transitions that involve passing through physical objects.

6.3 Asymmetry of Control and Awareness in Avatar Locomotion

An interesting and underexplored aspect of our system design concerns the asymmetry in users' awareness and control over avatar movement. In our current implementation, remote users are not shown how their surrogate avatar repositions itself within the local user's environment. This raises several important questions: Would remote users prefer to see how their avatar moves within the local environment? Would they prefer to exert more control over their avatar's locomotion, such as selecting placement or adjusting following behaviors during conversation? More broadly, would providing visual feedback through self-mirroring or offering mobility control enhance their sense of agency, ownership, and embodiment toward their avatar representation?

Our study offers preliminary insights into these issues. Participants experienced both stationary and nomadic conditions. During the nomadic sessions, they observed the remote avatar actively following them through the environment. Participants generally perceived this behavior as natural and supportive of ongoing conversation. Interestingly, experiencing the mobile surrogate seemed to help participants empathize with how their own avatar might behave when roles were reversed. As P10 noted, "After experiencing the mobile condition, I clearly understood that avatar positioning was context-dependent. Therefore, when I later experienced the stationary condition, I naturally accepted that my avatar would change according to the other participant's environment." As a result, participants largely accepted the autonomous movement of the avatar and did not report discomfort about the absence of direct control. Notably, participants also remarked that they did not perceive the remote avatar as an extension under their own control, as the avatar's facial expressions, speech, and body gestures remained tightly linked to the actions of the remote conversational partner. Despite these positive reactions, further exploration is needed to better understand how different feedback and control mechanisms influence users' sense of agency, ownership, and embodiment toward their avatar representation in symmetric telepresence systems.

6.4 Clarity Concerns of Body Motions in Telepresence

Participants expressed significant concerns regarding the transmission of body motions through avatars when multitasking, such as preparing meals during a conversation. These unrelated body motions could lead to misunderstandings since they might not be relevant or comprehensible to the conversation partner. To mitigate this issue, future research could explore the integration of virtual representations of side tasks into the avatar's display. For example, if the user is reading a book, a virtual book could appear in the avatar's hands, providing clear contextual cues. P4 highlighted the potential benefits, stating, "I hope that virtual items can appear to interact with the avatar, allowing me to better understand what the other person is doing." Moreover, when the user needs to address an external encounter, it could be visually indicated in a way that communicates the avatar's engagement with the event, enhancing understanding without compromising the conversation.

6.5 Privacy Concerns of Body Motions in Telepresence

We identify two primary types of privacy concerns associated with transmitting and receiving avatar motions in telepresence communications. Firstly, privacy issues arise when users transmit body motions that may include activities unrelated to the conversation. To address this, future developments could provide users the option to conceal non-essential body motions that could compromise privacy. This would involve using generative motion methods to modify avatar movements so that they reflect only the conversational aspects of the interaction. Focusing avatar movements solely on dialogue prevents distractions and protects user privacy by avoiding the display of potentially sensitive or irrelevant actions. Secondly, participants also expressed privacy concerns when interacting in sensitive environments, such as restrooms, during telepresence conversations. The presence of the partner's avatar in such private settings can feel intrusive, even though the remote partner does not have visibility into the user's actual environment and the avatar is just a virtual representation. To alleviate this discomfort, participants recommended an option to temporarily disable the partner's avatar to preserve privacy. It is also crucial to notify the remote partner whenever their avatar is disabled. This notification helps maintain open communication about changes in avatar visibility, ensuring both parties are aligned in their understanding of the interaction's context and maintaining mutual comfort levels throughout the conversation.

6.6 Theoretical Reflection on Spatial Behavior

Our adaptation objectives, grounded in insights from a formative study, extend foundational theories of spatial interaction—namely, Hall's proxemic theory [12] and Marshall et al.'s F-formation framework [25]—to the context of symmetric telepresence. While these theories traditionally describe co-located interactions, we reinterpret their principles to support dynamic avatar positioning in remote environments where shared physical context is absent. In doing so, we emphasize not only social norms of distance and orientation but also the integration of local environmental affordances—such as seating, obstacles, and eye-level alignment—into adaptive avatar behavior. Importantly, we do not merely reference these theories. Instead, our formative study surfaces user behaviors—such as maintaining eye-level, avoiding occlusion, and leveraging environmental affordances—that substantiate the social relevance of these spatial norms in telepresence settings. These insights were systematically encoded into our multi-objective optimization framework using quantifiable heuristics such as Sitting Affordance, Avoid Levitation, Maintain Eye-Level, and Social Distance. Thus, while our system does not yet detect proxemic zones or F-formations in real time, it functionally reproduces their social logic through adaptive behaviors. Future work could explore more formal incorporation of proxemic distances and real-time recognition of spatial formations to enhance avatar positioning, particularly in dynamic or mobile conversational contexts.

