# Immersive Social Teleoperation Interface with Semi-automatic Ingroup Navigation for Intuitive Communication

Akitomo Takeda
Kyoto University
Kyoto, Japan
takeda@robot.soc.i.kyoto-u.ac.jp

Stela Hanbyeol Seo
Kyoto University
Kyoto, Japan
stela.seo@i.kyoto-u.ac.jp

Satoru Satake
ATR
Souraku-gun, Japan
satoru@atr.jp

Takayuki Kanda
Kyoto University
Kyoto, Japan
ATR
Souraku-gun, Japan
kanda@i.kyoto-u.ac.jp

Figure 1: Our immersive teleoperation interface allows the operator to look around and have social interaction while automatically navigation the robot. We used a head-mounted display (right) to map the operator's head movement with the viewport on the 270-degree camera stream (middle) of the environment (left).

## Abstract

We introduce a novel immersive social teleoperation interface to perform social interaction intuitively and semi-automatic locomotion simultaneously. Teleoperated robots in complex, human-centric social environments present a significant challenge on simultaneously managing intricate navigation while engaging in natural social interaction. This dual task imposes a high cognitive load, hindering the fluidity and quality of social interaction. As existing systems typically prioritize either task-oriented teleoperation control or non-moving social interaction, failing to integrate dynamic locomotion with social engagement effectively. Our system combines a head-mounted display with a wide-angle, 270-degree video feed to support extensive situation awareness and a strong sense of presence to overcome these limitations by fostering an immersive and socially aware experience. Our interface performs low-level navigation of the robot by pointing at a place to go and selecting a person to follow. The operator can focus on high-level goals (social interactions). We evaluated our interface through a rigorous field experiment, using a testbed (remote guide) scenario developed through iterative pilot studies in a real-world shopping mall. Our findings demonstrate that the operator's task performance improves statistically significantly. We report other findings and discuss limitations and future improvements. In short, our novel interface, integrating immersive visualization with autonomous navigation, enables operators to achieve a more intuitive and engaging social interaction in dynamic remote social environments.

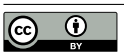
## CCS Concepts

• **Human-centered computing** → **User centered design**; **Field studies**; *Mixed / augmented reality*.

## Keywords

teleoperation, social interaction, autonavigation, head-mounted display

**ACM Reference Format:**
Akitomo Takeda, Stela Hanbyeol Seo, Satoru Satake, and Takayuki Kanda. 2026. Immersive Social Teleoperation Interface with Semi-automatic Ingroup Navigation for Intuitive Communication. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction (HRI '26), March 16–19, 2026, Edinburgh, Scotland, UK*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3757279.3785614




*HRI '26, Edinburgh, Scotland, UK*

ACM ISBN 979-8-4007-2128-1/2026/03
https://doi.org/10.1145/3757279.3785614

# 1 Introduction

Modern society is profoundly shaped by advanced technologies including telecommunication, immersive head-mounted displays (HMDs), and sophisticated mobile navigation platforms. Leveraging these advancements, particularly in robotics and automation algorithms, robots are increasingly capable of extending human presence and agency into distant environments. This progress has transformed the once-futuristic concept of remote work, exploration, and interaction into a tangible reality. However, despite these significant strides, a critical limitation persists: the challenge of simultaneously engaging in social interaction and efficiently navigating a robot within complex, human-centric social spaces.

Consider window shopping, a seemingly effortless task for people. When engaging in this activity, individuals can concurrently converse, observe their surroundings, and plan intricate navigation paths, all without conscious effort or formal training. For a robot, however, and particularly when performed via teleoperation, this seemingly simple task becomes surprisingly complex and difficult. This raises a fundamental question: can a teleoperator genuinely achieve such an interactive and socially fluid experience?

This difficulty can be attributed to two primary teleoperation challenges: (1) maintaining a high-level of situation awareness for the operator and (2) effectively performing navigation tasks within dynamic and complex social environments. Socially appropriate and safe navigation is critically dependent on an operator's comprehensive situation awareness. While manual navigation in such settings is cognitively taxing, we propose that parts of the navigation process can be effectively automated. Much like how people unconsciously coordinate their legs while walking, certain subconsciously performed low-level tasks can be offloaded to the robot. This allows the operator to focus on high-level decisions, such as following a companion, moving towards a specific point of interest, or initiating short bursts of manual maneuvering. Furthermore, by providing a continuous, comprehensive view of the robot's surroundings, mapped with the operator's head motion, operators can sustain a sufficient level of situation awareness.

In this work, we present a realistic teleoperation interface specifically designed to facilitate immersive social interactions by addressing the aforementioned challenges (Figure. 1). Through a rigorous field experiment, we provide empirical evidence of our interface's effectiveness, comparing it to an existing telepresence interface in a real-world daily task, window shopping. Our findings demonstrate the critical importance of thoughtful interface design, particularly when enabling operators to engage intuitively with a person while navigating the robot in socially interactive remote environments.

# 2 Related Work

## 2.1 Task-oriented Teleoperation: Prioritizing Performance and Awareness

A substantial body of research in teleoperation has focused on enabling operators to perform complex, instrumental tasks in remote or hazardous environments. Applications such as urban search and rescue [7, 16, 19], deep-sea exploration [6, 13], and industrial material handling [2, 14] demand systems that maximize operator performance and safety. A key focus in this area is enhancing the operator's situation awareness through advanced information visualization. Interfaces often present comprehensive macro- and micro-level data about the robot's state [17, 18], detailed environmental models [5], and other task-specific information [4].

