Displays 84 (2024) 102781

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Exploration of exocentric perspective interfaces for virtual reality collaborative tasks[☆]

Lei Chen [a,1], Junkun Long [b,1], Rongkai Shi [c,d], Ziming Li [c,d], Yong Yue [c], Lingyun Yu [c], Hai-Ning Liang [b,*]

[a] *School of Information Engineering, Hebei GEO University, Shijiazhuang, China*
[b] *Computational Media and Arts Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*
[c] *School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China*
[d] *Department of Computer Science, University of Liverpool, Liverpool, UK*

ARTICLE INFO

Keywords:
Virtual reality
Interface design
World-in-miniature
Collaborative and social computing
User studies

ABSTRACT

Exocentric views play a pivotal role in computer-mediated collaboration, especially in Collaborative Virtual Environments (CVEs), where focusing on the actions and operations of collaboration partners is crucial. The exocentric perspective offers users a vantage point to ascertain the whereabouts and actions of their partners, enhancing spatial awareness and social presence in CVEs. Moreover, interacting via the Exocentric Perspective Interface (ExPI) can help users complete searching and manipulation tasks remotely and efficiently. This work investigates the potential benefits of two representative ExPIs, World In Miniature (WIM) and 2D Map, for VR collaboration. We conducted a user study with 36 participants (18 pairs) to compare WIM and the 2D Map against a baseline in a VR collaborative task encompassing a series of searching and manipulation tasks with different task complexities (Simple, Medium, and Complex). For the Baseline (BL) condition, participants were not provided with an Exocentric Perspectives Interface (ExPI) but instead were given a map of the virtual environment (VE). The results indicate that these two ExPIs significantly improved task performance, usability, social presence, and user experience while reducing VR sickness. In addition, we also found that WIM outperformed 2D Maps, especially in complex collaborative environments. Based on the findings, three design implications are proposed to guide the design of future VR collaboration systems.

1. Introduction

Virtual Reality (VR) technology immerses users in computer-generated three-dimensional (3D) environments using Head-Mounted Displays (HMDs) and other supporting equipment, such as trackers. Collaborative Virtual Environments (CVEs) are VR platforms that allow multiple users to interact and collaborate in a shared virtual space, fostering teamwork and improved performance [1–3]. However, the expansive nature of CVEs in VR often poses many challenges, particularly in reducing spatial awareness. Spatial awareness, the insight users accrue over time regarding their environment or workspace, decreases as the environment expands, which would confuse and overload users [4]. A notable repercussion of this is the diminished social presence. If collaborators are out of users' vicinity, users lack a feeling of social presence—the sense of being with others [5,6]. Social presence stands as a cornerstone in collaborative endeavors. It enables efficient communication and a fluid workflow, positively impacting task performance and user experience in collaborative scenarios [7–9].

Employing an exocentric perspective that displays the entire VR scene could provide users with spatial awareness cues, aiding in discerning the positions and actions of their collaborators. This knowledge can enhance social presence [10]. Researchers have proposed several techniques to provide exocentric perspectives. World In Miniature (WIM) [11] is a visual technique that provides an exocentric perspective with a miniaturized version of the virtual world, allowing users to observe and travel efficiently in the Virtual Environment (VE). The utility of WIM has been extensively explored, particularly in tasks involving navigation and locomotion within VR. For example, Berger et al. [12] uncovered that WIM significantly outpaced continuous motion and teleportation, two prevalent VR locomotion techniques, in scenarios of long-distance navigation. In addition to WIM, 2D Map is another visual tool that provides an exocentric view, presenting a top

---

[☆] This paper was recommended for publication by Prof Guangtao Zhai.
[*] Corresponding author.
E-mail address: hainingliang@hkust-gz.edu.cn (H.-N. Liang).
[1] Both authors contributed equally to this research.

https://doi.org/10.1016/j.displa.2024.102781
Received 20 February 2024; Received in revised form 9 June 2024; Accepted 14 June 2024

view of the VE. Like WIM, most research investigating 2D Maps has been geared towards aiding self-location and orientation for navigation [13–15]. However, most research on these two techniques has focused on single-user tasks and mainly used these tools to project the VE rather than aid collaboration and interaction. Our review reveals a notable gap in leveraging WIM or 2D Map for multi-user, collaborative tasks within VR.

Therefore, this paper aims to bridge this research gap by examining the potential benefits of two Exocentric Perspective Interfaces (ExPIs), WIM and 2D Map, to provide spatial awareness cues and interactive operations for visual search and manipulation tasks in VR collaboration. We hypothesized that,

- ***H1:*** Compared to not using an ExPI (i.e., a baseline condition), WIM and 2D Map would lead to better performance and experience in terms of the time taken to complete the task (task performance), system usability, perceived sense of social presence, and VR sickness.
- ***H2:*** WIM would outperform 2D Map. To confirm this, a user study was conducted to compare a baseline, 2D Map, and WIM in different task complexities (Simple, Medium, and Complex) using a test environment involving visual search and manipulation tasks in collaborative VR.

The results showed that both ExPIs significantly improved collaborative performance and experience. Furthermore, WIM led to better performance and experience than 2D Map and was more favorable, especially when the VE had a higher complexity. Our contributions are as follows:

1. A prototype supporting collaborative activities in VR with the aid of interactive exocentric views — WIM and 2D Map;
2. The results of a user study comparing the two types of interactive exocentric views against baseline condition (the absence of exocentric views) in collaborative VR scenarios for visual search and manipulation tasks; and
3. Three implications on the design and use of these exocentric views for VR collaboration scenarios.

2. Related work

2.1. Exocentric views

There are mainly two exocentric perspective metaphors in VR: World In Miniature (WIM) and 2D Map. They can provide a shared perspective to improve collaborators' understanding of the workspace. We next discuss these two exocentric perspective metaphors in more detail.

2.1.1. Word In Miniature (WIM)

WIM was initially proposed by Stoakley et al. [11] to provide an additional viewport as a 3D map to the VE. The exocentric view is represented via a scaled-down model attached to a user's hand. Subsequently, many extensions have been explored for the first generation of WIMs, such as allowing scaling and scrolling [16], automatically optimizing the perspectives [17], handling complexity and occlusions [18], and integrating multiple levels of visuals into the WIM [19]. Most focused on object selection, manipulation, and navigation in VEs for single users [20,21].

Some research has paid attention to WIM in collaboration settings [22,23]. For example, Stafford et al. [24] presented Hand of God on a WIM, the first collaborative WIM. The tabletop display user had a god's eye view of the virtual world and communicated with the VR users. On the other hand, Irlitti et al. [23] explored the effect of an exocentric world in miniature visualization in extended reality environments. Recently, Zhao et al. [25] developed L-WiM, an interactive user interface that links multiple WiMs for collaborative astronomical data exploration tasks. Chheang et al. [26,27] presented the group WiM system to support a guide for team navigation in VR environments.

From this review, we can see that WIM can support VR users in their collaborative tasks. However, more research is needed on evaluating WIM for visual search and manipulating collaborative tasks in VR.

2.1.2. 2D Map

2D Maps have been widely used and explored for navigation and orientation purposes [28,29]. They provide a top view of the scene to support a quick search for information. Darken and Sibert [30] conducted an early study investigating the use of the 2D Map in VR. They presented users with a 2D overview of the VE, which led to enhanced navigation and orientation. Subsequently, some researchers have focused on specific applications and challenges associated with the 2D Map in VR. For example, Burigat and Chittaro [31] examined the user experience aspect of the 2D Map. Darken et al. [32] and Ruddle et al. [33] found that using an overview map improved users' performance of way-finding tasks in VEs. In addition, Sjölinder et al. [34] found that an overview map helps users better and more precisely understand the layout of an information space. However, most of these studies have focused on single-user navigation and orientation tasks.

Some collaborative systems are providing 2D Map aids. For example, the Enscape real-time realistic rendering system provides a mini-map interface supporting teleportation for collaboration [35]. Arkio, a drawing collaborative VR software, uses a virtual camera to take screenshots as a shared workspace view [35]. However, these studies only introduced the related systems and interfaces. There is no evaluation of the effect of using a 2D Map as a collaborative spatial tool on improving performance and experience. Besides, most 2D Map aids only provide view functions and exclude direct interaction with these maps. Schafer et al. [36] presented a radar view technique as shared representations to provide awareness information in spatial collaboration tasks. Their follow-up study [37] integrated 2D and 3D representations in a qualitative study dealing with collaborative tasks of rearranging furniture in a virtual space. However, this prior work did not involve VR environments.

Based on the previous work we have reviewed, there is a limited exploration of the interactive 2D Map in collaborative VR scenarios. The potential benefits of 2D Map as an interactive exocentric tool to enhance spatial co-presence awareness and collaborative experience in VR still need to be explored.

2.2. Spatial awareness and VR sickness

Spatial awareness is broadly defined as the ability to understand the body's position relative to its surroundings. Providing exocentric perspectives to enhance users' spatial presence to improve collaboration has been explored and discussed in prior research [11,38,39]. For example, Leigh et al. [40] proposed using multiple perspectives in CVEs. In addition to the default view, users could also view the VR world from an exocentric perspective. Cho et al. [41] developed an exocentric viewer system that focused on visualizing the state of a slave robot. Based on this system, the operator could easily recognize the workspace context. However, we found that these researchers only used the exocentric perspective as a view tool in a single-user setting. In addition, they contributed to proposing a new system but did not show empirical evidence for its usability.