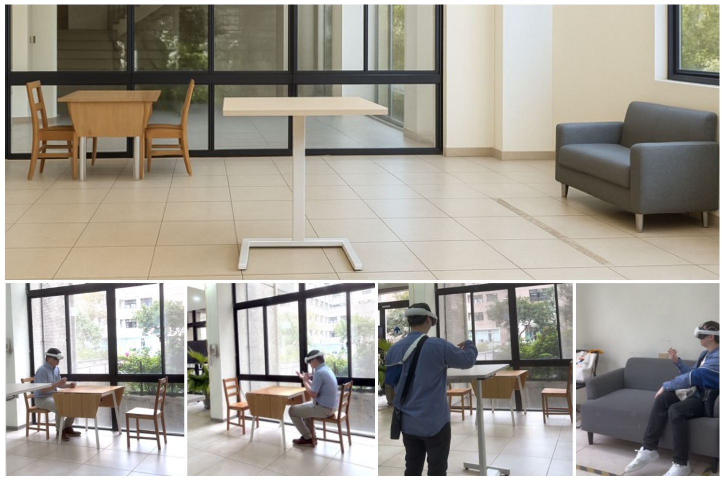


Fig. 10. Example snapshots from the exploratory deployment during the campus open day. Participants interacted with the Surrogate Avatar at various locations within the open environment.

7 EXPLORATORY STUDY

To complement our structured evaluation, we conducted an exploratory deployment of the Surrogate Avatar system in a naturalistic setting—a semi-public lobby space during a campus open day (Figure 10). This environment resembled the layout of the Nomad room used in the main study. Sixteen individuals participated, including five computer science faculty members with expertise in robotics, telecommunications, and HCI, as well as eleven alumni working in related fields. Participants interacted freely with the system without prescribed tasks or movement constraints. In this setup, participants engaged in conversation with a remote partner (one of the authors) who remained unseen, ensuring that the interaction relied solely on the avatar's representation. Users moved freely through the space—walking, sitting, standing, and transitioning between areas—while maintaining dialogue with the avatar. Although no formal metrics were collected, observational notes and participant feedback indicated that the system supported fluid communication during spontaneous movement.

Participants frequently praised adaptive behaviors such as automatic seating and maintaining appropriate proximity and orientation. Seated conversations, especially across a table, were consistently described as the most natural and satisfying. Most users reported low effort in managing the avatar. Those without prior telepresence experience intuitively likened the interaction to "talking on the phone while roaming"—accepting that the remote partner remains stationary while they move. This suggests participants readily adapted to asymmetric embodiment without needing explicit explanation. In contrast, participants with technical backgrounds showed greater curiosity. They valued the avatar's ability to follow them and proposed adding gesture-based controls, such as using a stop signal with "wait here" or beckoning to resume following. However, they also noted potential confusion, as the remote user sees the avatar as stationary and may misinterpret such commands without shared awareness of movement states.

Importantly, this open-ended deployment surfaced limitations that were not evident during the earlier controlled study. First, the current adaptation objective, which prioritizes maintaining the avatar within the user's forward field of view, makes it difficult to naturally support side-by-side conversational placements, such as sitting together on a sofa. Here, the user intentionally turned his body sideways to bring the avatar into an appropriate adjacent position (Figure 10). Second, the system's reliance on a static manual configuration of the environment prevented it from adapting to changes, such as relocated furniture. For example, one user moved a chair and later found that the avatar remained at its previous position, now awkwardly floating in the air. Third, the system lacked awareness of bystanders, at times placing avatars where support staff or other individuals were physically present, leading to spatial conflicts.

8 LIMITATIONS AND FUTURE WORK

While our study demonstrates the potential of the Surrogate Avatar system, several limitations remain. We outline key areas for improvement and directions for future research below.

8.1 Need for Baselines and Broader Validation

Our results suggest that the surrogate avatar system supports telepresence across mobility contexts, as reflected in consistent usability and presence scores. However, without a non-adaptive baseline (e.g., an avatar with a fixed spatial relation), the specific benefits of the adaptation technique remain unclear. Future work should incorporate baseline or ablation comparisons to isolate its contribution to co-presence and user experience. It is also important to examine how user factors—such as age, social roles, and familiarity—shape perceptions of avatar positioning [19], to inform the design of systems that adapt effectively across diverse contexts.