However, the design philosophy of these systems prioritizes technical mastery and task completion. The information flow is primarily robot-to-operator, centered on sensor data and control parameters. Consequently, the social dynamics of the remote environment are often overlooked. The operator is treated as a pilot managing a complex machine, not as a social agent. While these systems excel at providing the awareness needed to complete a task, they are not meant to support the nuanced perception and interaction required for fluid social engagement.

## 2.2 Social Telepresence: Facilitating Connection at the Cost of Mobility

In contrast to task-oriented systems, social telepresence robots are explicitly designed to facilitate remote social interaction and foster a sense of connection [15]. These platforms enable remote participation in meetings, medical consultations, and family gatherings [1, 15]. To achieve this, designers often prioritize the clarity of audio-visual communication and the representation of the remote user's presence, for instance, through a screen displaying their face.

A critical trade-off in many of these systems is the simplification or heavy constraint of mobility. To reduce the operator's cognitive load and technical burden, many social telepresence robots are designed to be relatively stationary or have limited navigational capabilities [3]. Navigation, if required, often depends on assistance from people in the remote environment. This design choice fundamentally alters the nature of social engagement, shifting it from a dynamic, mobile activity to one that more closely resembles a video call with a physical embodiment. It precludes a wide range of natural human interactions, such as walking alongside a companion, approaching a group, or non-verbally negotiating passage. This lack of dynamic physical agency presents a significant barrier to achieving a truly immersive and natural social presence.

## 2.3 The Challenge of Social Navigation in Shared Spaces

Navigating a robot in a human-centric environment introduces a distinct set of challenges related to safety, social conventions, and shared spatial understanding [10]. This challenge manifests in two ways for a teleoperator: the cognitive load of manual control and the social limitations of autonomous control.

Manually navigating a robot through a dynamic social space is a cognitively demanding task that directly competes for the operator's attention. The constant need to plan paths, avoid obstacles, and control the robot's low-level movements detracts from the mental resources available for social interaction, such as listening, speaking, and interpreting non-verbal cues.

To reduce this burden, much research has focused on autonomous, social-aware navigation. However, these systems are typically designed with the paramount goal of ensuring human safety. This translates into highly conservative and passive behaviors, where the robot defaults to yielding to people and maintaining a large personal space. This operational philosophy can create an uneven

power dynamic, where the teleoperator lacks the agency to lead, take initiative, or navigate assertively alongside a partner [12]. This passivity can disrupt the natural flow of co-navigation and hinder the operator's sense of presence and control.

### 2.4 Synthesizing Agency and Interaction: Our Contribution

The literature often separated into three different focuses: task-oriented systems provide navigational agency without social awareness; social telepresence systems provide social connection without mobile agency; and autonomous navigation prioritizes safety over the operator's interactive agency. This leaves a critical gap for applications requiring a teleoperator to be both a competent navigator and a fluid social actor. Systems that optimize for one domain often do so at the expense of the other [9, 11].

Our work addresses this gap by presenting a novel teleoperation interface that synthesizes these disparate goals. By providing a wide, immersive view mapped to head motion, we enhance the operator's social awareness. By offloading low-level navigation to autonomous modes (lead or follow), we reduce cognitive load while preserving their social agency. Our contribution lies in designing and empirically validating an interface that enables operators to seamlessly manage spatial navigation and social engagement, fostering a more balanced, intuitive, and embodied telepresence experience.

## 3 Immersive Teleoperation Interface for Social Navigation and Interaction

We design and implement a novel teleoperation interface to address a challenge of social interaction while navigating a robot. Our primary design goal is to create a system that fosters a strong sense of presence and social awareness for the remote operator while ensuring safe and fluid navigation in dynamic social environments.

After our iterative design process, our successful teleoperation system provides two major benefits: (1) it updates its view perspective by mapping the operator's head movement resulting in the high-level situation awareness which is critical for social navigation and (2) it allows the operator to offload the low-level navigation tasks allowing the operator to establish an immersive and intuitive social interaction. These benefits are realized through a combination of a wide-angle (270-degree field of view), real-time visual interface and an autonomous navigation control that offloads the cognitive burden of low-level path planning from the operator.

### 3.1 Improving Immersiveness from Traditional Telepresence Interfaces

We performed a pilot test with a mobile robot with a forward-facing, wide-angle camera that streamed video to a standard desktop monitor, much like a commercial telepresence robot interface. While this setup is common and widely used among commercial telepresence robots, our test quickly revealed significant limitations in the context of social interaction. As such, we shifted our focus toward creating a more holistically immersive experience. The primary bottleneck was the limited visual feedback without spatial information. To overcome this, we replaced the previous camera with a single 360-degree camera mounted on the robot. This provided a complete, panoramic view of the remote environment.