Interacting with the VE from an exocentric view also allows users to complete tasks effectively. It can also mitigate VR sickness, which can be caused by the frequent movements and head turns in the VE [42]. Berger et al. [12] showed that an interactive WIM interface could reduce the task completion time while causing less VR sickness compared with other techniques. Liao et al. [43] stated that using interactive 3D geo-browsers in pedestrian navigation benefited spatial knowledge acquisition and decision-making. However, they also only focused on single-user task scenarios. More research is needed on both interactive and view functions of exocentric views in collaboration.

2.3. Presence

Lee et al. [44] defined presence as a mental state where virtual objects are perceived as real. IJsselsteijn et al. [45] further divided it into two categories: spatial presence and social presence. Spatial presence can be described as the degree to which an individual experiences presence in the real world rather than the VE [46]. A high sense of spatial presence means people will not find the difference between the mediated environment and the physical world. This presence part is mainly related to the environment's vividness and spatial properties, which is not the main point of this paper.

On the other hand, social presence refers to the sense or feeling of being with someone else or others [47,48]. In the collaborative context, social presence refers in more detail to the ability of users to collect and maintain an understanding of their collaborators' actions in a shared workspace [49,50]. During collaboration in a shared environment, it is essential to know where the collaborators are and what they are doing, such as what object(s) they are interested in or interacting with [3,51]. When one user moves around or operates some objects in virtual environments, the collaborators do not understand others' spatial position and actions, resulting in a cognitive disconnect with associated communication [52]. That is, effectively expressing the users' intentions and noticing the collaborators' actions can promote collaboration performance and improve user experience.

Researchers have proposed valuable techniques to solve this problem [23,53]. Much research has explored the interaction process, presence, and awareness by sharing visual cues in collaborative scenarios. For example, virtual avatars have been explored to represent each collaborator and increase awareness of others in the shared VE [54]. As mentioned above, providing exocentric perspectives of the workspace to give users spatial awareness cues is another solution that should be noticed to enhance collaboration. However, we found that existing research only used an exocentric perspective of the workspace as a non-interactive observational tool. An exocentric perspective tool to enhance social presence and enable users to interact with it directly in VR collaboration scenarios has yet to be explored. Therefore, we aim to fill this gap in this work.

3. Design and implementation

In this section, we introduce two exocentric perspective interfaces (ExPIs) to improve social presence and provide rapid interaction for collaborative tasks in VEs, followed by a description of the test environment and the task we designed for evaluating the proposed interfaces.

3.1. Exocentric Perspective Interfaces (ExPIs)

An ExPI provides a "God's eye view" of the entire VE—it allows users to observe the whole scene, including elements and users' avatars, with an additional view in real-time. It is also an interactive interface allowing the user to input operations directly. Any changes made to the interface will be reflected in the VE. For instance, if a user moves an object to a new position in ExPI, the corresponding object in VE will also shift to the new position. In general, the ExPI serves as a reference frame and a means of interaction. Based on the prior work, we implemented two interactive visualization variations: WIM and 2D Map. WIM is a 3D map-like technique that provides a scaled-down VR scene (see Fig. 1a). In contrast, the 2D Map only provides a top view of the VE (see Fig. 1b). Next, we introduce the features of the proposed ExPIs.

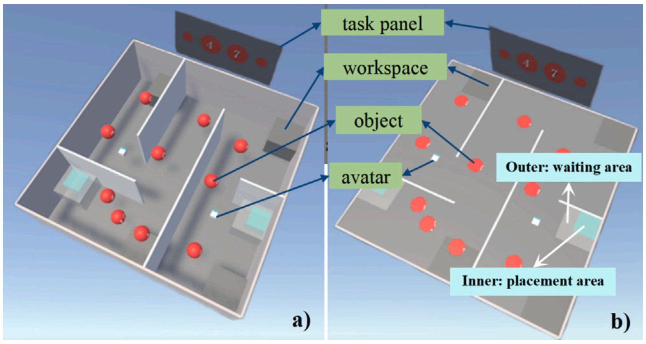

Fig. 1. An overview picture of two exocentric perspective interfaces, (a) WIM and (b) 2D Map.

3.1.1. Widget-based tools

In a VR collaborative task, users may be assigned different subtasks in different locations. To support wide usage scenarios, we make the ExPI "widget-like", mainly shown in three aspects. First, users have their own interface rather than a shared, fixed-positioned tool. It ensures that users can use the ExPI anytime and anywhere. Second, the ExPI is a paralleled interactive component rather than a separate mode so that users do not have to stop their current task and switch to a standalone interface for interaction, which enables fluid workflows. Finally, following the second feature, we allow users to place the ExPI in front of their viewpoint or pin it to their non-dominant hand. In the former case, users can translate or rotate the interface to use it efficiently without blocking their vision. In the latter case, attaching the interface to their non-dominant hand naturally follows the hand's movement.

3.1.2. Synchronous interactions

By providing an exocentric view or perspective, the current state of the object in VR scenes that needs to pay attention to can be dynamically provided by the exocentric view in real-time so that the changes of objects in VR scenes (including targets to be operated and operators) can be observed in the exocentric view in a time-saving and labor-saving manner. At the same time, these tools are designed to be interactable.

Users can directly manipulate objects in WIM or 2D Map with the raycasting tool, and then the corresponding objects in VR scenes will be changed synchronously (see Fig. 2). Users can manipulate the objects in WIM in the same way they interact in the VE. However, unlike the three-dimensional view provided by WIM, the 2D Map is a two-dimensional plane (only in an X–Z plane). So manipulating an object through a 2D Map can only move it on a plane, i.e., it can only change the X-axis and Z-axis position of the object. In addition to users' operations on the objects in the VE, users' locomotion behavior is also synchronized. When a user moves around the VE, his/her avatar will move simultaneously in all collaborators' ExPI. This approach was designed to give users a higher sense of social presence. Users could feel more connected with the collaborator by observing the avatar in real-time. The sense of social presence is an essential factor in collaboration and would affect users' experience and task performance.

3.2. Point to reveal more information

All spherical objects were uniformly sized (radius = 0.15 m), and each featured numerical markers on their front sides within the virtual environment (VE). To enhance the user experience, we implemented a technique where, upon pointing at a sphere using raycasting, the object would rotate to expose its front side (marked with a number) for quick verification. Once the ray moved away from the sphere, the object would return to its original position.

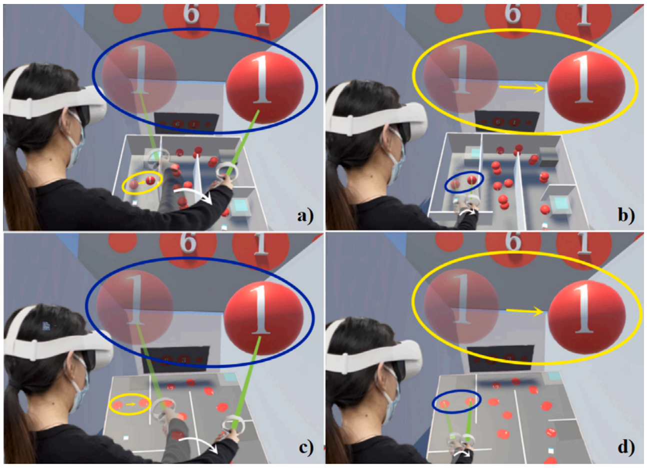

Fig. 2. The exocentric perspective interface synchronizes the changes in the virtual environment while the virtual environment also synchronizes the changes in the exocentric perspective interface. (a) and (b) show a user manipulating an object in the virtual environment and WIM, respectively; (c) and (d) show the same action for the virtual environment and 2D Map. The user's actions are in blue circles, while the synchronized changes are in yellow circles.

This functionality was also smoothly integrated into the ExPI. During the use of WIM, and especially the 2D Map, there might be a flattened perspective of the scene, which restricts users' view to the top of objects and the surrounding area. Whenever the ray intersected a sphere on the ExPI, its front side would be exposed to users. Additionally, the corresponding sphere in the VR scene would rotate as well, ensuring that its front side (with the number facing upwards) was clearly visible. This allowed users to easily identify the number associated with each sphere, even on the ExPI.

3.3. Test environment

As shown in Fig. 1, the test environment was a 6 m × 6 m maze-like VE with several interactable spherical objects, six workspaces, and a task panel. The red spherical objects were equal-sized (radius = 0.15 m) and marked with numbers. Workspaces were located at six fixed positions with inactive or active states. Users cannot interact with the inactive workspaces visualized, which is dark gray. Active workspaces consisted of waiting and placement areas in translucent gray and translucent blue, which would be used for the task (introduced soon in the following section). This setting can effectively simulate tasks that require users to operate simultaneously. For example, when a user first finds a target object, he/she can place it in the waiting area of a workspace and then observe/inquire about the other user's states and behaviors. Only when both users are ready can they start the next operation, which is to place it in the placement area at the same time. This process simulates the simultaneous operation of this task state very well. Besides, the task panel was a large rectangular plane to demonstrate the task specifications. The user's avatar was a cube (edge length = 0.25 m). The front side of the cube was in blue to indicate the direction the avatar was looking at.