8.2 Avatar Anonymity and Embodiment

To avoid biases related to avatar appearance and streamline the study protocol, we adopted avatar anonymity as a methodological choice. Prior research suggests that avatar personalization and visual self-recognition can enhance embodiment, agency, and identification. Although participants reported high social presence and engaged naturally with their partner's avatar, the absence of visual self-representation may still have limited their sense of control or connection to the avatar's body. This trade-off between experimental control and embodied realism should be acknowledged as a study limitation and motivates future work comparing anonymous and personalized avatar designs.

8.3 Avatar Stability and Hand Tracking Limitations

Our system occasionally exhibited avatar instability due to the limited field of view of HMD-based hand tracking. When participants rested their hands outside the tracking zone—on laps or at their sides—the system lost hand data, causing visual artifacts or postural jitter. While we sought to mitigate this by instructing participants to keep their hands within view, intermittent tracking loss was still observed. This highlights a broader limitation of current HMDs in full-body tracking [15]. Future work could incorporate wrist-worn IMUs or predictive models to improve avatar stability.

8.4 Environmental Awareness and Dynamic Changes

Our system assumes a static and fully labeled environment, which limits its responsiveness to dynamic changes. In the exploratory study, users occasionally moved furniture—such as repositioning chairs—but the avatar remained anchored to its original location, leading to awkward placements like floating above the floor. Similarly, the system lacked awareness of bystanders, at times placing avatars where others were physically present, causing spatial conflicts. These issues stem from the system's inability to access real-time scene updates. Future implementations could leverage emerging capabilities in commercial HMDs that support controlled spatial sensing to detect moving objects, furniture reconfiguration, and the presence of others, enabling more socially appropriate and context-sensitive avatar adaptation.

8.5 Asymmetric Contexts: Interpretability and Privacy Concerns

Surrogate Avatar enables user mobility in non-shared environments but introduces asymmetries in contextual awareness. Participants in our study often used intuitive commands like "stay here" or gestures to direct the avatar, yet these actions could confuse remote users who lacked access to the local cues influencing avatar behavior. Beyond interpretability, this asymmetry also raises privacy concerns—avatar movements, such as reaching toward objects disconnected from the remote user's context, may unintentionally reveal private details of the local environment. Future systems should digitally augment critical cues to support mutual understanding and offer users control over contextually sensitive avatar behaviors.

8.6 Limited Support for Multi-Party Conversations

Our system currently supports only dyadic interactions, adapting avatar behavior for a single local-remote user pair. In contrast, real-world telepresence often involves group conversations with more complex spatial requirements—such as maintaining mutual visibility, appropriate distances, and shared formations (e.g., circular or L-shape F-formations). The existing adaptation framework does not address these dynamics. Future work should extend the system to handle multi-party scenarios, exploring how to coordinate multiple avatars while preserving social coherence and reducing user burden.

9 CONCLUSION

This work envisions telepresence systems that flexibly support diverse conversational settings without requiring strict environmental synchronization—allowing users to remain grounded in their own space while still feeling socially connected. We have presented Surrogate Avatar, a contextually mediated avatar that embodies the remote user and autonomously positions itself in socially and environmentally suitable locations within the local user's environment. This design aims to enhance the mobility of symmetric telepresence while preserving situated co-presence. Our findings demonstrate that the Surrogate Avatar effectively supports symmetric telepresence, adeptly adapting to the social and environmental contexts of participants in both stationary and nomadic settings. Future research should further explore dynamic context transitions—not only within symmetric telepresence, but also across hybrid and asymmetric formats—enabling users to fluidly shift between context-independent and context-rich modes as conversational needs evolve. Additionally, integrating adaptive privacy mechanisms and broader social use cases could help extend the applicability of this framework to more diverse telepresence scenarios.

ACKNOWLEDGMENTS

This research was supported by the National Science and Technology Council of Taiwan under grants 113-2223-E-A49-006-MY3, 113-2628-E-A49-013-MY2, 113-2218-E-002-044, 113-2634-F-002-007 and 111-2221-E-002-145-MY3.

REFERENCES

[1] Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. RemoteFusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Hong Kong, Hong Kong) *(VRCAI '13)*. Association for Computing Machinery, New York, NY, USA, 235–242. https://doi.org/10.1145/2534329.2534331

[2] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. A framework for designing intelligent task-oriented augmented reality user interfaces. In *Proceedings of the 10th international conference on Intelligent user interfaces*. 317–319.