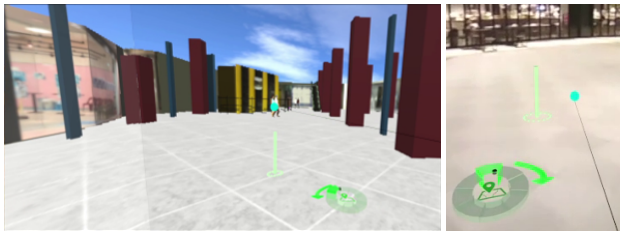


Figure 2: The leading mode appears in a simulation (left) and in an evaluation session (right). The movement indicator, which appears on the bottom, shows the detailed movement.

Visualizing 360-degree video feed on a standard monitor creates cognitive overload to understand the relations shown in the visual feed. We use this opportunity to leverage a HMD. Streaming and rendering a 360-degree video feed in a game engine is technologically straightforward, and by streaming the video and directly mapping the operator's head movement into viewport, the operator can see where they look at. There are no requirements of cognitive mapping of the visual data and the real-world location of an object shown in the video. This change dramatically improved the sense of presence, as operators in pilot studies reported feeling as though they were truly there. However, a significant bottleneck remained: the control mechanism (i.e., the cognitive load of manual navigation). Even with an immersive view, operators were still preoccupied with the low-level task of steering the robot, which detracted from their ability to fully engage in social interaction.

### 3.2 Improving Social Navigation through Semi-autonomy

Regarding the navigation control problem, our goal is to reduce the cognitive load associated with manual maneuvering, freeing the operator to focus on their high-level goals and social engagement. People do not consciously plan every footstep; people simply decide their destination or what to do (e.g., follow or lead a person), and the body executes the low-level movements unconsciously.

To realize this functionality in our social teleoperation systems, we implemented an autonomous navigation system on top of a point-and-go paradigm. Given the immersive nature of the HMD, we designed the control interface to be minimal. Instead of a complex dashboard, the operator uses a simple handheld controller to designate high-level navigation goals (a point of interest or a person) within the equirectangular view.

For instance, the operator can pull the trigger while pointing a goal direction. Then the system places a virtual marker, performs autonomous navigation toward the direction. Another option is to point at a person (who is within 5 meter range) and pull the trigger letting the system perform autonomous navigation toward the person's side. The navigation stack will continuously plan to stand beside the targeted person (often social interaction partner). These two different navigation systems allow the operator to delegate the task of navigation and dedicate their full attention to observing the environment and interacting with their companion.

Figure 3: The following mode appears in a simulation (left and middle) and in an evaluation session (right). When an operator points at a person nearby the robot, the augmented highlight appears to indicate the target person. Once the operator selects the person by pulling the trigger, the mode activates and the robot starts moving.

### 3.3 Reducing the Network Bandwidth Requirement for the Visual Data

While performing system pilots, we exceeded the local 5G bandwidth limits and the operators did not look at their direct rear. To address the network bandwidth limitation, while the 360-degree camera can stream a 3840x1920 equirectangular video at 29.97 frames per second, we reduced the resolution down to its quarter size and slowed its frames delivery. Further, we programmatically cropped the streaming data to a 270-degree forward-facing view, removing the rear 90-degree view. This design choice prevents the socio-perceptual dissonance that can arise from a superhuman rear-view capability while retaining a wide, immersive perspective critical for social and situation awareness. Our final streaming data size is 1440x960 at 15 frames per second with 270-degree view angles.

### 3.4 The Final Version Interface

The final version of our immersive teleoperation interface integrates these iterative refinements into a cohesive system designed to prioritize social presence and fluid interaction. Our core components of the system are detailed below while detailing our iterative design process (i.e., pilot studies) in the appendix.

**Immersive Visual System:** The operator wears a HMD that presents a live, 270-degree panoramic video stream from the robot (Figure 1, middle). This provides a complete and uninterrupted view, allowing to achieve the high level situation and social awareness. Viewpoint control is managed intuitively through the operator's head movements, decoupled from the robot's locomotion.

**Minimalist Heads-Up Display:** A simple, non-intrusive heads-up display is rendered within the HMD to display essential information and provide a targeting reticle for navigation commands without breaking the sense of immersion (Figure. 2). We color the indicator based on the mode (black for the manual mode or not moving, green for the leading mode, and blue for the following mode). The indicator consists of eight blocks which are enlarged and highlighted when the robot is moving toward the direction. The small ball inside of the enlarged block represents the detailed fine-grained direction. The arrow indicator beside the blocks only appears when the robot is turning. This indicator can be used as an anchor to know the robot's body direction as the heads-up display is always located in the middle of the equirectangular screen.

**Navigation Mode:** The operator uses a handheld controller to issue high-level commands. The controller vibrates when the robot is moving in the field. The commands include:

- **Leading Mode (Point-To-Go):** This mode automates navigation to a specific point. The operator aims a virtual laser pointer with the right controller and pulls the trigger to set a destination, marked by a persistent flag icon (Figure. 2). The robot's navigation stack then autonomously plans and executes a path to the goal. Since the operator can override the goal by setting a new destination, they do not need to accurately pin their destination. In this mode, the movement indicator turns green and shows a map in the middle, and an augmented pin to the destination appears in the space. If the pin could not be shown because it is out of the operator's field of view, a triangle arrow mark appears at the edge of the screen.
- **Following Mode (Person-Following):** This mode automates to move alongside a partner. The operator uses the laser pointer to select a person (Figure. 3, left), who is then tagged with a persistent triangular marker on top of their head (Figure. 3, middle). The movement indicator turns to blue and shows people in the middle. The system then autonomously moves the robot to locate beside the partner.
- **Manual Mode:** The operator has direct control over the robot's holonomic movement. The left controller pad governs forward/backward and rotation.