The test environment supports two types of locomotion. Users can point and teleport far destinations using the Point & Teleport locomotion technique [55] or naturally move around. As mentioned before, the change of avatars representing two users will also be synchronously shown in WIM and 2D Map. On the other hand, to simplify the test environment, When a user employs raycasting to direct attention towards the sphere, the sphere will automatically orient itself to display the side with the numerical label, ensuring that the label is conveniently visible to the user. The object would return to its previous state when the ray moved away. This feature was also extended to ExPI (as mentioned in Section 3.2).

3.4. Task design

Each time participants entered the test environment, their avatars and objects were instantiated at random positions. Meanwhile, the system would activate two out of six workspaces and show two numbers on the task panel.

Participants are required to complete a compound collaborative task involving visual searching and object translation in the test environment. They must find the target objects with the given numbers shown on the task panel and place them into the active workspaces simultaneously and separately. Setting the task this way mainly aims to simulate real-life scenarios where collaborators need to observe the partner's behavior while operating some items, such as assembly training. When users are not co-located and visible to each other, providing an external perspective to help them see the partner's actions in real-time would be more conducive to collaboration efficiency and experience. In this case, the spatial awareness will be more important. We want to explore the effect of exocentric perspective metaphors on this type of collaborative task.

We had two designs to encourage collaboration and prevent "loners" from completing the task alone. First, the task required participants to put the target objects into the workspaces at the same time. This forced the two participants to actively communicate with each other and follow their partner's actions promptly. Second, each active workspace can only receive one object. The two participants had to discuss which workspace they should go for. In our pilot tests, we noticed that the participants who first got the target forgot to wait for their partner and placed the objects directly. To avoid such a case, we divided the workspace into an outer waiting area (in translucent gray) and an inner placement area (in translucent blue) to give participants explicit cues (see Fig. 1b). The participants who got the target faster could wait at the waiting area and later put the object into the placement area when the partner was ready. Overall, this task simplifies real collaborative scenarios that require users to complete two sub-tasks in parallel individually. For example, two users simultaneously click two buttons to turn on a machine at different locations. In this case, users must always consider their partner's actions.

Specifically, each trial task consisted of the following steps: (1) The system randomly showed two targets (denoted as T1 and T2) on the task panel. Two workspaces (W1 and W2) became on-state for participants to place the targets. (2) Two participants (P1 and P2) were required to search the two target objects in the VR room, 2D Map, or WIM. By communicating, they were asked to search for different targets to finish the task quickly, for example, P1 for T1 and P2 for T2. (3) When one target was found (e.g., T1) by one participant (e.g., P1), P1 needed to place T1 in the waiting area of one workspace (e.g., W1). Then, P2 was asked to find T2 and place T2 in the waiting area of the other workspace, W2. Before that, P2 needed to confirm which one of the two workspaces had been used by P1. P2 can know this by communicating with P1 or observing it on 2D Map and WIM if they have one. (4) P1 would observe T2's behavior. Once P1 saw that P2 was also ready (placing T2 in the waiting area of W2), they communicated to ask if they were prepared to move the targets into the placement areas. For instance, T1 might say, "Are you ready?", "When I count to "three", we simultaneously move them into placement areas, ok?". By doing so, participants simultaneously moved two targets to the placement areas. (5) When two targets were both in the placement areas, two participants clicked a button in the controller to confirm it. Then, one trial was finished, and the next one started (see Fig. 3).

4. User study

We conducted a user study to evaluate the performance and experience of using the ExPI to complete VR collaborative tasks against a baseline condition without an ExPI. We aim to address the following research questions (*RQ*):

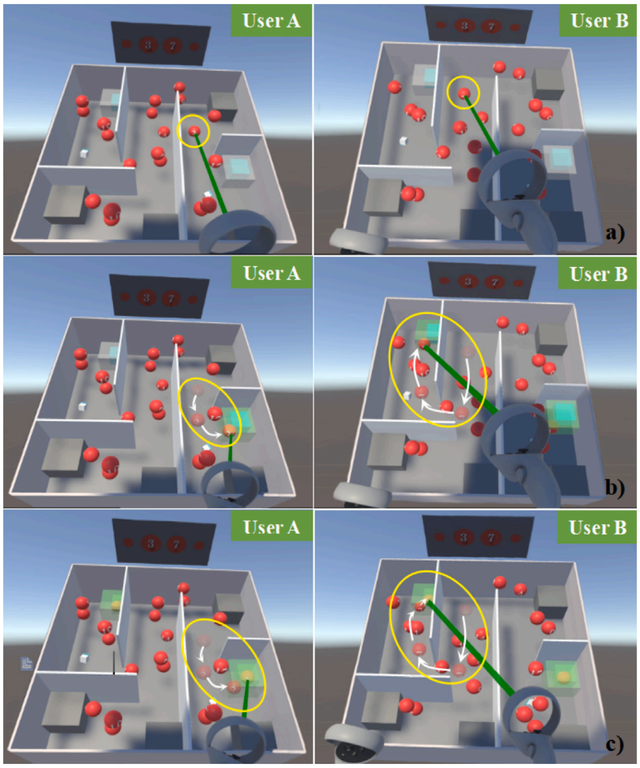

Fig. 3. An example of all the steps involved in completing the given collaborative tasks using WIM. (a) User A found "Object 3" and User B found "Object 7", the two target objects specified on the task panel. (b) After negotiation, User A selected and translated "Object 3" to the waiting area of the active workspace closest to her and waited for User B. At the same time, User B was translating "Object 7" to another active workspace. (c) User A and User B simultaneously placed the two targets in the placement area of two active workspaces and confirmed the completion of the task.

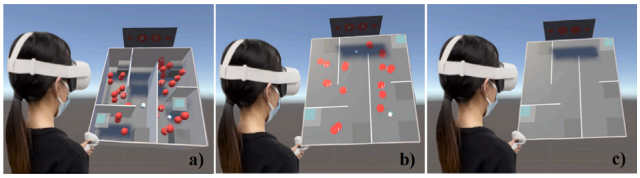

Fig. 4. One participant is observing details demonstrated in the (a) WIM, (b) 2D Map, and (c) map in the BL condition.

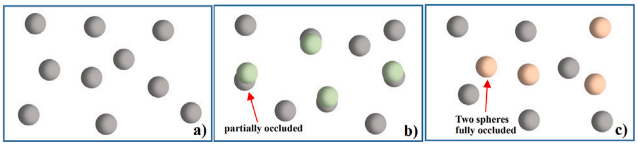

Fig. 5. Sketch maps (from a top view) of how the objects were occluded in the three complexity conditions: (a) no occlusion, (b) partial occlusion, and (c) full occlusion. The object would rotate to display its front side (marked with the number) using raycasting.

- ***RQ1***: *How does the ExPI affect task performance, and how is its usability in collaborative tasks with different complexities?* The ExPI, whether with WIM or 2D Map visualization variations, can support collaborative tasks. The WIM can help users complete tasks faster because the users can interact with and control the interactive 3D map intuitively. However, when the targets are very close to the users, they might find it more convenient to manipulate the object in VE directly than using the WIM. Besides, its effect on task performance and users' perceived usability is unclear, especially at different difficulty levels.
- ***RQ2***: *How does the ExPI affect the social presence in VR collaborative tasks?* The ExPI, as an interactive interface, offers a distinctive perspective for collaborative tasks in VR environments. By allowing users to observe real-time information about their own actions and those of their collaborators, it is expected that the ExPI has the potential to enhance social presence, particularly the co-presence among users within the VR environment.
- ***RQ3***: *How does the ExPI affect the VR sickness when completing the VR collaborative tasks?* VR sickness is generally induced when moving in VR environments [56]. An ExPI helps users observe and interact with the workspace remotely and keep updates on collaborators' states, which can potentially avoid frequent movement and head turns in VEs. Thus, it is supposed to reduce VR sickness in collaborative tasks.

4.1. Experiment design

This experiment followed a 3×3 within-subjects design with Technique (BL, 2D Map, and WIM) and Complexity (Simple, Medium, and Complex) as two independent variables. The order of Technique conditions was counterbalanced using a Latin square design to minimize learning effects. In each Technique condition, the order of Complexity was randomized. There were 3 trails for each condition. In total, each pair of participants needs to complete 27 trials (= 3 Technique × 3 Complexity × 3 trials).

Firstly, we planned to compare the two ExPI (WIM and 2D Map) with a BL (baseline) condition (see Fig. 4). For the Baseline (BL) condition, participants were not provided with an ExPI but instead were given a map of the VE (see Fig. 4c). This map was non-interactive and did not display objects or users, resembling a traditional map that depicts static terrain. Consequently, in the BL condition, participants had to manually manipulate the targets and move in the VE to complete each trial, as no assisting tools or clues were provided. In contrast, participants in the WIM and 2D Map conditions could accomplish these tasks using the ExPI. In short, we compare the use of ExPI (WIM and 2D Map) against its absence (Baseline).

Secondly, we wanted to evaluate the techniques with different environmental complexities. Thus, we set up three Complexity conditions: Simple (10 objects with no occlusion), Medium (20 objects with partial occlusion), and Complex (30 objects with full occlusion). We defined task complexity in this way based on prior work [51].