[3] Soojin Choi, Seokpyo Hong, Kyungmin Cho, Chaelin Kim, and Junyong Noh. 2023. Online Avatar Motion Adaptation to Morphologically-similar Spaces. *Computer Graphics Forum* 42, 2 (2023), 13–24. https://doi.org/10.1111/cgf.14740 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14740

[4] Barrett Ens, Eyal Ofek, Neil Bruce, and Pourang Irani. 2015. Spatial Constancy of Surface-Embedded Layouts across Multiple Environments. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/2788940.2788954

[5] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. 2021. XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 290, 11 pages. https://doi.org/10.1145/3411764.3445349

[6] João Marcelo Evangelista Belo, Mathias N. Lystbæk, Anna Maria Feit, Ken Pfeuffer, Peter Kán, Antti Oulasvirta, and Kaj Grønbæk. 2022. AUIT – the Adaptive User Interfaces Toolkit for Designing XR Applications. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 48, 16 pages. https://doi.org/10.1145/3526113.3545651

[7] Daniel Immanuel Fink, Johannes Zagermann, Harald Reiterer, and Hans-Christian Jetter. 2022. Re-locations: Augmenting personal and shared workspaces to support remote collaboration in incongruent spaces. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (2022), 1–30.

[8] Kraig Finstad. 2010. Response interpolation and scale sensitivity: evidence against 5-point scales. *J. Usability Studies* 5, 3 (May 2010), 104–110.

[9] Ran Gal, Lior Shapira, Eyal Ofek, and Pushmeet Kohli. 2014. FLARE: Fast layout for augmented reality applications. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 207–212. https://doi.org/10.1109/ISMAR.2014.6948429

[10] Jens Emil Sloth Grønbæk, Ken Pfeuffer, Eduardo Velloso, Morten Astrup, Melanie Isabel Sønderkær Pedersen, Martin Kjær, Germán Leiva, and Hans Gellersen. 2023. Partially Blended Realities: Aligning Dissimilar Spaces for Distributed Mixed Reality Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 456, 16 pages. https://doi.org/10.1145/3544548.3581515

[11] Jens Emil Sloth Grønbæk, Juan Sánchez Esquivel, Germán Leiva, Eduardo Velloso, Hans Gellersen, and Ken Pfeuffer. 2024. Blended Whiteboard: Physicality and Reconfigurability in Remote Mixed Reality Collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 798, 16 pages. https://doi.org/10.1145/3613904.3642293

[12] Edward T. Hall. 1966. *The Hidden Dimension*. Doubleday, Garden City, NY.

[13] Ann Huang, Pascal Knierim, Francesco Chiossi, Lewis L Chuang, and Robin Welsch. 2022. Proxemics for Human-Agent Interaction in Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing*

Systems (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 421, 13 pages. https://doi.org/10.1145/3491102.3517593

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

[15] Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, and Hideki Koike. 2020. MonoEye: Multimodal Human Motion Capture System Using A Single Ultra-Wide Fisheye Camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 98–111. https://doi.org/10.1145/3379337.3415856

[16] Heeyoon Jeong and Gerard Jounghyun Kim. 2023. Table2Table: Merging "Similar" Workspaces and Supporting Adaptive Telepresence Demonstration Guidance. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 402–406. https://doi.org/10.1109/VRW58643.2023.00088

[17] Jiho Kang, Dongseok Yang, Taehei Kim, Yewon Lee, and Sung-Hee Lee. 2023. Real-time Retargeting of Deictic Motion to Virtual Avatars for Augmented Reality Telepresence. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 885–893. https://doi.org/10.1109/ISMAR59233.2023.00104

[18] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article 46, 8 pages. https://doi.org/10.1145/2582051.2582097

[19] Ikhwan Kim and Junghan Sung. 2024. New proxemics in new space: proxemics in VR. *Virtual Reality* 28, 2 (2024), 85. https://doi.org/10.1007/s10055-024-00982-5

[20] Seonji Kim, Dooyoung Kim, Jae-Eun Shin, and Woontack Woo. 2024. Object Cluster Registration of Dissimilar Rooms Using Geometric Spatial Affordance Graph to Generate Shared Virtual Spaces. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 796–805. https://doi.org/10.1109/VR58804.2024.00099

[21] Sarah Krings, Enes Yigitbas, Ivan Jovanovikj, Stefan Sauer, and Gregor Engels. 2020. Development framework for context-aware augmented reality applications. In *Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. 1–6.