This design effectively offloads the cognitive burden of low-level maneuvering, freeing the operator to focus their attention on observing the environment, engaging in conversation, and making high-level strategic decisions. Our pilot test shows that the operator can effectively shift their role between leading or following with a social participant, fostering a more genuine sense of embodiment and enabling more intuitive social navigation and interaction.

**Following Mode Computation:** The system uses the past two-second position data to compute a partner's speed (total moved distance over time), and the past three-second position data to calculate orientations (average angle of the point pairs) if they have moved a cumulative of 30cm or more. Using this velocity, the robot predicts the partner's position in three-second future and identifies two potential destinations located 0.8m to the left and right of the partner's future position and navigates toward the closer one.

**Following Mode Velocity:** The robot's velocity is determined by the following logic:

- If the robot is behind the partner: It moves at maximum velocity.
- Otherwise (if the robot is not behind):

  - If the robot and partner have the same orientation:
    * If the partner is stationary, the robot approaches at a constant velocity.
    * If the partner is moving, the robot matches the partner's velocity.
  - If the orientations differ: The robot moves at maximum velocity.
- Exceptions: In all cases, the robot gradually decelerates within 0.8m of its destination or if a person is in close proximity.

## 4 Robot Platform and Safety Systems

The teleoperation system consists of a custom-built robotic platform and an operator interface, connected via a high-throughput, low-latency 5G wireless link. The design prioritizes safe, natural, and fluid movement within dynamic social environments.

The robotic platform is built upon an omni-directional drive system (Figure. 5, left). This holonomic base enables simultaneous translation and rotation, allowing the robot to perform complex, human-like maneuvers such as sidestepping to yield space or turning to face a conversational partner without altering its primary trajectory. This capability is fundamental to achieving a more natural and less disruptive physical presence.

A LiDAR sensor (Figure. 5, right) generates a continuous 3D point cloud of the surroundings. This data is processed in real-time to detect, segment, and track people within the environment. This capability serves as the perceptual foundation for the interface's person-following navigation mode.

To ensure safe operation in shared spaces, the robot incorporates a dual-layered sensing architecture for robust obstacle and person detection. The platform is outfitted with a ring of laser distance sensors that continuously monitor the robot's immediate vicinity. If any object breaches a predefined safety threshold, the system triggers an automatic emergency stop, overriding all operator or autonomous commands to prevent imminent collisions. In addition, an on-site researcher continuously observes the situation and pulls the emergency stop if any unexpected situation happens (e.g., a child gets close to the robot).

## 5 Evaluation

We conducted a within-subject user study to evaluate our immersive social teleoperation interface.

### 5.1 Conditions

In this evaluation, we have two conditions, our proposed immersive teleoperation interface and a commercially available telepresence interface. The order was counterbalanced across participants. We chose a commercial telepresence robot (Double 3 Telepresence Robot) and its control interface as the baseline. It provides a point-to-go interface and stops automatically near obstacles. However, in our outdoor test, the robot misrecognized bright sunlight as an obstacle and could not move further. As such, we implemented a mimicry interface with our robot (e.g., view angle and interface components, such as feedback on top of the screen during navigation, Figure 4). For the baseline interface, participants use keyboard and mouse (like the original Double Robotics interface), whereas they use HTC Vive Pro 2 for our proposed interface (Figure 6).

Figure 4: We implemented a mock telepresence interface working with our mobile teleoperation robot.

### 5.2 Scenario

To empirically evaluate our interface, we need a scenario that would compel an operator to simultaneously engage in social interaction and navigate a complex, dynamic environment. The chosen task needed to create a genuine need for the high-level situation awareness and low cognitive load navigation that our system was designed to support. As such, we developed a remote guide scenario that effectively simulated these demands.

Initially, we considered a simple window shopping task, where the operator and a remote partner would walk together and converse. However, pilot studies revealed that operators often defaulted to a robot pilot mindset, focusing on technical control rather than social engagement whereas the remote partner was often accommodating, waiting patiently for the robot, and simplifying the navigation challenges. As this changes social dynamics (robot being socially inferior), we refined the scenario into a remote guide task. In this setup, the teleoperator acts as a host guiding a curious visitor through an interesting location. Another participant beside the robot (i.e., walker) is now our confederate acting as a visitor. We requested the visitor to act like a self-centered and easily distracted individual, frequently pointing out objects of interest and asking questions. As a result, the scenario maintains the nature of window shopping and leads the knowledgeable person of the shopping mall to guide the other person around.

This design imposed two critical demands on the operator: (1) they had to maintain constant social engagement to provide explanations and guide their partner and (2) they had to maintain high-level of situation awareness to spot the objects their partner was interested in and navigate effectively towards them. The operator could not simply pause navigation to talk; they had to integrate conversation and movement seamlessly, forcing them to rely on the interface's features for autonomous navigation and immersive awareness.