Given that ExPIs primarily provide a top view, the occlusion only occurred on the Y-axis in our experimental design. That is, when the coordinate values of the X and Z axes of two objects are the same, an object would be occluded by those with larger Y-axis values. In the Simple condition, no occlusion happened to any of the 10 objects. In the Medium condition, 20 objects were presented in the test environment, and half were partially occluded for 40% to 60% of their size. The Complex condition involved 30 objects, half of which were fully occluded. Fig. 5 demonstrates occlusion conditions from the top of the test environment.

4.2. Participants

We recruited 36 participants (15 females, 21 males) aged 18 to 28 ($M = 21.83$, $SD = 2.76$) from a local university. Only two pairs did not know each other before participating in the experiment. All participants had normal or corrected vision, and none reported known visual or vestibular disorders, such as color or night blindness. Eighteen participants (50%) who used VR HMDs less than once per month were not frequent VR users. The experiment was classified as low-risk research and was approved by the University Ethics Committee. All participants gave their consent to participate in the experiment.

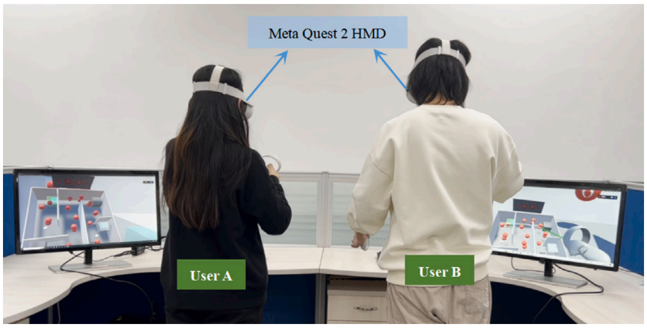

Fig. 6. Experimental setup. Two participants are completing tasks using a WIM, one of the exocentric perspective interfaces.

4.3. Apparatus and setup

As shown in Fig. 6, two Meta Quest 2 VR HMDs were used for the experiment. Quest 2's display is RGB LCD 1832 × 1920 display resolution per eye @ 90 Hz, and its FOV (Field of View) is 113.46° diagonally. They were connected to two desktop computers (Windows 10, 21H2) with an Intel Core i7-7700k CPU @ 2.9 GHz, 16 GB RAM, and an NVIDIA GeForce GTX 1080 Ti GPU. The cables for connection were long enough for the participants' movements during the experiment. The Quest 2 controllers were used as the input device. Users' heads and hands were tracked in 6 DoF (Degrees of Freedom) by the Quest 2 sensors to reproduce users' physical actions in the virtual environment via their avatars. The test environment and the techniques were developed with Unity (version 2021.3.8f1c1) and Photon Unity Networking (version 3.10).

4.4. Procedure

The whole experiment took about one hour to complete for each pair and involved 4 phases:

- ***Introduction*** (~3 min): we informed participants of the experiment's goal and ethics regulations. Then, we asked them to complete the consent form, a short questionnaire to collect demographic data, and an SSQ about the user's sickness level before the experiment [56].
- ***Training*** (~20 min): participants received several practice trials to familiarize them with the VR device, techniques, controls, and the task.
- ***Formal trials*** (~35 min): participants completed the experimental trials and filled in questionnaires after each condition. There was a short break between the two conditions.
- ***Interview*** (~3 min): we conducted a semi-structured interview to collect further feedback and comments.

4.5. Measurements

The three techniques, BL, 2D Map, and WIM, were compared using objective and subjective measurements.

- *Completion Time*: the completion time for each trial was recorded in a system log file to measure task performance (*RQ1*). In each trial, the completion time was defined as the time from the task panel showing a task till both participants clicked the button to confirm they completed the task.
- *System Usability Scale (SUS)* [57]: we used a SUS questionnaire to measure the usability of the interface (*RQ1*). SUS consists of 10 items on a 5-point Likert scale (1: Strongly Disagree ~5: Strongly Agree). Based on the ratings, we calculated the overall SUS scores [57] (ranging from 0 to 100) and used them for analysis. The higher the SUS score, the higher the perceived usability.
- *Social Presence Questionnaire* [58]: the NMM (Networked Minds Measure) Social Presence Questionnaire [58] was used to measure social presence (*RQ2*). The questionnaire contains nine 7-scale items (1: Strongly Disagree ~7: Strongly Agree). The results from the questionnaire are further summarized into three sub-scores: Co-Presence (CP), Attention Allocation (AA), and Perceived Message Understanding (PMU). In addition, we also calculated the Overall score. The results also show that the higher, the better.
- *Simulator Sickness Questionnaire (SSQ)* [56]: we used the SSQ to measure VR sickness (*RQ3*). SSQ is used to determine how severe users' VR sickness symptoms are. It asks participants to provide subjective severity ratings of 16 symptoms on a scale from 0 (no perception) to 3 (severe perception) after exposure to VR. The ratings were further processed to output four sub-scales: Nausea (SSQ-N), Oculomotor (SSQ-O), Disorientation (SSQ-D), and Total Severity (SSQ-TS).
- *User Experience Ratings*: we also asked participants to rate the techniques in terms of collaborative experience (CE), that is, their satisfaction with using the given technique to complete the collaboration task (1: Very Unsatisfied ~7: Very Satisfied). In addition, we also asked participants about the ease of use of the technique (EU)—how difficult they felt using the given technique to complete the task [59] (1: Very Easy ~7: Very Difficult).[2]
- *Preference and Feedback*: at the end of the experiment, participants were asked to choose the most preferred technique for each condition and provide feedback regarding their experience with each technique.

Besides the above measures, we recorded participants' behaviors and communications during collaboration (how pairs in each condition tended to work together and interact with each technique).

5. Results

In this section, we report the results, starting with the objective measures—completion time (Section 5.1), followed by subjective ratings collected from questionnaires (Sections 5.2 to 5.5).

We applied two-way repeated-measure Analysis of Variance tests (short as RM-ANOVA hereafter) for completion time and SUS scores to compare the effects of TECHNIQUE and COMPLEXITY. Shapiro–Wilk tests showed that both completion time and SUS scores were normally distributed ($p > .05$). We set significance level α to .05 and applied Bonferroni corrections to post hoc pairwise comparisons. In addition, Greenhouse–Geisser adjustments were used to adjust the degrees of freedom when the sphericity assumption was violated, and effect sizes (η_p^2) were reported whenever feasible.

For the data collected from the social presence questionnaire, SSQ, and user experience ratings that were ordinal, not normally distributed, and only compared among TECHNIQUE conditions, we applied non-parametric Friedman tests. Post hoc analysis was conducted using Wilcoxon signed-rank tests with Bonferroni corrections. M, SD, and Mdn are short for mean, standard deviation, and median, respectively.

5.1. Completion time

RM-ANOVA revealed significant main effects of TECHNIQUE ($F_{1.686,89.334} = 177.508$, $p < .001$, $\eta_p^2 = .770$) and COMPLEXITY ($F_{2,106} = 19.725$, $p < .001$, $\eta_p^2 = .271$), and a significant interaction effect

---

[2] Though the SUS questionnaire also has questions related to the ease-of-use of the technique, it is not recommended to analyze the subscales individually [60]. Thus, we adapted the question from the Single Easement Questionnaire [59] to cover this measure.

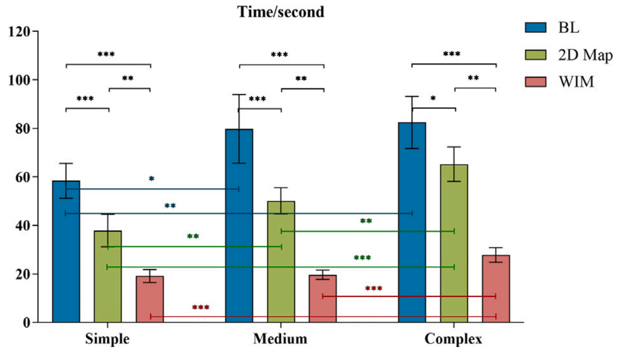

Fig. 7. Mean completion time (in seconds) by Technique and Complexity. Error bars indicate standard deviations, and asterisks indicate statistically significant effects (***: $p < .001$; **: $p < .01$; *: $p < .05$).

Table 1
Mean (standard deviation) of completion time in seconds.

Technique	Simple	Medium	Complex
BL	58.42 (26.39)	79.80 (51.81)	82.50 (39.35)
2D Map	37.88 (24.47)	50.14 (19.67)	65.24 (26.09)
WIM	19.17 (9.65)	19.66 (6.96)	27.82 (10.97)

Table 2
Mean (standard deviation) of SUS scores.

Technique	Simple	Medium	Complex
BL	42.50 (18.02)	38.82 (17.53)	35.21 (16.17)
2D Map	63.33 (18.74)	56.67 (17.38)	50.56 (18.17)
WIM	87.22 (11.13)	85.49 (12.01)	82.64 (13.54)

between Technique × Complexity on completion time ($F_{2.451,129.928} = 2.962$, $p = .045$, $\eta_p^2 = .053$). The results were summarized in Fig. 7 and Table 1.

Pairwise comparisons showed that WIM led to a significantly shorter completion time than 2D Map and BL for all Complexity conditions (all $p < .001$). Also, 2D Map was significantly faster than BL ($p < .001$ for Simple and Medium, and $p = .037$ for Complex).