[22] Changyang Li, Wanwan Li, Haikun Huang, and Lap-Fai Yu. 2022. Interactive augmented reality storytelling guided by scene semantics. *ACM Trans. Graph.* 41, 4, Article 91 (July 2022), 15 pages. https://doi.org/10.1145/3528223.3530061

[23] Ruyi Li, Yaxin Zhu, Min Liu, Yihang Zeng, Shanning Zhuang, Jiayi Fu, Yi Lu, Guyue Zhou, Can Liu, and Jiangtao Gong. 2024. TeleAware Robot: Designing Awareness-augmented Telepresence Robot for Remote Collaborative Locomotion. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 2, Article 70 (may 2024), 33 pages. https://doi.org/10.1145/3659622

[24] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 147–160. https://doi.org/10.1145/3332165.3347945

[25] Paul Marshall, Yvonne Rogers, and Nadia Pantidi. 2011. Using F-formations to analyse spatial patterns of interaction in physical environments. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 445–454. https://doi.org/10.1145/1958824.1958893

[26] Thomas Neumayr, Hans-Christian Jetter, Mirjam Augstein, Judith Friedl, and Thomas Luger. 2018. Domino: A descriptive framework for hybrid collaboration and coupling styles in partially distributed teams. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[27] Benjamin Nuernberger, Eyal Ofek, Hrvoje Benko, and Andrew D. Wilson. 2016. SnapToReality: Aligning Augmented Reality to the Real World. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1233–1244. https://doi.org/10.1145/2858036.2858250

[28] Allan Oliveira and Regina B. Araujo. 2012. Creation and visualization of context aware augmented reality interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Capri Island, Italy) *(AVI '12)*. Association for Computing Machinery, New York, NY, USA, 324–327. https://doi.org/10.1145/2254556.2254618

[29] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[30] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965

[31] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3290605.3300458

[32] Mose Sakashita, Ruidong Zhang, Xiaoyi Li, Hyunju Kim, Michael Russo, Cheng Zhang, Malte F. Jung, and François Guimbretière. 2023. ReMotion: Supporting Remote Collaboration in Open Space with Automatic Robotic Embodiment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 363, 14 pages. https://doi.org/10.1145/3544548.3580699

[33] Ludwig Sidenmark, Tianyu Zhang, Leen Al Lababidi, Jiannan Li, and Tovi Grossman. 2024. Desk2Desk: Optimization-based Mixed Reality Workspace Integration for Remote Side-by-side Collaboration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 44, 15 pages. https://doi.org/10.1145/3654777.3676339

[34] Maurício Sousa, Daniel Mendes, Alfredo Ferreira, João Madeiras Pereira, and Joaquim Jorge. 2015. Eery Space: Facilitating Virtual Meetings Through Remote Proxemics. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 622–629.

[35] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 1327–1339. https://doi.org/10.1145/3064663.3064707

[36] Yujie Tao, Cheng Yao Wang, Andrew D Wilson, Eyal Ofek, and Mar Gonzalez-Franco. 2023. Embodying Physics-Aware Avatars in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 254, 15 pages. https://doi.org/10.1145/3544548.3580979

[37] Theophilus Teo, Louise Lawrence, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300431

[38] Jerald Thomas and Evan Suma Rosenberg. 2019. A General Reactive Algorithm for Redirected Walking Using Artificial Potential Functions. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 56–62. https://doi.org/10.1109/VR.2019.8797983

[39] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. https://doi.org/10.1145/3332165.3347872

[40] Rishi Vanukuru, Suibi Che-Chuan Weng, Krithik Ranjan, Torin Hopkins, Amy Banic, Mark D. Gross, and Ellen Yi-Luen Do. 2023. DualStream: Spatially Sharing Selves and Surroundings using Mobile Devices and Augmented Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 138–147. https://doi.org/10.1109/ISMAR59233.2023.00028

[41] Xuanyu Wang, Hui Ye, Christian Sandor, Weizhan Zhang, and Hongbo Fu. 2022. Predict-and-Drive: Avatar Motion Adaption in Room-Scale Augmented Reality Telepresence with Heterogeneous Spaces. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3705–3714. https://doi.org/10.1109/TVCG.2022.3203109

[42] Lillian Yang, Brennan Jones, Carman Neustaedter, and Samarth Singhal. 2018. Shopping Over Distance through a Telepresence Robot. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 191 (nov 2018), 18 pages. https://doi.org/10.1145/3274460

[43] Leonard Yoon, Dongseok Yang, Jaehyun Kim, ChoongHo Chung, and Sung-Hee Lee. 2022. Placement Retargeting of Virtual Avatars to Dissimilar Indoor Environments. *IEEE Transactions on Visualization and Computer Graphics* 28, 3 (2022), 1619–1633. https://doi.org/10.1109/TVCG.2020.3018458

[44] Jacob Young, Tobias Langlotz, Matthew Cook, Steven Mills, and Holger Regenbrecht. 2019. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 1908–1918. https://doi.org/10.1109/TVCG.2019.2898737

[45] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 65 (jun 2020), 16 pages. https://doi.org/10.1145/3397331

A APPENDIX

Notation and Sampling Parameters

In the heuristic components of each adaptation objective, random displacements are generated by sampling from normal distributions $\mathcal{N}(\mu, \sigma)$. Since each objective involves different movement scales and sensitivity, the corresponding parameters are empirically tuned to achieve a trade-off between local exploration and convergence stability. This approach ensures that the optimization process remains both flexible and efficient across diverse spatial constraints.