### 5.3 Environment

We evaluate our proposed immersive teleoperation interface during teleoperated social interaction. The experiment was conducted in the open outdoor space of a local shopping mall (Figure. 1, middle). This large, open space was chosen to provide a realistic public environment with sufficient room for a person and a robot to move around together. The open space is surrounded by several storefronts, including a retail shop, a restaurant, and a cafe, which served as points of interest for the task. The second floor of the building

Figure 5: Our mobile robot platform (left) and the closeup shot of the back pole with additional sensors (right). We mounted a lidar sensor, a 360-degree camera, a directional microphone, and a local 5G network router. The robot is equipped with an omni-drive base.

(e.g., a children's gym and a dentist) are also visible and can be observed through the robot's camera. We chose to conduct our experiment only on weekdays as pedestrian traffic was significantly lower compared to the weekend.

In Japan, video and audio recording in public spaces for research purposes is not prohibited. However, we take people's privacy seriously. In addition to a request from the management of the shopping mall to install a poster stating the ongoing experiment and the videorecording, we modify videos for public presentations so that viewers cannot identify visitors' faces. Furthermore, for our supplementary video, we carefully selected scenes where we could not hear any private conversations and blurred the faces of any passersby if they were recognizable. We will continue to abide by this policy for this work. We attached the poster installed in the experiment space along with an English translation as an appendix.

## 5.4 Procedure

Each pair of participants engaged in a brief icebreaking session. They introduced themselves and discussed neutral topics such as their hometowns and hobbies to build initial rapport. To protect their anonymity in audio and video recordings, participants were asked to decide their nicknames to use throughout the experiment.

After this session, the procedure was divided for the two roles. The participant assigned as the operator received a short training session on where to introduce the environment and how to use the interfaces. Concurrently, the participant in the visitor role was escorted to the outdoor space by an on-site researcher. The researcher familiarizes the visitor with the designated experimental area and practices the actions to perform as a confederate.

The visitor (confederate) has a total of five unexpected actions to perform throughout a session: playing with a phone while not listening to the explanation, not following the robot (or moving to an opposite direction from the robot), looking at different places, asking about their fashion (specifically a keyholder attached on a bag), and suddenly asking other point of interest. We believe

Figure 6: The head-mounted display headset and joysticks. We put stickers on joysticks to provide tactile sense of direction on the pad to operators while wearing the headset.

all these unexpected situations could be addressed if the operator maintains a high level of situation awareness.

Once both participants were ready, the interaction trials began. The session in average lasts about 10 minutes – we did not set any time limits. Immediately after each interaction trial, we administered a post-condition questionnaire.

After finishing all trials, we administered a post-study questionnaire and conducted semi-structured interviews if time permits.

## 5.5 Measurements

Our main measurement is the count of operators' reactions to the visitor's unexpected situations. If they handled the situation, we count it; if they failed to notice, we do not count the situation. We believe that their performance task is closely related to their cognitive load as if under high cognitive load condition, they may not be able to think of other tasks (i.e., high focus and missing to observe unexpected situations).

The self-reported questions in the post-condition questionnaire contains the Coordination Rapport Scale [8] and a set of questions (how intuitive operation was, were they able to constantly understand the robot's surroundings, were they able to have a smooth conversation, and were they able to guide the visitor considerably well) in 5-level Likert scale. The score from the Coordination Rapport Scale has been normalized by the number of questions.

The post-study questionnaire asks about their prior experience with gaming, robotics, their familiarity with the shopping mall before the experiment, and their interface preferences and reasons.

## 5.6 Participants

We recruited 13 pairs of participants (26 = Total number of participants) for this iterative exploration study (6 females for the operator role with age $M$=32.85,$SD$=10.94, and 11 females for the visitor role with age $M$=42.15,$SD$=8.73). We booked several more sessions; however, we had to cancel the sessions because the gear(s) in the robot base started malfunctioning and the repair could not be done in time. Each participant received an honorarium of 4000 JPY for their time. The experimental protocol was reviewed and approved by the Institutional Research Ethics Board of Kyoto University.

# 6 Results

We conducted a within-subject user study comparing our proposed interface to a mock commercial telepresence interface as a baseline.

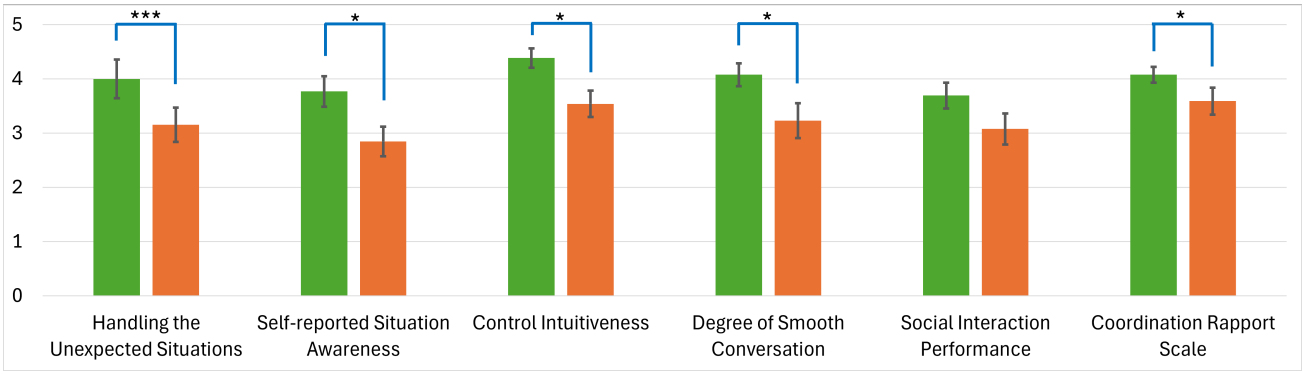



Figure 7: The two-tailed paired samples t-test reveals statistical significant differences between our proposed interface and the baseline interface. Error bar indicates standard error, *** indicates <.001 whereas * indicates <.05.