Besides, BL took significantly less time to complete trials in Simple tasks than in Medium ($p = .028$) and Complex tasks ($p = .002$). Similarly, 2D Map took significantly less time in Simple tasks than in Medium ($p = .006$) and Complex ($p < .001$) tasks and significantly less time in Medium tasks than in Complex tasks ($p = .001$). Finally, WIM took significantly less time in Simple and Medium tasks than in Complex trials (both $p < .001$).

5.2. System Usability Scale (SUS)

Fig. 8 and Table 2 summarize the SUS results. Results from RM-ANOVA showed a significant main effect of Technique ($F_{2,70} = 105.303$, $p < .001$, $\eta_p^2 = .751$), a significant main effect of Complexity ($F_{1.209,42.310} = 30.231$, $p < .001$, $\eta_p^2 = .463$), and a significant interaction effect on SUS scores ($F_{1.898,66.436} = 5.126$, $p = .010$, $\eta_p^2 = .128$). Post hoc pairwise comparisons revealed that for all Complexity conditions, WIM obtained a significantly higher score than 2D Map and BL, and 2D Map got a significantly higher score than BL (all $p < .001$).

When using BL to complete the given tasks, it was rated better in Simple tasks than in Medium and Complex tasks and better in Medium tasks than in Complex tasks (all $p < .001$). 2D Map got higher scores in Simple tasks than in Medium ($p = .001$) and Complex ($p < .001$) tasks and a higher score in Medium tasks than Complex tasks ($p = .002$). As for WIM, it got a higher score in Simple and Medium tasks than in Complex tasks ($p = .007$ and $.021$, respectively).

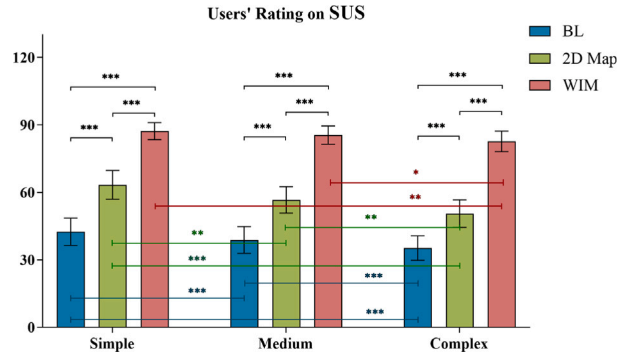

Fig. 8. Mean SUS scores by Technique and Complexity. Error bars indicate standard deviations, and asterisks indicate statistically significant effects (***: $p < .001$; **: $p < .01$; *: $p < .05$).

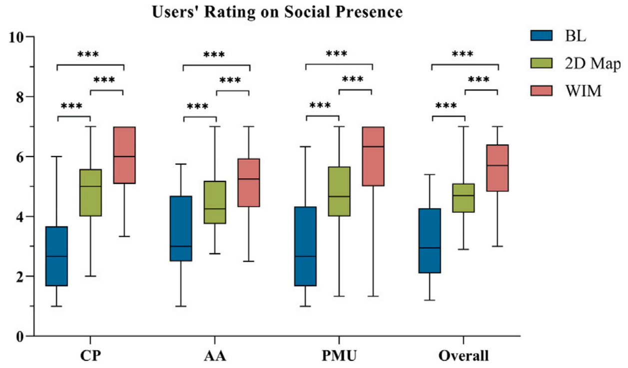

Fig. 9. Users' ratings from social presence questionnaire by Technique. (***: $p < .001$; CP: Co-Presence, AA: Attention Allocation, and PMU: Perceived Message Understanding).

5.3. Social presence questionnaire

Fig. 9 shows the users' ratings on social presence for all techniques. Friedman test showed that there was a significant difference in Overall scores among the three techniques ($\chi^2(2) = 66.526$, $p < .001$). Post hoc Wilcoxon signed-rank tests revealed that WIM ($Mdn = 5.70$) was rated significantly higher than 2D Map ($Mdn = 4.70$; $p < .001$) and BL ($Mdn = 2.95$; $p < .001$). Additionally, 2D Map scored significantly higher than BL ($p < .001$).

Friedman tests yielded significant main effects of Technique on CP ($\chi^2(2) = 60.941$, $p < .001$), AA ($\chi^2(2) = 31.581$, $p < .001$), and PMU ($\chi^2(2) = 57.662$, $p < .001$). Post hoc analysis revealed that WIM was rated significantly higher than 2D Map and BL, and 2D Map was significantly higher than BL for all three sub-scales (all $p < .001$). Table 3 summarizes the results.

5.4. Simulator Sickness Questionnaire (SSQ)

Fig. 10 and Table 3 demonstrate the results in SSQ. Results from Friedman tests showed that there were significant effects of Technique on SSQ-N ($\chi^2(2) = 16.021$, $p < .001$), SSQ-O ($\chi^2(2) = 15.942$, $p < .001$), SSQ-D ($\chi^2(2) = 12.923$, $p = .002$), and SSQ-TS ($\chi^2(2) = 17.802$, $p < .001$).

Regarding the total severity (SSQ-TS), post hoc tests showed that WIM ($Mdn = 0.00$) induced less VR sickness than 2D Map ($Mdn = 14.96$, $p < .001$) and BL ($Mdn = 13.09$, $p < .001$). For SSQ-N, post hoc tests showed that WIM ($Mdn = 0.00$) induced significantly lower nauseating sensation than 2D Map ($Mdn = 9.54$) and BL ($Mdn = 4.77$) (both p <.001). For SSQ-O, WIM ($Mdn = 0.00$) was rated significantly lower than 2D Map ($Mdn = 7.58$, $p = .001$) and BL ($Mdn = 7.58$, $p <$

Table 3
Friedman test results for social presence questionnaire, Simulator sickness questionnaire, and user experience ratings.

Median				Friedman tests	
Item	BL	2D Map	WIM	Chi-square statistics	Post hoc results
Social presence questionnaire (the higher, the better)					
CP	2.67	5.00	6.00	$\chi^2(2) = 60.941$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)
AA	3.00	4.25	5.25	$\chi^2(2) = 31.581$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)
PMU	2.67	4.67	6.33	$\chi^2(2) = 57.662$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)
Overall	2.95	4.70	5.70	$\chi^2(2) = 66.526$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)
Simulator sickness questionnaire (the lower, the better)					
SSQ-N	4.77	9.54	0	$\chi^2(2) = 16.021$, $p < .001$	WIM < 2D Map, WIM < BL (both $p < .001$)
SSQ-O	7.58	7.58	0	$\chi^2(2) = 15.942$, $p < .001$	WIM < 2D Map ($p = .001$), WIM < BL ($p < .001$)
SSQ-D	13.92	13.92	0	$\chi^2(2) = 12.923$, $p = .002$	WIM < 2D Map ($p < .001$), WIM < BL ($p = .003$)
SSQ-TS	13.09	14.96	0	$\chi^2(2) = 17.802$, $p < .001$	WIM < 2D Map, WIM < BL (both $p < .001$)
User experience ratings (the higher, the better)					
CE	2.00	4.00	7.00	$\chi^2(2) = 72.000$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)
EU	2.00	4.00	6.50	$\chi^2(2) = 71.042$, $p < .001$	WIM > 2D Map, WIM > BL, 2D Map > BL (all $p < .001$)

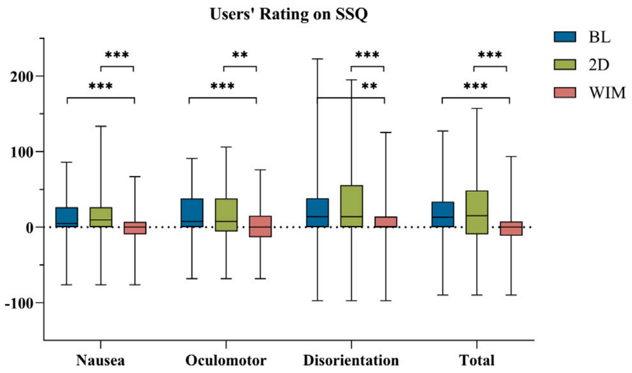

Fig. 10. Users' ratings on SSQ regarding all techniques with significance (***: $p < .001$; **: $p < .01$).

.001), indicating a lower sense of oculomotor. Similarly, in terms of SSQ-D, WIM ($Mdn = 0.00$) was rated significantly lower than 2D Map ($Mdn = 13.92$, $p < .001$) and BL ($Mdn = 13.92$, $p = .003$), indicating a lower sense of disorientation. No significant difference between 2D and BL was found for all four sub-scales (all $p > .05$).

5.5. User experience, preference, and feedback

Friedman tests showed significant main effects of Technique on the CE ($\chi^2(2) = 72.000$, $p < .001$) and EU ($\chi^2(2) = 71.042$, $p < .001$). Post hoc analysis with Wilcoxon signed-rank tests revealed that WIM was rated significantly higher than 2D Map and BL, and 2D Map also got significantly higher scores than BL for both items (all $p < .001$). The results indicate that participants felt WIM and 2D Map were better than BL regarding collaborative experience and ease of use. Between the two interfaces, WIM provided a better experience than BL. The results were summarized in Table 3.