Algorithm 1 Server

```
receive userPos per 100ms                          ▷ receive user's position
receive localEnv per 100ms                         ▷ receive user's local environment
apply userPos to objs per 100ms
apply localEnv to objs per 100ms
optimizationThresh ← threshold for adaptation
currentPos ← avatar's current position
objs ← list of all objectives
weights ← list of weights for objectives
function SOLVER(currentPos)
    if optimizationThresh > EVALUATOR(currentPos) then
        return currentPos
    end if
    for i ← 1 to 1500 do
        obj ← objs[random(0, objs.size − 1)]
        newPos ← obj.Heuristic(currentPos)
        if EVALUATOR(currentPos) > EVALUATOR(newPos) then
            currentPos ← newPos
        end if
    end for
    send currentPos                                ▷ send avatar position to client
    return
end function
function EVALUATOR(position)                       ▷ evaluator
    cost ← 0
    for i ← 0 to objs.length − 1 do
        cost ← cost + weights[i] × objs[i].Cost(position)
    end for
    return cost                                    ▷ return totalCost
end function
```

Algorithm 2 Client

```
localEnv ← local environment
userBodyPosture ←user's body posture
avatarBodyPosture ← avatar's body posture
userPos ← user's position
avatarHeadRot ← avatar's head rotation
avatarBodyRot ← avatar's body rotation
optimizationThresh ← threshold for adaptation
bodyObjs ← body objective list
headObjs ← head objective list
weights ← weights of all objectives
avatarPos ← RECEIVE(avatarPos)          ▷ apply avatar position if receive from server
function ContinuousUpdate
    receive avatarBodyPosture           ▷ receive remote user's body posture
    apply avatarBodyPosture             ▷ use remote user's body posture to render avatar
    receive remoteVoice                 ▷ receive remote user's voice
    send userBodyPosture                ▷ send body posture to remote user
    send voice                          ▷ send voice to remote user
    return
end function
function PeriodicUpdate
    send localEnv                       ▷ send local environment to server
    send userPos                        ▷ send user's position to server
    return
end function
function FrequentUpdate
    Solver(avatarHeadRot, headObjs, weights)
    Solver(avatarBodyRot, bodyObjs, weights)
    apply avatarHeadRot and avatarBodyRot
    return
end function
function Solver(currentRot, objs)
    if optimizationThresh >Evaluator(currentRot, adaptationTarget) then
        return currentRot
    end if
    for i ← 1 to 750 do
        obj ← objs[random(0, objs.size − 1)]
        newRot ← obj.Heuristic(currentRot)
        if Evaluator(currentRot) > Evaluator(newRot) then
            currentRot ← newRot
        end if
    end for
```

```
    return
end function
function Evaluator(rotation, objs, weights)                     ▷ evaluator
    cost ← 0
    for i ← 0 to objs.size − 1 do
        cost ← cost + weights[i] × objs[i].Cost(rotation)
    end for
    return cost                                                  ▷ return total_cost
end function
```

Algorithm 3 Environment Occlusion Objective

```
keypoints ← list of avatar local-space keypoints
userPos ← user position
avatarPos ← avatar position
avatarTRS ← avatar transform matrix
cost ← cost
function Cost
    occlusionCount ← 0
    for each keypoint in keypoints do
        wK ← avatarTRS · keypoint                          ▷ get keypoint in world coordinate
        if Raycast(origin = uPos, target = wK) hits then
            occlusionCount ← occlusionCount + 1            ▷ increase cost
        end if
    end for
    return occlusionCount/keypoints.size                   ▷ return normalized cost
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    randomVec ← random unit vector
    if randomValue < 0.5 then
        shuffleKs ← random shuffle of keypoints
        for each keypoint in shuffleKs do
            wK ← aTRS · keypoint                           ▷ get keypoint in world coordinate
            hitInfo ← (hit, hitpoint, surfaceNorm)         ▷ surfaceNorm is the normal at the hit
point
            if hitInfo.hit then
                hitInfo ← Raycast(origin = userPos, target = wK)
                move ← surfaceNorm ×N(1, 0.5)
                return hitpoint +move                      ▷ move away from surface
            end if
        end for
        return avatarPos
    else
        move ← randomVec × N(1, 0.5)
        return avatarPos + move                            ▷ move at random
    end if
end function
```