Further, to complement our quantitative findings, we conducted semi-structured interviews with the operators to gather qualitative insights into their experiences with both interfaces.

To quantify operator reactions, two researchers rated them during each session (one observing the operator and the other with the visitor/confederate). They marked a scripted, unexpected situation as unhandled if the operator did not provide any direct reaction within approximately 30 seconds. The Cohen's kappa of two researchers' coding is $\kappa = 0.944$. For the final result, after the experiment, they compared their results for each situation and reviewed video recordings to resolve any disagreements.

## 6.1 Verification of Hypothesis

We conducted two-tailed paired samples t-test to assess our interface's superiority in terms of handling the unexpected situations (situation awareness), self-reported situation awareness, control intuitiveness, degree of smooth conversation, social interaction performance, and the coordination rapport scale (Figure 7). The paired samples statistics reveal that the operator were able to handle more number of unexpected situations with our proposed system ($M$=4.00, $SD$=1.29) than the baseline interface ($M$=3.15, $SD$=1.14). The paired samples t-test results reveal a statistically significant mean difference between the proposed and baseline interfaces, $t(12)$=5.50, $p$<.001. Cohen's D showed a large effect size of $d$=1.525. This supports our main challenge in improving the operator's situation awareness using our interface.

In addition, the paired samples statistics reveal that statistically significant mean differences between the proposed and baseline interfaces in terms of self-reported situation awareness $t(12)$=2.52, $p$=.027 with a medium effect size of $d$=0.699 ($M$=3.77, $SD$=1.01 vs. $M$=2.85, $SD$=0.99); control intuitiveness $t(12)$=3.81, $p$=.002 with a large effect size of $d$=1.057 ($M$=4,38, $SD$=0.65 vs. $M$=3.54, $SD$=0.88); degree of smooth conversation $t(12)$=2.67, $p$=.020 with a medium effect size of $d$=0.740 ($M$=4.08, $SD$=0.76 vs. $M$=3.23, $SD$=1.17); and coordination rapport scale $t(12)$=2.36, $p$=.036 with a medium effect size of $d$=0.655 (normalized $M$=4.08, $SD$=0.53 vs. $M$=3.59, $SD$=0.89).

While there were no statistical significant differences, it showed a trend toward significance on self-reported social interaction performance $t(12)$=1.98, $p$=.071 ($M$=3.69, $SD$=0.86 vs. $M$=3.08, $SD$=1.04).

## 6.2 Interview Results

From our semi-structured interviews, we found that the feedback consistently highlighted the advantages of our proposed system in terms of intuitive control, enhanced awareness, and a stronger sense of presence, which corroborated our quantitative results. We have organized the key themes from these interviews below.

**Intuitive Control and Reduced Cognitive Load:** Operators frequently praised the intuitive control scheme of our proposed interface. The separation of autonomous commands (right controller) and manual controls (left controller) was described as "easy to understand." The point-to-go feature was particularly well-received for guiding tasks. One operator noted, "With the HMD, you just move your wrist in the direction you want to go and point, and it moves for you, so guiding is easy" (P3). Another commented that "pointing the controller was an easy way to specify a destination, even when turning" (P12). Additionally, the haptic feedback from the controller vibrating during movement was cited as a helpful and clear cue for understanding the robot's state.

**Enhanced Situation and Social Awareness:** The most significant feedback centered on the dramatic improvement in situation and social awareness. Operators attributed this primarily to the ability to decouple their viewpoint from the robot's movement. One participant explained, "With the HMD, it's easy to look back [at my partner] while moving forward, whereas with the baseline, I had to stop completely to check" (P6). This capability was crucial for maintaining engagement during the co-navigation task.

This enhanced awareness directly translated into more fluid and natural social interactions. Operators reported that it was easier to see their partner's facial expressions and gauge their reactions. As one participant stated, "With the HMD, I felt like I was right there, looking at my partner's face while talking, so I could even make more lively reactions" (P7). In contrast, the baseline interface was described as "too much of a multitasking challenge to operate the robot and talk at the same time" (P4).

The wide field of view and quick, head-mapped viewpoint control also allowed operators to rapidly reorient themselves and locate their partner if they became separated. One operator contrasted the two experiences sharply: "With the baseline, I couldn't move while watching my partner, so I often felt like I was just talking to

myself. With the HMD, I could see my partner and how they were walking, which allowed me to respond to both the environment and the topics they wanted to discuss" (P13).

**Stronger Sense of Presence:** Finally, operators described a stronger sense of presence and embodiment when using the HMD interface. This was summarized effectively by a participant who remarked, "The baseline felt like I was operating a robot by looking at a screen, but the HMD felt like I was directly seeing the world myself" (P10). This heightened sense of "being there" appeared to be a key factor in enabling the more natural and effective social interactions reported by operators.