We asked participants to choose the most preferred technique for each Complexity condition based on their preference. The preferences for Simple and Medium were similar, while the preference for the Complex condition was completely one-sided. For the Simple condition, 3 participants (8.33%) ranked BL as the most favored technique, 9 voted for 2D Map (25%), and 24 voted for WIM (66.67%). For the Medium condition, 2 participants (5.56%) voted BL as the most favored technique, 5 voted for 2D Map (13.89%), and 24 voted for WIM (80.56%). Meanwhile, for the Complex condition, all 36 participants preferred to use WIM. In the next section, we discussed our observations during the experiments and all other feedback collected from the interview.

6. Discussion

In this research, we explored the effects of two ExPIs, WIM, and 2D Map, for VR collaborative tasks with different levels of task complexity (Simple, Medium, and Complex). Overall, the results showed that the ExPIs benefited from the collaboration. We next discuss the findings in more detail combined with experimental observations and communications of participants.

6.1. Answers to the research questions

Results showed that using an ExPI for VR collaborative tasks yielded significantly higher efficiency and usability than not using one (i.e., baseline) for tasks with different complexities (***RQ1***). The results supported our ***H1***. Also, WIM (3D exocentric perspective) performed better than 2D Map (2D exocentric perspective) in terms of task efficiency, usability, and VR sickness mitigation. Significant differences existed in completion time between Medium (partial occlusion) and Complex (full occlusion) conditions for 2D Map, but not for WIM and BL. These results supported our ***H2***.

We inferred that the occlusion issues would be more obvious for the 2D Map because of its 2D view, which would greatly influence the task performance. Besides, there were similar findings for the perceived usability—the simpler the task was, the higher the perceived usability, which helps explain the significant reduction in task completion time between Complex and Simple trials. Some participants mentioned that they only found the difference between Simple and Complex conditions and between Simple and Medium conditions, but not for Complex and Medium conditions. One possible reason is that the difference between partial and full occlusion is limited. While occlusion occurs in the VE, it makes the task more complex than without occlusion in visual searching and object translation tasks in a VR collaborative scenario. Based on these results, WIM was more effective and useful for VR collaborative tasks involving visual searching and object manipulation, especially when the objects are partially or fully occluded. A 2D Map is also efficient and useful compared to a baseline case; however, it introduces ambiguity when the objects overlap and are viewed from the top.

In our study, the primary function of ExPIs is to offer real-time awareness of collaborators' activities, enabling basic operational support. Although it could handle some basic operations, our experimental focus is on promoting efficient and smooth team collaboration rather than pursuing highly complex control operations. Our aim is to create a simple and easy-to-use tool to supplement rather than replace direct operations. In this work, due to the relatively straightforward task, which only involves placing the target object in the designated work area, the experiment does not cover the steps that require more precise operations. Therefore, we did not particularly emphasize the accuracy issues caused by subtle differences in actions under different experimental conditions. In situations where precise operation is required, on-site direct operation will likely have advantages. Considering this, in our future work, we will consider adding more functionalities to ExPIs to allow it to support more complex tasks, such as zoom-in-out functionality for specific regions, making it a more effective supplementary tool for direct operations. At the same time, we will consider operational accuracy as an additional evaluation indicator. In this way, we can better understand the performance of ExPIs under different operational requirements to further optimize their functions and allow them to adapt to a wider range of collaborative needs.

The frequency of interaction can serve as an indicator of a user's workload and the ease of use of the implemented interaction method. Our observations of participants' behaviors revealed that the WiM and the 2D Map might necessitate more recurrent adjustments due to orientation manipulation, possibly leading to a higher number of interactions compared to the baseline condition. The 2D Map, however, confronts challenges with object occlusion and navigating a two-dimensional projection of a three-dimensional space. Consequently, we observed increased interactions among participants as they addressed occlusion problems. Conversely, the baseline condition could facilitate more direct manipulation, potentially resulting in a distinct interaction pattern. Moreover, individual participants exhibited diverse strategies, such as swiftly scanning the scene, meticulously inspecting each object, relocating nonessential items from their field of view, or momentarily releasing and re-grasping targets to enhance precision and accuracy. These observations indicate that interaction frequency could impact user experience. Thus, future research should contemplate incorporating interaction frequency as an ancillary metric to offer a more comprehensive evaluation of efficiency and user experience across various interaction modalities.

Compared to the baseline condition, both ExPIs received higher scores on all social presence sub-scales and supported the collaboration in our study setup (***RQ2***). The results also supported our ***H1***. We found that WIM received significantly higher ratings than 2D Map on overall social presence scores and all sub-scales between the two interfaces. This supported our ***H2***. The results suggest that WIM with a 3D exocentric perspective allows users to observe the operations and behavior intentions of the collaborator more accurately. By contrast, a 2D map only provides the movement on a plane (X–Z plane in our study), which may not show changes so clearly as in WIM. Participants' communication during collaboration also confirmed this point. When using the 2D Map, we observed that despite the collaborator's manipulation of the target, some participants still verbally inquired about the status of the task. For instance, questions such as "Have you found the target?" and "Have you put it in the waiting area?" were frequently asked, indicating a reliance on verbal communication to confirm the progress of the collaboration. This suggests that, even when visual cues and the spatial awareness provided by the 2D Map were available, participants still found it beneficial to verbally confirm information. In contrast, we did not observe similar verbal inquiries in the WIM condition. This indicates that the non-verbal feedback and implicit communication provided by WIM might have been sufficient for the participants to assess the status of the task without the need for explicit verbal confirmation. This finding suggests that the design of WIM might have facilitated a more implicit and efficient form of collaboration, reducing the need for explicit verbal communication. While visual cues and spatial awareness can provide valuable information, they might not always be sufficient, and explicit verbal communication can still play a crucial role. On the other hand, interfaces like WIM that provide implicit feedback might promote a more efficient and natural form of collaboration, reducing the cognitive load associated with explicit communication. In the BL condition, the absence of spatial awareness cues significantly altered the nature of participant interactions. Without the visual aids provided by the WIM and 2D Map, participants were forced to rely almost exclusively on verbal communication to coordinate their actions. Participants frequently asked questions like "Where are you?", "Are you ready?", and "Can we put them into the placement area now?" to establish a shared understanding of their relative positions and readiness to proceed. Additionally, there is a need to explicitly verify workspace locations, as can be seen from questions such as "Is it the one in the upper right corner?" or "Is it the one in the middle when facing the task panel?". These observations further underscored the importance of considering both explicit and implicit communication channels in collaborative VR environments. Based on the results, we can conclude that users can see the operations and positions of all avatars with an ExPI, which provides a significantly stronger feeling of being connected and helps them better understand their collaborator's actions and intentions. In addition, WIM is more beneficial for providing spatial awareness and a sense of co-presence, as shown in Section 5.3. For the 2D Map, participants commented that noticing small changes, including users' operations and movements, was not obvious, resulting in a lower sense of social presence.

We found that WIM significantly reduced VR sickness in the tasks compared to the 2D Map and BL conditions (***RQ3***). The results supported our ***H2***. We did not find a significant difference between 2D Map and BL on VR sickness. This did not support our ***H1***. Participants' behaviors during the experiment can help explain these results. With WIM, participants could find the target objects and translate them to the target workspace with fewer or even no movements and head turns; thus, it significantly mitigated the VR sickness. However, when using a 2D Map, participants may need to move around frequently for disambiguation.

Regarding collaborative experience, participants also expressed a preference for WIM. They thought WIM was easier to use than 2D Map and BL. Besides, 2D Map got higher ratings in collaborative experience and ease-of-use scores than BL. We also found that they always preferred WIM regardless of task complexity levels. The more complex the scenarios, the more participants preferred WIM.

Overall, our ***H1*** was partially supported and ***H2*** was fully supported. Providing an ExPI is recommended for enhancing spatial awareness and improving usability and task efficiency. Specifically, WIM, a 3D interactive exocentric interface used in this study, is particularly helpful when completing visual searching and manipulation tasks in collaborative VR. When few objects are in scenes, and the occlusion issue is not severe, 2D Map can also provide collaborative aids.

6.2. Design implications

Based on the results of our user study, we distilled the following design implications (*DI*) for future collaborative VR systems:

- ***DI1***: An ExPI can enhance the sense of social presence and collaborative experience for VR collaboration. It provides explicit spatial awareness cues to users. Users can notice each other's state in a timely.
- ***DI2***: An ExPI enables rapid interaction during VR collaboration. With an ExPI, users can interact with the elements in the VE remotely, simplifying the task and mitigating VR sickness.
- ***DI3***: WIM and 2D Map are recommended for the simple complexity collaborative VE. When the environment becomes complex, such as having a large number of interfering objects and the target object is fully occluded, a WIM-type technique is more usable, given its higher disambiguation capability.

6.3. Limitations and future work

We identified the following limitations from our research, which could serve as directions for future work. Firstly, we used a simplified collaborative task for experimental purposes. Future work could be conducted to evaluate the ExPIs in the field and examine their use in different collaboration scenarios. Secondly, to remove confounding variables, we used unified spherical objects and avatars in the test environment. ExPIs can be more useful in a high-fidelity test environment. For example, if users' hand and arm movements are rendered in the test environment and observable via ExPIs, users can see their collaborators' actions more clearly. In addition, to obtain a better understanding of users' specific interaction patterns, our follow-up work will include additional related metrics, such as interaction frequency and users' perceived workload.