Algorithm 4 Social Distance Objective

```
avatarPos ← avatar position
userPos ← user position
horizontalDist ← horizontal distance between avatarPos and userPos
prefDist ← (1, 1.5)                                    ▷ min/max preferred social distance
normThresh ← (0.5, 1.0)                                ▷ thresholds for cost normalization
function Cost
    if horizontalDist < prefDist.first then
        diff ← prefDist.first − horizontalDist
        cost ← diff/normThresh.first
    else if horizontalDist < prefDist.second then
        cost ← 0
    else
        diff ← horizontalDist − prefDist.second
        cost ← diff/normThresh.second                  ▷ normalize cost
    end if
    return min(cost, 1)
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.5 then
        move ← (0, 0, 0)
        dir ← Normalize(avatarPos − userPos)
        if horizontalDist > prefDist.second then
            move ← −dir × N(horizontalDist − prefDist.second, 0.5)
        else if horizontalDist < prefDist.first then
            move ← dir × N(prefDist.first − horizontalDist, 0.5)
        end if
        return avatarPos + move                        ▷ adjust position toward preferred distance
    else
        randomX ← N(0, 0.5)
        randomZ ← N(0, 0.5)
        return avatarPos + (randomX, 0, randomZ)       ▷ random move
    end if
end function
```

Algorithm 5 Field of View Objective

```
userStatus ← user's status: Idle or Walking
avatarPos ← avatar position
userPos ← user position
userFacingDir ← user facing direction
dirToAvatar ← direction from user to avatar
FOVAngle ← (45, 80)                                  ▷ threshold for preferred FOVAngle
FOVCost ← cost for FOVAngle interval
cost ← cost
if m == I then
    FOVCost ← (0, 0.5)
else
    FOVCost ← (0.5, 0)
end if
function Cost
    θ ← angle(userFacingDir, dirToAvatar)    ▷ angle between facing direction and Avatar
    if θ < FOVAngle.First then
        cost ← FOVCost.First
    else if θ < FOVAngle.Second then
        cost ← FOVCost.Second
    else
        cost ← 1
    end if
    return cost                                      ▷ return cost
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.5 then                        ▷ move avatar toward user front
        dist ← magnitude(avatarPos − userPos)
        offset ← userFacingDir × dist
        targetPosition ← userPos + offset
        moveVector ← targetPosition − avatarPos
        return avatarPos + moveVector × N(0.5, 0.5)
    else                                             ▷ move at random
        randomVec ← random unit vector
        randomMove ← randomVec × N(0.5, 0.5)
        return avatarPos + randomMove
    end if
end function
```

Algorithm 6 Sitting Affordance Objective

```
chairs ← list of all chairs in the environment
avatarPos ← Avatar Position
cost ← cost
d ← distance to check chair below avatar
function Cost
    hitInfo ← (hit, hitObject)
    hitInfo ← Raycast(origin = avatarPos, downVector, d)
    hit ← Raycast(origin = avatarPos, downVector, d)
    if hitInfo.hit ∧ hitInfo.hitObject is Chair then
        cost ← 0
    else
        cost ← 1
    end if
    return cost
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.33 then
        chair ← closest chair from user
        return chair.pos
    else
        randomIndex ← random int ∈ [0, chairs.size)
        return chairs[randomIndex].pos
    end if
end function
```

Algorithm 7 Levitation Objective

```
avatarPos ← avatar position
avatarTRS ← avatar transform matrix
keyPoints ← list of keypoints on avatar bottom bounding box
levitatingThresh ← allowed height gap
surfaces ← list of all surfaces in environment
function COST
    levitatingPoints ← 0
    for each keyPoint ∈ keyPoints do
        wK ← avatarTRS · keypoint                         ▷ get keypoint in world coord.
        downVec ← downVector × levitatingThresh
        rayEnd ← wK − downVec
        hitInfo ← (hit, hitObject)
        hitInfo ← Raycast(origin = wK, target = rayEnd)
        if hitInfo.hit ∧ hitInfo.hitObject is Surface then
        else
            levitatingPoints ← levitatingPoints + 1
        end if
    end for
    cost ← 1 − levitatingPoints / keyPoints.size
    return cost
end function
function HEURISTIC
    if cost > 0 and surfaces is not empty then
        randomValue ← random value ∈ [0, 1)
        if randomValue < 0.33 then
            surface ← get closest surface to avatar
            avatarPos ← surface.position
        else
            randomIndex ← random int ∈ [0, surfaces.size)
            avatarPos ← surfaces[randomIndex].pos
        end if
        adjustY ← avatarHeight + threshold
        avatarPos.y ← avatarPos.y + adjustY / 2
    end if
    if randomValue < 0.5 then
        randomX ← N(0, 0.5)
        randomZ ← N(0, 0.5)
        return avatarPos + (randomX, 0, randomZ)
    end if
    return avatarPos
end function
```