## 7 Discussions and Limitations

Our proposed immersive and intuitive teleoperation interface is certainly over benefits compared to traditional commercially available telepresence interface in terms of the operator's awareness and social interactability. However, there are several clear limitations as well as discussion points from our results.

A primary set of limitations stems from the hardware and the resulting perceptual experience. Our system's reliance on a HMD currently limits its accessibility, as HMDs are not yet ubiquitous and can be cumbersome for prolonged usage. While not a measure vocal, some people mentioned in the interview that the headset is a bit bulky and heavy for them for prolonged usage.

Furthermore, HMDs are known to induce motion sickness. This issue was exacerbated in our study by the robot's physical vibrations, which caused camera shake and reportedly led to stronger feelings of motion sickness during our technical pilots. We tried to mitigate this by vibrating the controllers when the robot is moving, this still is not a solved issue. While we are not sure, we think that if the interaction duration was longer than our sessions, some people may have to quit due to motion sickness.

The performance of our interface is heavily dependent on a stable, high-bandwidth network connection to support real-time video streaming to the HMD. While our field experiment was conducted in an environment with reliable network coverage, deployments in areas with intermittent or low-bandwidth connectivity could suffer from increased latency and degraded video quality, which would severely diminish the user experience and the effectiveness of the interface. This can lead to overall increasing cognitive loads and frustrations to maintain the situation awareness. Fortunately, recent improvement on video encoding technologies and algorithms (e.g., VP9) may solve this limitation in near future.

Our system's audio capabilities need to be improved. During our implementation, we opted for a single-channel microphone, reasoning that common telepresence platforms such as Double Robotics interface or audio communication tools such as Google Meet transmit mono audio. In retrospect, this was a notable limitation. The lack of spatial audio made it difficult for operators to localize sounds in the remote environment, such as determining the direction of a companion's voice or other ambient social cues. In a dynamic social setting, the ability to discern the directionality of sound is crucial for maintaining situation awareness and achieving a genuine sense of social presence. Future iterations should incorporate a microphone array and binaural audio transmission to more accurately replicate the remote auditory scene.

While the control interface of a commercial telepresence robot is a good initial baseline condition, in retrospect, we acknowledge that we could have included additional conditions, such as a simple manual joystick with an HMD, to analyze the control scheme's effectiveness. Further, we had some limitations including the small sample size ($N$=13) and gender balance of our participants. We do not think these limitations would play a huge role in invalidating our result; however, we could have addressed them if we had a backup robot base. In our future work, we would like to prepare a backup robot and carefully recruit participants for having a better gender balance.

## 8 Conclusion and Future Work

The challenge of simultaneously navigating a remote robot and engaging in natural social interaction remains a significant barrier to achieving truly embodied telepresence. Task-oriented systems often neglect social nuance, while social telepresence platforms frequently sacrifice mobility. In this work, we addressed the challenge of interacting with a partner via a teleoperated robot in a social environment by designing, implementing, and evaluating a novel teleoperation interface that prioritizes both immersive awareness and fluid social navigation. Our system's core principles are the decoupling of the operator's navigation viewpoint from the robot's movement via a HMD and the offloading of cognitive load through autonomous navigation modes.

Our field experiment provided strong empirical evidence of our system's effectiveness. Compared to a baseline interface mimicking a commercial telepresence robot interface, our proposed system improved operators' situation awareness, control intuitiveness, and the overall quality of social interaction. Our work highlights the critical importance of human-centered interface design in enabling teleoperated robots to function not just as tools, but as effective social proxies (avatars) in dynamic, human-centric spaces.

Building on these findings, our future work will proceed along several promising avenues including integrating a more sophisticated audio system, mitigating the physical and perceptual discomfort associated with HMD, and tackling motion sickness issues. In addition, we aim to expand the social capabilities of our system beyond our testbed interactions. For example, exploring the effectiveness of our interface in a group setting or in a much busy and crowded environment seems a good starting point. We also plan to explore the integration of more expressive non-verbal communication channels, such as mapping operator gestures to robot movements or social cues, to provide the operator with a richer set of tools for social expression and engagement. By pursuing these directions, we hope to continue breaking down the barriers to much more intuitive and effective remote social interaction, bringing us closer to a future where physical distance no longer limits us.

## Acknowledgments

This work was supported by JST Moonshot R&D Grant JPMJMS2011, Japan, and was in part supported by JSPS KAKENHI Grant No. 24H00722, Japan.

We would like to express our gratitude to Dr. Shinichi Arakawa, Information Networking, Osaka University for their support on the local 5G network.

## References


[1] Kourosh Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, Jerry Pratt, Eiichi Yoshida, Serena Ivaldi, and Daniele Pucci. 2023. Teleoperation of Humanoid Robots: A Survey. *IEEE Transactions on Robotics* 39, 3 (2023), 1706–1727. doi:10.1109/TRO.2023.3236952

[2] Yevheniy Dmytriyev, Marco Carnevale, and Hermes Giberti. 2024. Enhancing flexibility and safety: collaborative robotics for material handling in end-of-line industrial operations. *Procedia Computer Science* 232 (2024), 2588–2597. doi:10.1016/j.procs.2024.02.077 5th International Conference on Industry 4.0 and Smart Manufacturing (ISM 2023).