In the future, we also plan to explore VR scenes on a larger scale. In these more complicated spaces with more items, users may need more time and effort to adapt and interact with ExPI. We want to add scalable features for ExPIs (e.g., Scaled Scrolling Worlds in Miniature [16]), which allows for dynamic adjustment of the WIM size. Besides, the interaction methods within WIM, including controller raycasting and gesture-supported drag-and-drop, are crucial topics worthy of further exploration. Furthermore, gaze analysis was not included in our study. It could provide additional insights into users' visual attention patterns [61,62], and incorporating gaze data could be beneficial in future studies. Finally, comparing communication patterns and strategies in cooperative and competitive tasks is intriguing. Understanding the nuances and differences in communication behaviors across different VR interactions will offer valuable insights.

7. Conclusion

In this paper, we evaluated two Exocentric Perspective Interfaces (ExPIs), World In Miniature (WIM) and 2D Map, on task performance, system usability, social presence, and VR sickness for collaboration tasks in Virtual Reality (VR). A user study compared them against a baseline condition with Simple, Medium, and Complex levels of task complexity. The study followed a within-subjects design (i.e., 3 techniques × 3 task complexities) with 18 pairs of participants. Based on the results, we found that the two ExPIs improved collaboration performance and experience. Specifically, WIM performed better than 2D Map on task efficiency. Users were more positive about the usability of WIM than 2D Map, especially in complex tasks. We also found that both ExPIs were useful for enhancing the sense of social presence and user experience when completing visual searching and manipulation tasks. Finally, WIM induced less VR sickness compared to 2D Map and baseline. Overall, we conclude that WIM and 2D Map are beneficial when completing visual searching and collaborative manipulation tasks. WIM is particularly useful when the collaborative environment is complex. Overall, ExPIs are useful in enhancing collaborative tasks in VR and should be provided in such scenarios.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank the participants who volunteered their time to join the user study. We also thank the reviewers whose insightful comments and suggestions helped improve our paper. This work was partially funded by the PhD Research Startup Foundation of Hebei GEO University, China (#BQ2024080), the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering, China (#SZS2022004), and the National Natural Science Foundation of China (#62272396).

References

[1] H.-N. Liang, F. Lu, Y. Shi, V. Nanjappan, K. Papangelis, Evaluating the effects of collaboration and competition in navigation tasks and spatial knowledge acquisition within virtual reality environments, Future Gener. Comput. Syst. 95 (2019) 855–866, http://dx.doi.org/10.1016/j.future.2018.02.029.

[2] L. Chen, H.-N. Liang, F. Lu, J. Wang, W. Chen, Y. Yue, Effect of collaboration mode and position arrangement on immersive analytics tasks in virtual reality: A pilot study, Appl. Sci. 11 (21) (2021) http://dx.doi.org/10.3390/app112110473.

[3] J. Lacoche, N. Pallamin, T. Boggini, J. Royan, Collaborators awareness for user cohabitation in co-located collaborative virtual environments, in: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, VRST '17, Association for Computing Machinery, New York, NY, USA, 2017, http://dx.doi.org/10.1145/3139131.3139142.

[4] C. Gutwin, S. Greenberg, A descriptive framework of workspace awareness for real-time groupware, Comput. Support. Coop. Work (CSCW) 11 (2002) 411–446, http://dx.doi.org/10.1023/A:1021271517844.

[5] F. Biocca, C. Harms, J.K. Burgoon, Toward a more robust theory and measure of social presence: Review and suggested criteria, Presence: Teleoper. Virtual Environ. 12 (5) (2003) 456–480, http://dx.doi.org/10.1162/105474603322761270, arXiv:https://direct.mit.edu/pvar/article-pdf/12/5/456/1623957/105474603322761270.pdf.

[6] H.H. Clark, S.E. Brennan, Grounding in communication, in: Perspectives on Socially Shared Cognition, American Psychological Association, 1991, pp. 127–149, http://dx.doi.org/10.1037/10096-006.

[7] S. Kim, M. Billinghurst, G. Lee, The effect of collaboration styles and view independence on video-mediated remote collaboration, Comput. Support. Coop. Work (CSCW) 27 (3) (2018) 569–607, http://dx.doi.org/10.1007/s10606-018-9324-2.

[8] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, M. Billinghurst, Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–13, http://dx.doi.org/10.1145/3290605.3300403.

[9] R.E. Kraut, M.D. Miller, J. Siegel, Collaboration in performance of physical tasks: Effects on outcomes and communication, in: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, CSCW '96, Association for Computing Machinery, New York, NY, USA, 1996, pp. 57–66, http://dx.doi.org/10.1145/240080.240190.

[10] B. Avery, C. Sandor, B.H. Thomas, Improving spatial perception for augmented reality X-Ray vision, in: 2009 IEEE Virtual Reality Conference, 2009, pp. 79–82, http://dx.doi.org/10.1109/VR.2009.4811002.

[11] R. Stoakley, M.J. Conway, R. Pausch, Virtual reality on a WIM: Interactive worlds in miniature, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '95, ACM Press/Addison-Wesley Publishing Co., USA, 1995, pp. 265–272, http://dx.doi.org/10.1145/223904.223938.

[12] L. Berger, K. Wolf, WIM: Fast locomotion in virtual reality with spatial orientation gain & without motion sickness, in: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, MUM '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 19–24, http://dx.doi.org/10.1145/3282894.3282932.

[13] S. Kratz, I. Brodien, M. Rohs, Semi-automatic zooming for mobile map navigation, in: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '10, Association for Computing Machinery, New York, NY, USA, 2010, pp. 63–72, http://dx.doi.org/10.1145/1851600.1851615.

[14] C.W. Nielsen, M.A. Goodrich, Comparing the usefulness of video and map information in navigation tasks, in: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI '06, Association for Computing Machinery, New York, NY, USA, 2006, pp. 95–101, http://dx.doi.org/10.1145/1121241.1121259.

[15] L. Chittaro, R. Ranon, L. Ieronutti, VU-flow: A visualization tool for analyzing navigation in virtual environments, IEEE Trans. Vis. Comput. Graphics 12 (6) (2006) 1475–1485, http://dx.doi.org/10.1109/TVCG.2006.109.

[16] C. Wingrave, Y. Haciahmetoglu, D. Bowman, Overcoming world in miniature limitations by a scaled and scrolling WIM, in: 3D User Interfaces, 3DUI'06, 2006, pp. 11–16, http://dx.doi.org/10.1109/VR.2006.106.

[17] R. Trueba, C. Andujar, F. Argelaguet, Complexity and occlusion management for the world-in-miniature metaphor, in: Smart Graphics, 2009, pp. 155–166.

[18] D. Coffey, N. Malbraaten, T.B. Le, I. Borazjani, F. Sotiropoulos, A.G. Erdman, D.F. Keefe, Interactive slice WIM: Navigating and interrogating volume data sets using a multisurface, multitouch VR interface, IEEE Trans. Vis. Comput. Graphics 18 (10) (2012) 1614–1626, http://dx.doi.org/10.1109/TVCG.2011.283.

[19] J.W. Nam, K. McCullough, J. Tveite, M.M. Espinosa, C.H. Perry, B.T. Wilson, D.F. Keefe, Worlds-in-wedges: Combining worlds-in-miniature and portals to support comparative immersive visualization of forestry data, in: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces, VR, 2019, pp. 747–755, http://dx.doi.org/10.1109/VR.2019.8797871.

[20] Y. Luo, J. Wang, Y. Pan, S. Luo, P. Irani, H.-N. Liang, Teleoperation of a fast omnidirectional unmanned ground vehicle in the cyber-physical world via a VR interface, in: Proceedings of the 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI '22, Association for Computing Machinery, New York, NY, USA, 2023, http://dx.doi.org/10.1145/3574131.3574432.

[21] K. Danyluk, B. Ens, B. Jenny, W. Willett, A design space exploration of worlds in miniature, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, New York, NY, USA, 2021, http://dx.doi.org/10.1145/3411764.3445098.

[22] H. Benko, E. Ishak, S. Feiner, Collaborative mixed reality visualization of an archaeological excavation, in: Third IEEE and ACM International Symposium on Mixed and Augmented Reality, 2004, pp. 132–140, http://dx.doi.org/10.1109/ISMAR.2004.23.

[23] A. Irlitti, T. Piumsomboon, D. Jackson, B.H. Thomas, Conveying spatial awareness cues in xR collaborations, IEEE Trans. Vis. Comput. Graphics 25 (11) (2019) 3178–3189, http://dx.doi.org/10.1109/TVCG.2019.2932173.

[24] A. Stafford, W. Piekarski, B.H. Thomas, HOG on a WIM, in: 2008 IEEE Virtual Reality Conference, 2008, pp. 289–290, http://dx.doi.org/10.1109/VR.2008.4480805.

[25] L. Zhao, N. Cao, S. He, H.-N. Liang, L. Yu, L-wim: Collaborative exploration in immersive environments, in: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct, 2022, pp. 118–123, http://dx.doi.org/10.1109/ISMAR-Adjunct57072.2022.00031.

[26] V. Chheang, F. Heinrich, F. Joeres, P. Saalfeld, B. Preim, C. Hansen, Group WiM: A group navigation technique for collaborative virtual reality environments, in: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW, 2022, pp. 556–557, http://dx.doi.org/10.1109/VRW55335.2022.00129.

[27] V. Chheang, F. Heinrich, F. Joeres, P. Saalfeld, R. Barmaki, B. Preim, C. Hansen, Wim-based group navigation for collaborative virtual reality, in: 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR, IEEE, 2022, pp. 82–92, http://dx.doi.org/10.1109/AIVR56993.2022.00018.