Algorithm 8 Eye Level Objective

```
acceptableDist ← 0.1                                    ▷ tolerated range
normThresh ← 0.25                                       ▷ for cost normalization
avatarPos ← avatar position
cost ← cost
avatarEyeY ← avatar eye height
userEyeY ← user eye height
function Cost
    diff ← |avatarEyeY − userEyeY|
    if diff ≤ acceptableDist then
        return 0                                        ▷ well aligned
    else
        cost ← min((diff−acceptableDist)/normThresh, 1)
        return cost                                     ▷ normalized cost
    end if
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.5 then
        adjustY ← 0
        if avatarEyeY > userEyeY then
            adjustY ← −N(avatarEyeY − userEyeY − acceptableDist, 0.5)
        else
            adjustY ← N(userEyeY − avatarEyeY − acceptableDist, 0.5)
        end if
        avatarPos.y ← avatarPos.y + adjustY
    else
        avatarPos.y ← avatarPos.y + N(0.5, 0.5)         ▷ random move
    end if
    return avatarPos
end function
```

Algorithm 9 Spatial Consistency Objective

$avatarPos \leftarrow$ avatar position
$prevAvatarPos \leftarrow$ avatar previous position
$dirToPrevPos \leftarrow$ direction from avatar's current position to previous position
$normThresh \leftarrow$ threshold for cost normalization
$cost \leftarrow$ cost
function Cost
    $distToPrevPos \leftarrow HorizontalDist(prevAvatarPos, avatarPos)$ ▷ horizontal distance
    $cost \leftarrow Min(1, distToPrevPos/normThresh)$
    return $cost$ ▷ return cost
end function
function Heuristic
    $randomValue \leftarrow$ random value $\in [0, 1)$
    if $randomValue < 0.5$ **then** ▷ pick heuristic at random
        return $avatarPos + dirToPrevPos \times \mathcal{N}(0.5, 0.5)$
    else
        $randomVec \leftarrow$ random unit vector
        return $avatarPos + randomVec \times \mathcal{N}(0.5, 0.5)$
    end if
end function

Algorithm 10 Head Toward User Objective

```
avatarHeadDir ← current avatar's head direction
dirToUser ← direction from avatar to user
anglethresh ← (15, 30)                                   ▷ face user angle threshold
angleNormThresh ← threshold for Cost normalization
function Cost
    angleDiff ← angle(avatarHeadDir, dirToUser)   ▷ angle between avatar's head direction
and direction from avatar to user
    if angleDiff < anglethresh.First then
        return 0
    else if angleDiff < anglethresh.Second then
        return min(1, angleDiff/angleNormThresh)
    else
        return 1
    end if
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.5 then                            ▷ rotate head to face user
        return dirToUser
    else                                                 ▷ rotate head randomly
        θ ← random angle ∈ (−π, π)
        avatarHeadDir ← rotY(avatarHeadDir, θ)
        return avatarHeadDir
    end if
end function
```

Algorithm 11 Body Toward Head Objective

```
bodyDir ← avatar's body direction
headDir ← avatar's head direction
angleRange ← (45, 80)                       ▹ angle thresholds for head-body angle difference
angleCost ← (0, 0.5)                        ▹ Cost for head-body angle difference
cost ← cost
function Cost
    angleDiff ← horizontalAngle(bodyDir, headDir) ▹ angle between avatar's head direction
and body Direction
    if angleDiff < angleRange.First then
        cost ← angleCost.First
    else if angleDiff < angleRange.Second then
        cost ← angleCost.Second
    else
        cost ← 1
    end if
    return cost
end function
function Heuristic
    randomValue ← random value ∈ [0, 1)
    if randomValue < 0.5 then                ▹ pick heuristic at random
        return headDir                       ▹ rotate avatar's body to head
    else
        θ ← random(−π, π)                    ▹ angle around y-axis
        bodyDir ← rotY(bodyDir, θ)
        return bodyDir                       ▹ avatar's body rotate at random
    end if
end function
```