[3] Teppo Jakonen and Heidi Jauni. 2024. Managing activity transitions in robot-mediated hybrid language classrooms. *Computer Assisted Language Learning* 37, 4 (2024), 872–895. arXiv:https://doi.org/10.1080/09588221.2022.2059518 doi:10.1080/09588221.2022.2059518

[4] Yixiang Jin, Daniel Alonso Paredes Soto, John Anthony Rossiter, and Sandor M. Veres. 2021. Advanced Environment Modelling for Remote Teleoperation to Improve Operator Experience. In *Proceedings of the International Conference on Artificial Intelligence and Its Applications* (Virtual Event, Mauritius) *(icARTi '21)*. Association for Computing Machinery, New York, NY, USA, Article 16, 8 pages. doi:10.1145/3487923.3487939

[5] Mitsuhiro Kamezaki, Junjie Yang, Ryuya Sato, Hiroyasu Iwata, and Shigeki Sugano. 2021. A situational understanding enhancer based on augmented visual prompts for teleoperation using a multi-monitor system. *Automation in Construction* 131 (2021), 103893. doi:10.1016/j.autcon.2021.103893

[6] Oussama Khatib, Xiyang Yeh, Gerald Brantner, Brian Soe, Boyeon Kim, Shameek Ganguly, Hannah Stuart, Shiquan Wang, Mark Cutkosky, Aaron Edsinger, Phillip Mullins, Mitchell Barham, Christian R. Voolstra, Khaled Nabil Salama, Michel L'Hour, and Vincent Creuze. 2016. Ocean One: A Robotic Avatar for Oceanic Discovery. *IEEE Robotics & Automation Magazine* 23, 4 (2016), 20–29. doi:10.1109/MRA.2016.2613281

[7] Kai Kruckel, Florian Nolden, Alexander Ferrein, and Ingrid Scholl. 2015. Intuitive visual teleoperation for UGVs using free-look augmented reality displays. *IEEE International Conference on Robotics and Automation (ICRA)* (2015), 4412–4417. doi:10.1109/ICRA.2015.7139809

[8] Ting-Han Lin, Hannah Dinner, Tsz Long Leung, Bilge Mutlu, J. Gregory Trafton, and Sarah Sebo. 2025. Connection-Coordination Rapport (CCR) Scale: A Dual-Factor Scale to Measure Human-Robot Rapport. arXiv:2501.11887 [cs.RO] https://arxiv.org/abs/2501.11887

[9] Henrique Martins, Ian Oakley, and Rodrigo Ventura. 2015. Design and evaluation of a head-mounted display for immersive 3D teleoperation of field robots. *Robotica* 33, 10 (2015), 2166–2185. doi:10.1017/S026357471400126X

[10] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. 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, 418–431. doi:10.1145/2818048.2819922

[11] Yeonju Oh, Ramviyas Parasuraman, Tim McGraw, and Byung-Cheol Min. 2018. 360 VR based robot teleoperation interface for virtual tour. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

[12] Samwel Opiyo, Jun Zhou, Emmy Mwangi, Wang Kai, and Idris Sunusi. 2021. A review on teleoperation of mobile ground robots: Architecture and situation awareness. *International Journal of Control, Automation and Systems* 19, 3 (2021), 1384–1407.

[13] Amy Phung, Gideon Billings, Andrea F. Daniele, Matthew R. Walter, and Richard Camilli. 2023. Enhancing scientific exploration of the deep sea through shared autonomy in remote manipulation. *Science Robotics* 8, 81 (2023), eadi5227. arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.adi5227 doi:10.1126/scirobotics.adi5227

[14] David Portugal, Maria Eduarda Andrada, André G. Araújo, Micael S. Couceiro, and João Filipe Ferreira. 2021. *ROS Integration of an Instrumented Bobcat T190 for the SEMFIRE Project*. Springer International Publishing, Cham, 87–119. doi:10.1007/978-3-030-75472-3_3

[15] Daniel J Rea, Stela H Seo, and James E Young. 2020. Social robotics for nonsocial teleoperation: Leveraging social techniques to impact teleoperator performance and experience. *Current Robotics Reports* 1, 4 (2020), 287–295.

[16] Stela H. Seo, Daniel J. Rea, Joel Wiebe, and James E. Young. 2017. Monocle: Interactive detail-in-context using two pan-and-tilt cameras to improve teleoperation effectiveness. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Lisbon, Portugal). IEEE, 962–967. doi:10.1109/ROMAN.2017.8172419

[17] Stela H. Seo, James E. Young, and Pourang Irani. 2017. Where are the robots? In-feed embedded techniques for visualizing robot team member locations. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 522–527. doi:10.1109/ROMAN.2017.8172352

[18] Stela H. Seo, James E. Young, and Pourang Irani. 2021. How are Your Robot Friends Doing? A Design Exploration of Graphical Techniques Supporting Awareness of Robot Team Members in Teleoperation. *International Journal of Social Robotics* 13 (7 2021), 725–749. Issue 4. doi:10.1007/s12369-020-00670-9

[19] Ashish Singh, Stela H. Seo, Yasmeen Hashish, Masayuki Nakane, James E. Young, and Andrea Bunt. 2013. An interface for remote robotic manipulator control that reduces task load and fatigue. In *2013 IEEE RO-MAN*. 738–743. doi:10.1109/ROMAN.2013.6628401


Received 2025-09-30; accepted 2025-12-01