[28] M. Kraus, H. Schäfer, P. Meschenmoser, D. Schweitzer, D.A. Keim, M. Sedlmair, J. Fuchs, A comparative study of orientation support tools in virtual reality environments with virtual teleportation, in: 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR, IEEE, 2020, pp. 227–238, http://dx.doi.org/10.1109/ISMAR50242.2020.00046.

[29] S. Chen, F. Miranda, N. Ferreira, M. Lage, H. Doraiswamy, C. Brenner, C. Defanti, M. Koutsoubis, L. Wilson, K. Perlin, et al., Urbanrama: Navigating cities in virtual reality, IEEE Trans. Vis. Comput. Graph. 28 (12) (2021) 4685–4699, http://dx.doi.org/10.1109/TVCG.2021.3099012.

[30] R.P. Darken, J.L. Sibert, A toolset for navigation in virtual environments, in: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology, Association for Computing Machinery, New York, NY, United States, 1993, pp. 157–165, http://dx.doi.org/10.1145/168642.168658.

[31] S. Burigat, L. Chittaro, Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids, Int. J. Hum.-Comput. Stud. 65 (11) (2007) 945–958, http://dx.doi.org/10.1016/j.ijhcs.2007.07.003, URL https://www.sciencedirect.com/science/article/pii/S1071581907000985.

[32] R.P. Darken, J.L. Sibert, Navigating large virtual spaces, Int. J. Hum.-Comput. Interact. 8 (1) (1996) 49–71, http://dx.doi.org/10.1080/10447319609526140.

[33] R.A. Ruddle, S.J. Payne, D.M. Jones, The effects of maps on navigation and search strategies in very-large-scale virtual environments, J. Exp. Psychol.: Appl. 5 (1) (1999) 54, http://dx.doi.org/10.1037/1076-898X.5.1.54.

[34] M. Sjölinder, K. Höök, L.-G. Nilsson, G. Andersson, Age differences and the acquisition of spatial knowledge in a three-dimensional environment: Evaluating the use of an overview map as a navigation aid, Int. J. Hum.-Comput. Stud. 63 (6) (2005) 537–564, http://dx.doi.org/10.1016/j.ijhcs.2005.04.024.

[35] D. Ververidis, S. Nikolopoulos, I. Kompatsiaris, A review of collaborative virtual reality systems for the architecture, engineering, and construction industry, Architecture 2 (3) (2022) 476–496, http://dx.doi.org/10.3390/architecture2030027, URL https://www.mdpi.com/2673-8945/2/3/27.

[36] W.A. Schafer, D.A. Bowman, A comparison of traditional and fisheye radar view techniques for spatial collaboration, in: Proceedings of the Graphics Interface 2003 Conference, June 11-13, 2003, Halifax, Nova Scotia, Canada,Canadian Human-Computer Communications Society and A K Peters Ltd., 2003, pp. 39–46, URL http://graphicsinterface.org/wp-content/uploads/gi2003-5.pdf.

[37] W.A. Schafer, D.A. Bowman, Integrating 2D and 3D views for spatial collaboration, in: Proceedings of the 2005 ACM International Conference on Supporting Group Work, GROUP '05, Association for Computing Machinery, New York, NY, USA, 2005, pp. 41–50, http://dx.doi.org/10.1145/1099203.1099210.

[38] M. Billinghurst, H. Kato, I. Poupyrev, The MagicBook: a transitional AR interface, Comput. Graph. (2001).

[39] K. Kiyokawa, H. Takemura, N. Yokoya, A collaboration support technique by integrating a shared virtual reality and a shared augmented reality, in: IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028), Vol. 6, 1999, pp. 48–53, http://dx.doi.org/10.1109/ICSMC.1999.816444.

[40] J. Leigh, A. Johnson, C. Vasilakis, T. DeFanti, Multi-perspective collaborative design in persistent networked virtual environments, in: Proceedings of the IEEE 1996 Virtual Reality Annual International Symposium, 1996, pp. 253–260, http://dx.doi.org/10.1109/VRAIS.1996.490535.

[41] K. Cho, K. Ko, H. Shim, I. Jang, Development of VR visualization system including deep learning architecture for improving teleoperability, in: 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI, 2017, pp. 462–464, http://dx.doi.org/10.1109/URAI.2017.7992776.

[42] D. Monteiro, H.-N. Liang, X. Tang, P. Irani, Using trajectory compression rate to predict changes in cybersickness in virtual reality games, in: 2021 IEEE International Symposium on Mixed and Augmented Reality, ISMAR, 2021, pp. 138–146, http://dx.doi.org/10.1109/ISMAR52148.2021.00028.

[43] H. Liao, W. Dong, C. Peng, H. Liu, Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers, Cartogr. Geogr. Inf. Sci. 44 (6) (2017) 474–490.

[44] K.M. Lee, Presence, explicated, Commun. Theory 14 (1) (2004) 27–50.

[45] W.A. IJsselsteijn, H. De Ridder, J. Freeman, S.E. Avons, Presence: concept, determinants, and measurement, in: Human Vision and Electronic Imaging V, Vol. 3959, SPIE, 2000, pp. 520–529.

[46] J. Steuer, F. Biocca, M.R. Levy, et al., Defining virtual reality: Dimensions determining telepresence, Commun. Age Virtual Real 33 (1995) 37–39.

[47] F. Biocca, C. Harms, Defining and measuring social presence: Contribution to the networked minds theory and measure, Proc. Presence 2002 (2002) 7–36, http://dx.doi.org/10.1080/15230406.2016.1174886.

[48] T. Huang, Y. Li, H.-N. Liang, Avatar type, self-congruence, and presence in virtual reality, in: Proceedings of the Eleventh International Symposium of Chinese CHI, CHCHI '23, 2024, pp. 61–72, http://dx.doi.org/10.1145/3629606.3629614.

[49] C. Gutwin, S. Greenberg, M. Roseman, Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation, in: People and Computers XI: Proceedings of HCI'96, Springer London, London, 1996, pp. 281–298, http://dx.doi.org/10.1007/978-1-4471-3588-3_18.

[50] W. Xu, R. Zhen, D. Monteiro, V. Nanjappan, Y. Wang, H. Liang, Exploring the effect of display type on co-located multiple player gameplay performance, immersion, social presence, and behavior patterns, in: Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - GRAPP, 2024, pp. 159–169, http://dx.doi.org/10.5220/0012469000003660.

[51] L. Chen, Y. Liu, Y. Li, L. Yu, B. Gao, M. Caon, Y. Yue, H.-N. Liang, Effect of visual cues on pointing tasks in co-located augmented reality collaboration, in: Proceedings of the 2021 ACM Symposium on Spatial User Interaction, SUI '21, Association for Computing Machinery, New York, NY, USA, 2021, http://dx.doi.org/10.1145/3485279.3485297.

[52] J. Müller, R. Rädle, H. Reiterer, Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 6481–6486, http://dx.doi.org/10.1145/3025453.3025717.

[53] D. Gergle, R.E. Kraut, S.R. Fussell, Using visual information for grounding and awareness in collaborative tasks, Hum. Comput. Interact. 28 (1) (2013) 1–39, http://dx.doi.org/10.1080/07370024.2012.678246, arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07370024.2012.678246. URL https://www.tandfonline.com/doi/abs/10.1080/07370024.2012.678246.

[54] T. Piumsomboon, G.A. Lee, J.D. Hart, B. Ens, R.W. Lindeman, B.H. Thomas, M. Billinghurst, Mini-me: An adaptive avatar for mixed reality remote collaboration, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–13, http://dx.doi.org/10.1145/3173574.3173620.

[55] E. Bozgeyikli, A. Raij, S. Katkoori, R. Dubey, Point & teleport locomotion technique for virtual reality, in: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, in: CHI PLAY '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 205–216, http://dx.doi.org/10.1145/2967934.2968105.

[56] R.S. Kennedy, N.E. Lane, K.S. Berbaum, M.G. Lilienthal, Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness, Int. J. Aviat. Psychol. 3 (3) (1993) 203–220.

[57] J. Brooke, et al., SUS-A quick and dirty usability scale, Usability Eval. Ind. 189 (194) (1996) 4–7.

[58] C. Harms, F. Biocca, Internal consistency and reliability of the networked minds measure of social presence, in: Seventh Annual International Workshop: Presence, Vol. 2004, Universidad Politecnica de Valencia Valencia, Spain, 2004.

[59] J. Sauro, J.S. Dumas, Comparison of three one-question, post-task usability questionnaires, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, Association for Computing Machinery, New York, NY, USA, 2009, pp. 1599–1608, http://dx.doi.org/10.1145/1518701.1518946.

[60] J.R. Lewis, The system usability scale: Past, present, and future, Int. J. Hum. Comput. Interact. 34 (7) (2018) 577–590.

[61] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, H.-N. Liang, Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, 2023, http://dx.doi.org/10.1145/3544548.3581042.

[62] J. Moreno-Arjonilla, A. López-Ruiz, J.R. Jiménez-Pérez, J.E. Callejas-Aguilera, J.M. Jurado, Eye-tracking on virtual reality: a survey, Virtual Real. 28 (2024) http://dx.doi.org/10.1007/s10055-023-00903-y.